

Verarbeitung und mentale Repräsentation von Idiomen im Erwachsenen- und Kindesalter

DISSERTATION

der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
RUTH KESSLER, M.Sc.
aus Bad Hersfeld

Tübingen
2020

Gedruckt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Eberhard Karls Universität Tübingen.

Tag der mündlichen Qualifikation: 29.01.2021

Stellvertretender Dekan: Prof. Dr. József Fortágh

1. Berichterstatter: Prof. Dr. Claudia K. Friedrich

2. Berichterstatter: Prof. Dr. Andrea Weber

3. Berichterstatter: Prof. Dr. Eva Smolka

Zusammenfassung

Die vorliegende Dissertation untersucht Verarbeitungsmechanismen von Idiomen (z.B. *die Katze aus dem Sack lassen*) als Exemplar für feststehende Mehrworteinheiten (MWE) bei erwachsenen Muttersprachler:innen und sprachlernenden Kindern. Idiome sind konventionalisierte Redewendungen mit einer figurativen Phrasen-Bedeutung. Die vorliegende Dissertation befasst sich mit den Fragen, wie Idiome im mentalen Lexikon repräsentiert werden und wie sich diese Repräsentation mit zunehmendem Alter verändert.

Sowohl für den Erwerb als auch die Repräsentation im Erwachsenenalter wird angenommen, dass MWE als Mehrwort-Repräsentationen im mentalen Lexikon abgespeichert sind. Erwachsene wie auch Kinder verarbeiten häufige und bekannte MWE schneller als Nicht-MWE. Erwachsene Muttersprachler:innen sind ebenfalls sensitiv gegenüber der Mehrwort-Konfiguration und können einzelne Wörter innerhalb dieser vorhersagen. Entsprechende Befunde fehlen bislang für Kinder. Daher wurde in den **Studien I, II** und **III** untersucht, ob Kinder und Erwachsene idiom-finale Wörter anhand von Idiomfragmenten vorhersagen können (z.B. das Wort *Sack* anhand des Idiomfragmentes *Hannes ließ die Katze aus dem...*).

Da Einzelwort-Bestandteile innerhalb von MWE nicht notwendig für das Verstehen der Phrasen-Bedeutung sind, spielt es eine Rolle, ob diese dennoch analysiert werden. Hinsichtlich dieser Fragestellung gibt es widersprüchliche Befunde für Erwachsene. **Studie I** knüpft daran an und untersucht unter welchen Bedingungen Einzelwort-Bedeutungen innerhalb von Idiomen aktiviert werden. Bisher liegen diesbezüglich keine empirischen Befunde für Kinder vor, deshalb untersuchten **Studie II** und **III**, ob Kinder Einzelwort-Bedeutungen innerhalb von Idiomen aktivieren. Dieses wurde anhand von Verarbeitungsunterschieden zwischen Wörtern gemessen, die entweder semantisch relatiert (*Korb*) oder unrelatierten (*Bauch*) zu dem erwarteten Wort innerhalb des Idiomfragmentes waren (*Sack* in *Hannes ließ die Katze aus dem...*).

Die Fragestellungen wurden anhand eines semantischen Erwartungsparadigmas in Eye-Tracking- und EKP-Experimenten gemessen. Die empirischen Ergebnisse der **Studien I, II** und **III** zeigen, dass sowohl Kinder als auch Erwachsene idiom-finale Wörter vorhersagen. Daraus kann man schlussfolgern, dass Idiome bereits früh im Erwerb als Mehrwort-Repräsentation im mentalen Lexikon abgespeichert werden. Dennoch scheinen Erwachsene idiom-finale Wörter zeitlich früher vorherzusagen als Kinder. Sowohl Erwachsene als auch Kinder aktivieren die Einzelwort-Bedeutungen innerhalb von Idiomen. Jedoch erscheint diese Aktivierung für die

jeweiligen Altersgruppen zeitlich unterschiedlich während der Sprachverarbeitung zu erfolgen. Erwachsene Muttersprachler:innen zeigen eine frühe Aktivierung der Einzelwort-Bedeutung, die auf eine automatische Bedeutungsaktivierung hindeutet. Im Gegensatz dazu tritt der Effekt bei Kindern zeitlich später auf. Dies könnte darauf hindeuten, dass die mentale Repräsentation von Idiomen noch nicht sehr stark ist. Die Vermutung wird aufgestellt, dass die Unterschiede zwischen den Altersgruppen durch eine Stärkung Mehrwort-Repräsentation im mentalen Lexikon erklärt werden können. Die Stärkung der Repräsentation könnte daher resultieren, dass Kinder mit zunehmendem Alter entsprechende Ausdrücke häufiger erfahren.

Danksagungen

Mein besonderer Dank gilt Prof. Dr. Claudia Friedrich für die Betreuung dieser Dissertationsarbeit. Sie hat mich stets bei meinen Forschungsvorhaben unterstützt und mir den Freiraum gegeben, diese umzusetzen. Vielen Dank, dass du mir häufig geholfen hast, das auszudrücken, was ich eigentlich gemeint habe. Und dafür, dass du mich dazu ermutigt hast mein Dissertation mit meinen Abbildungen zu illustrieren.

Ich bedanke mich bei Prof. Dr. Andrea Weber für die Betreuung dieser Dissertationsarbeit. Vielen Dank für die schnellen und hilfreichen Rückmeldungen zu meinen Ausarbeitungen, die Zusammenarbeit in Projekt B9 und die Unterstützung bei der Organisation unserer Workshops.

Der Sonderforschungsbereich 833 (ID 75650358) machte es möglich diese Forschungsprojekte in dem Projekt B9 durchzuführen. Außerdem konnten wir in diesem Projekt zwei sehr bereichernde Workshops in Tübingen veranstalten. Dafür bin ich sehr dankbar.

Ein großer Dank gilt der gesamten Abteilung für Entwicklungspsychologie an der Universität Tübingen: An Dr. Anne Bauch sowieso für alles, die gemeinsamen Mittagspausen, jede Menge Faultiere, post-it und skype-Nachrichten und Käse-Abende. An Dr. Anne Rau für die schöne Zeit im gemeinsamen Büro, für die super Koordination der Studien. Ohne dich wäre ich wohl teilweise aufgeschmissen gewesen und hätte den Karren in den Dreck gefahren. An Dr. Ulrike Schild insbesondere dafür, dass du mir bei meinen R-Problemen immer unter die Arme gegriffen hast. An Dr. Kasia Patro-Nürk und Bori German für den frischen Wind in der Arbeitsgruppe. An alle Helferlein, die mich bei der Erstellung und Durchführung der Experimente unterstützt haben, ohne die es doch ein gutes Stückchen anstrengender geworden wäre: Babette Jakobi, Birte Herter, Charlotte Veil, Holger Schneider, Johanna Schläger, Julia Blumenthal, Julia-Larissa Maier, Kerstin Jendrysik, Matteo Marks, Miriam Wild, Sören Koch, Stacie Boswell und Thanh Lan Truong.

Vielen Dank an Sara Beck, für die vielen idiom-talks und die schönen und bereichernden Workshops, die wir gemeinsam organisiert haben - *we nailed it*. An Edith, die mir sehr viel Motivation gegeben hat am Ball zu bleiben und mich weiterzuentwickeln. An Beate für die ganzen unkomplizierten Abrechnung meiner doch sehr komplizierten Dienstreisen.

Herzlichen Dank an alle Erwachsenen und Kinder, die an den Experimenten teilgenommen haben, und an die Eltern und Schulen, die dies ermöglicht haben.

Ich danke meiner Familie dafür, dass sie einfach meine Stärken kennen und mich darin unterstützen. Dafür, dass ihr einfach immer selbstverständlich für mich da seid, und für alle Konsultationen.

Danke an all meine Freunde, die mich begleitet und unterstützt haben. Ein ganz besonderer Dank geht an meine jetzigen und ehemaligen Mitbewohner, die mir immer zur Seite gestanden haben. An Claudio für alle phd-talks, es war wirklich sehr bereichernd und beruhigend mit jemandem in einem ähnlichen Boot zu sitzen. An Noah und Ole für die persönlichen Einsichten in den Spracherwerb, für Minjago, Kuhgabeln und Tischbeine an Betten. An Judith für die Schokobons, die mich immer wieder erheitert haben – richtig gutes Zeug! An Sammy, für die offenen Ohren und die gemeinsame Zeit in unserem co-working space. An Clemens und Steffen dafür, dass unsere WG einfach nur noch nerdiger geworden ist. Nochmal an Steffen dafür, dass ich in seinem Arbeitszimmer meine Dissertation fertig schreiben durfte. An Julle und Dorian dafür, dass mich einfach so im Exil aufgenommen habt, ich mich mit ein bisschen Gartenarbeit ablenken konnte und genug Isolation zum Schreiben hatte. An Lisa dafür, dass du einfach so eine treue Seele bist.

Many thanks to Dr. Yuliya Shapovalova for encouraging and sympathizing with me and Tony's Chocolate supplies. Vielen Dank an Andrea Weibel fürs Gegenlesen des Mantelteils. An Dr. Simon Bretschneider für die Latexvorlage. Herzlichen Dank an Dorit Reitmeyer für die Unterkunft im Exil und die Quark-Himbeermarmelade-Schoko-Eingebung.

Zuallerletzt möchte ich mich bei Dorothea Knopp dafür bedanken, dass du mich ganz am Anfang dabei motiviert hast mein zweites Studium einfach gleich abzubrechen und doch lieber direkt mit der Promotion anzufangen. Ich hoffe, das hat gefruchtet.

Anmerkungen der Autorin

Die vorliegende Dissertation gliedert sich in mehrere Kapitel, die teilweise empirische Forschungsprojekte in Form von wissenschaftlichen Manuskripten mit Koautorinnen enthalten. Diese sind als getrennt lesbare Manuskripte gestaltet. Daraus ergeben sich inhaltliche Überlappungen zwischen den einzelnen Kapiteln.

Studie I in Kapitel 4.2 basiert auf dem auf dem Manuskript „Activation of literal word meanings in idioms: Evidence from eye tracking and ERP experiments“, das in der Fachzeitschrift *Language and Speech* veröffentlicht wurde. Das Manuskript wurde gemeinsam mit den Koautorinnen Prof. Dr. Andrea Weber und Prof. Dr. Claudia K. Friedrich erstellt. Die wissenschaftliche Idee und Konzeption der Experimente entstand als Zusammenarbeit der drei Autorinnen. Ruth Keßler führte hauptsächlich die Datenerhebung und -analyse durch. Stacie Boswell, Birte Herter, Babette Jakobi, Sören Koch, Matteo Marks, Dr. Anne Rau und Charlotte Veil waren an der Rekrutierung der Versuchspersonen, der Datenerhebung und -auswertung beteiligt. Ruth Keßler führte die schriftliche Ausarbeitung in Absprache mit Prof. Dr. Claudia Friedrich und Prof. Dr. Andrea Weber durch.

Studie II in Kapitel 6.2 basiert auf dem auf dem Manuskript „Delayed prediction of idiom constituent meaning points to weak holistic multi-word representation in children“, das in der Fachzeitschrift *Language, Cognition and Neuroscience* zur Veröffentlichung eingereicht wurde. Das Manuskript wurde gemeinsam mit der Koautorin Prof. Dr. Claudia K. Friedrich erstellt. Die wissenschaftliche Idee und Konzeption der Experimente entstand als Zusammenarbeit der zwei Autorinnen. Ruth Keßler führte hauptsächlich die Datenerhebung und -analyse durch. Julia Blumenthal, Babette Jakobi, Sören Koch, Julia-Larissa Maier, Dr. Anne Rau, Johanna Schläger, Holger Schneider, Charlotte Veil und Miriam Wild waren an der Rekrutierung der Versuchspersonen, der Datenerhebung und -auswertung beteiligt. Ruth Keßler führte die schriftliche Ausarbeitung in Absprache mit Prof. Dr. Claudia Friedrich durch. Im Rahmen dieser Arbeit wurden die Bachelorarbeiten von Babette Jakobi und Julia-Larissa Maier, sowie die Masterarbeit von Johanna Schläger betreut. Entsprechend bestehen Überlappungen zwischen den genannten Abschlussarbeiten und der vorliegenden Dissertationsarbeit.

Studie III in Kapitel 7.2 basiert auf dem auf dem Manuskript „L1 and L2 Learners keep their eyes on the prize: Eye-tracking evidence during idiom recognition“, das für die *Proceedings of Linguistic Evidence 2020* zur Veröffentlichung eingereicht wurde. Das Manuskript

wurde gemeinsam mit der Koautorin Sara D. Beck erstellt und erscheint ebenfalls in ihrer Dissertation (*Native and Non-native Idiom Processing: Same Difference*). Die wissenschaftliche Idee entstand gemeinsam mit Sara D. Beck als Teil des Projekt B9 (SFB 833). Ruth Keßler führte Experiment 1 durch, Experiment 2 wurde in Zusammenarbeit durchgeführt. Ruth Keßler führte die Analyse der Daten in Absprache Sara D. Beck durch. Die schriftliche Ausarbeitung erfolgte gemeinsam, hauptsächlich jedoch durch Sara D. Beck. Stacie Boswell, Birte Herter, Kerstin Jendrysik, Thanh Lan Truong und Miriam Wild waren an der Rekrutierung der Versuchspersonen, der Datenerhebung und -auswertung beteiligt. Prof. Dr. Claudia K. Friedrich und Prof. Dr. Andrea Weber haben eine frühere Version dieses Manuskriptes korrekturgelesen.

In der vorliegenden Arbeit werden empirische Forschungsarbeiten zitiert und beschrieben. Zur Illustration dieser werden Beispiele für das jeweils verwendete, linguistische Material gegeben. Diese werden zuerst in der ursprünglich verwendeten Sprache genannt und dann ins Deutsche übersetzt und gegebenenfalls deren Bedeutung gegeben. Die Beispiele werden im Text mit Anführungszeichen und in kursiver Schrift gekennzeichnet. In Klammern werden diese nur kursiv gekennzeichnet. Kritische Wörter oder Phrasen können durch Unterstreichung hervorgehoben werden. Aus Gründen der gendergerechten Sprache wird für die Bezeichnung von Personengruppen, die nicht in geschlechtsneutraler Form vorliegt, ein Gender-Doppelpunkt verwendet.

Inhaltsverzeichnis

Zusammenfassung	III
Danksagungen	V
Anmerkungen der Autorin	VII
1 Einleitung	1
1.1 Modelle zur Repräsentation und zum Erwerb von Mehrworteinheiten	3
1.2 Fragestellung der vorliegenden Dissertation	7
2 Ausgewählte Methoden zur Messung von Sprachverarbeitungsmechanismen	9
2.1 Offline-Methoden	11
2.2 Online-Methoden	12
3 Verarbeitung und Repräsentation von Mehrworteinheiten im Erwachsenenalter	20
3.1 Verarbeitungsvorteile für Mehrworteinheiten	20
3.2 Flexibilität der Wort-Konfiguration innerhalb von Mehrworteinheiten	23
3.3 Verarbeitung der Einzel-Bestandteile innerhalb von Mehrworteinheiten	24
4 Studie I	31
4.1 Forschungsfrage	31
4.2 Manuskript: Activation of literal word meanings in idioms: Evidence from eye tracking and ERP experiments	35
4.3 Schlussfolgerungen	70
5 Verarbeitung und Repräsentation von Mehrworteinheiten im Erstspracherwerb	77
5.1 Erfahrungsbedingter Verarbeitungsvorteil für Mehrworteinheiten	77
5.2 Verarbeitung der Einzel-Bestandteile innerhalb von Mehrworteinheiten im Spracherwerb	79
6 Studie II	81
6.1 Forschungsfrage	81
6.2 Manuskript: Delayed prediction of idiom constituent meaning points to weak holistic multi-word representation in children	84
6.3 Schlussfolgerungen	117
7 Studie III	120
7.1 Forschungsfrage	120
7.2 Manuskript: L1 and L2 learners keep their eyes on the prize: Eye-tracking evidence during idiom recognition	122

7.3 Schlussfolgerungen	142
8 Interpretation und Zusammenfassung	144
Anhang	XI
Abbildungsverzeichnis	XIII
Tabellenverzeichnis	XVII
Literaturverzeichnis	XVIII

1 Einleitung

den Reigen eröffnen

Um Sprache verstehen und produzieren zu können, muss sprachliches Wissen gemeinsam mit den Informationen, wie man dieses verwendet, verfügbar sein. Das psycholinguistische Modell des mentalen Lexikons befasst sich mit der Frage, wie gelerntes, sprachliches Wissen im Langzeitgedächtnis abgespeichert ist. Das heißt, das Modell des mentalen Lexikons beschreibt, wie sprachliche Einheiten im Langzeitgedächtnis repräsentiert und organisiert sind (Aitchison, 2012).

In dem Modell des mentalen Lexikons wird angenommen, dass sprachliche Einheiten eine mentale Repräsentation besitzen. Die Stärke einer solchen Repräsentation hängt mit der Auftretenshäufigkeit im Sprachgebrauch zusammen (vgl. Diependaele, Lemhöfer & Brysbaert, 2013; Joseph, Nation & Liversedge, 2013). Sprachliche Einheiten, die häufig im Sprachgebrauch auftreten, weisen eine stärkere Repräsentation auf als selten auftretende, sprachliche Einheiten. Mit den Repräsentationen sprachlicher Einheiten sind außerdem deren phonologische (die sprachlichen Laute betreffende), morphologische (die Wortstruktur betreffend), semantische (die Bedeutung betreffende), syntaktische (die Funktion im Satz betreffende) und orthographische (die Rechtschreibung betreffende) Merkmale verknüpft. Es wird angenommen, dass die Speicherung sprachlicher Einheiten netzartig im mentalen Lexikon strukturiert ist (Collins & Loftus, 1975). In dieser Struktur bestehen enge Verknüpfungen zwischen Wörtern, die zueinander related sind. Beispielsweise entstehen solche Relationen durch phonologische (*Hund* – *Mund*) oder semantische Ähnlichkeiten (*Hund* – *Katze*). Sobald die Informationen einzelner Einheiten im mentalen Lexikon aktiviert werden, werden gleichzeitig related Einheiten entlang der Verknüpfungen ko-aktiviert.

Hinsichtlich der Größe der sprachlichen Einheiten, die im mentalen Lexikon abgespeichert sind, gibt es divergierende Annahmen. Traditionelle Ansätze nehmen an, dass sprachliche Einheiten hauptsächlich als Wörter abgespeichert sind (generative Ansätze, z.B. Pinker, 1991). Neuere Ansätze stellen eine ausschließlich wort-basierte Repräsentation infrage (gebrauchsbasierte Ansätze, z.B. Abbot-Smith & Tomasello, 2006; Bod, 2006; Bybee & McClelland, 2005). Diese Ansätze postulieren, dass die Größe der Einheiten im mentalen Lexikon von der Größe der Einheiten im sprachlichen Input abhängig ist. Daher können Einheiten im mentalen Lexikon sowohl aus einzelnen als auch aus mehreren Wörtern bestehen. Im sprachlichen Input häufig wiederkehrende Wort-Konfigurationen (z.B. *das Tanzbein schwingen*) – sogenannte Mehrworteinheiten – werden demzufolge auch gemeinsam mental abgespeichert.

Mehrworteinheiten (MWE) bezeichnen feststehende, formelhafte Ausdrücke, die aus mehreren Wörtern bestehen (Überblick siehe Carrol & Conklin, 2015; Conklin & Schmitt, 2012; Siyanova-Chanturia, 2013; Siyanova-Chanturia & Martinez, 2015; Wray, 2005). Zu MWE zählen beispielsweise Kollokationen (*schwarzer Kaffee*), Binomiale (*Kind und Kegel*), Idiome (*die Katze aus dem Sack lassen*) oder auch feste Sprachkonventionen (*danke gleichfalls; Sehr geehrte Damen und Herren*). Sowohl die Wort-Konfiguration als auch die Phrasen-Bedeutung von MWE sind sehr stark konventionalisiert. Das bedeutet, dass die Wörter innerhalb der MWE, deren Verbindungen und die damit assoziierte Bedeutung relativ fixiert sind und über verschiedene Kontexte hinweg konstant bleiben. In der Regel sind erwachsene Muttersprachler:innen mit der konventionalisierten Konfiguration und Bedeutung von MWE sehr gut bekannt. Die unterschiedlichen Arten der MWE können sich hinsichtlich der Bildhaftigkeit ihrer Phrasen-Bedeutung unterscheiden. Die meisten MWE weisen eine wörtliche, kompositionelle Phrasen-Bedeutung auf (*danke gleichfalls*, Bedeutung: man bedankt sich und wünscht dem Gegenüber das Gleiche). Im Gegensatz haben andere MWE eine nicht-wörtliche Phrasen-Bedeutung (z.B. *schwarzer Kaffee*, Bedeutung: Kaffee ohne Milch und Zucker, oder *die Katze aus dem Sack lassen*, Bedeutung: ein Geheimnis verraten).

Merkmale von MWE sind unter anderem eine hohe Häufigkeit und Bekanntheit (vgl. Siyanova-Chanturia & Martinez, 2015; Wray, 2005, 2012). Die Auftretenshäufigkeit von bestimmten Ausdrücken kann anhand von Korpusdaten ermittelt werden. Im Gegensatz dazu wird die Bekanntheit von Ausdrücken anhand von subjektiven Bewertungen ermittelt (vgl. Bonin, Meot & Bugajska, 2013; Sprenger, la Roi & van Rij, 2019; Tabossi, Arduino & Fanari, 2011). Obwohl Häufigkeit und Bekanntheit unterschiedliche Konzepte beschreiben, sind sie dennoch eng miteinander verknüpft: häufig auftretende Ausdrücke werden als sehr bekannt bewertet (Sprenger et al., 2019). Für einige MWE ist charakteristisch, dass sie eine hohe Auftretenshäufigkeit haben. Andere MWE wie beispielsweise Idiome treten allerdings nicht sehr häufig auf (Moon, 1988). Trotzdem sind Idiome sehr bekannte Ausdrücke. Dies könnte beispielsweise daraus resultieren, dass sie andere besondere Merkmale aufweisen, die sehr salient sind (z.B. nicht-wörtliche Bedeutung, emotionale Valenz, siehe Reuterskiöld & Van Lancker Sidtis, 2013; Siyanova-Chanturia & Martinez, 2015).

Aufgrund ihrer besonderen Merkmale sind Idiome prominent für die Erforschung der mentalen Repräsentation von MWE. Idiome sind feststehende Redewendungen, deren figurative Bedeutung nicht anhand der einzelnen Phrasen-Bestandteile kompositionell erschlossen werden kann. Beispielsweise lässt sich die Bedeutung des Idioms „*die Katze aus dem Sack lassen*“ (ein Geheimnis verraten) nicht anhand der einzelnen Wörter „*die*“, „*Katze*“, „*aus*“, „*dem*“, „*Sack*“ und „*lassen*“ und deren gemeinsamer Interpretation erschließen. In diesem Fall ist eine Aktivierung der einzelnen Elemente überflüssig. Da die figurative Bedeutung von Idiomen stark konventionalisiert ist (Caillies & Declercq, 2011), könnte angenommen werden, dass

bekannte Idiome im mentalen Lexikon direkt mit deren Phrasen-Bedeutung verknüpft sind (z.B. Carrol & Littlemore, 2020; Titone & Connine, 1999; Titone, Lovseth, Kasparian & Tiv, 2019). Demzufolge wäre es nicht nur nicht möglich sondern auch nicht notwendig die Phrasen-Bedeutung kompositionell zu erschließen. Insgesamt könnte dies bedeuten, dass Idiome als große Einheiten im mentalen Lexikon abgespeichert sind.

Da insgesamt ein großer Anteil des alltäglichen Sprachgebrauchs aus wiederkehrenden, konventionalisierten MWE besteht, ist deren Repräsentation im mentalen Lexikon relevant für Annahmen bezüglich des mentalen Lexikons. Falls MWE tatsächlich als Mehrwort-Repräsentationen abgespeichert sind und direkte Verbindungen zu ihrer Bedeutung haben, würde dies gegen wort-basierte Ansätze sprechen (vgl. Pinker, 1991). Im Gegensatz dazu würde dies gebrauchsbasierte Ansätze (vgl. Bod, 2006) unterstützen. Gemäß diesen Ansätzen steht die Häufigkeit, mit der man einen Ausdruck im sprachlichen Input erfährt oder gebraucht, in Zusammenhang mit der Stärke der Repräsentation. Dementsprechend ist ausschlaggebend, wie sich die Repräsentation von MWE entwickelt. Daher stellt sich nicht nur die Frage, wie erwachsene Muttersprachler:innen sondern auch sprachlernende Kinder MWE im mentalen Lexikon repräsentieren.

1.1 Modelle zur Repräsentation und zum Erwerb von Mehrworteinheiten

Bezüglich der mentalen Repräsentation und des Erwerbs von MWE wurden verschiedene Modelle aufgestellt. Diese unterscheiden sich grundlegend darin, welchen Status im mentalen Lexikon sie MWE zuschreiben und inwiefern die interne Struktur dieser (z.B. die einzelnen Wörter) in deren Repräsentation verfügbar ist. Deshalb lassen sich die Modelle auf einem Spektrum zwischen nicht-kompositionellen, hybriden und kompositionellen Modellen einordnen (Abbildung 1).

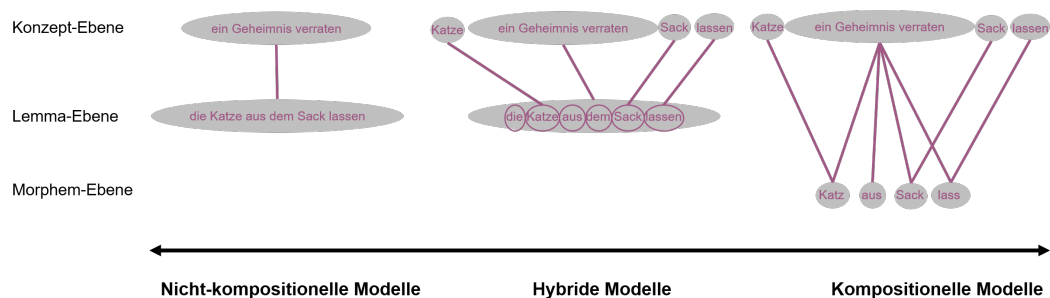


Abbildung 1: Schematische Darstellung der Modelle zur Repräsentation von MWE.

Nicht-kompositionelle Modelle nehmen generell an, dass die Wort-Konfiguration von MWE direkt mit der Phrasen-Bedeutung im mentalen Lexikon verknüpft ist und die Bedeutung daher nicht kompositionell generiert werden muss (z.B. Bobrow & Bell, 1973; Swinney & Cutler,

1979). Dadurch werden MWE vergleichbar zu langen Wörtern abgespeichert. Dies begründet sich beispielsweise im Verarbeitungsvorteil von MWE gegenüber neuartigen, kompositionellen Ausdrücken (siehe Kapitel 3.1), da die Generierung der Phrasen-Bedeutung länger dauern würde als der Abruf dieser aus dem mentalen Lexikon (Swinney & Cutler, 1979). Ähnlich gehen nicht-kompositionelle Modelle davon aus, dass MWE und insbesondere Idiome als eine Einheit erworben werden (Ackerman, 1982; Qualls, Treaster, Blood & Hammer, 2003; Reuterskiöld & Van Lancker Sidtis, 2013)

Hybride Modelle postulieren ebenfalls eine Mehrwort-Repräsentation von MWE. Gegenüber nicht-kompositionellen argumentieren sie allerdings, dass ebenfalls einzelne Bestandteile verfügbar sein können (vgl. Siyanova-Chanturia, 2015). Diese Annahme baut auf Befunden auf, die eine Aktivierung der Einzel-Bestandteile zeigen (Arnon & Cohen Priva, 2014; Beck & Weber, 2016a; Cacciari & Tabossi, 1988; Holsinger, 2013; Sprenger, Levelt & Kempen, 2006).

Kompositionelle Modelle gehen im Gegensatz zu nicht-kompositionellen oder hybriden Modellen davon aus, dass die Wort-Konfigurationen von MWE keinen besonderen Status im mentalen Lexikon aufweisen, da sie beispielsweise den gleichen syntaktischen Regeln folgen wie wörtliche, neuartige Ausdrücke (z.B. Konopka & Bock, 2009; Mancuso, Elia, Laudanna & Vietri, 2020; Smolka, Rabanus & Rösler, 2007; Tabossi, Fanari & Wolf, 2009).

Im Folgenden werden einzelne Modelle für die Repräsentation im Erwachsenen- und Kindesalter beschrieben. Aufgrund ihrer Prominenz bezieht sich ein Großteil der Modelle auf die Repräsentation und den Erwerb von Idiomen.

Ausgewählte Modelle zur Repräsentation von Mehrworteinheiten im Erwachsenenalter

Das *Dual-Route-Modell* (z.B. Van Lancker Sidtis, 2012; Wray, 1992) als ein Beispiel für nicht-kompositionelle Modelle postuliert, dass formelhafte und neuartige Ausdrücke dichotom verarbeitet und repräsentiert werden. Während formelhafte Ausdrücke als holistische Einheit repräsentiert werden, wird die Bedeutung neuartiger Ausdrücke kompositionell generiert. Laut dieses Modells besteht eine direkte Abbildung der Phrasen-Bedeutung auf die konventionalisierte Wort-Konfiguration wider. Durch diese direkte Abbildung sind die kompositionelle Interpretation oder die Aktivierung von Einzelwort-Bedeutungen nicht notwendig, sondern alle Bestandteile werden gleichzeitig als Teil der Konfiguration aktiviert. Daher ist ein direkter Abruf aus dem mentalen Lexikon möglich. Wray (2012) argumentiert, dass die direkte Abbildung durch wiederholten Gebrauch der Ausdrücke zustande kommt.

Ein Beispiel für hybride Modelle ist das *Superlemma-Modell* von Sprenger und Kolleg:innen (2006) für die Produktion von Idiomen. Gemäß diesem Modell sind Idiome wie

auch andere formelhafte Ausdrücke als sogenannte Superlemmas repräsentiert. Ein Superlemma wird aus der Verknüpfung der Einzelwort-Lemmas eines Idioms anhand von phrasalen Funktionen aufgebaut. Sowohl Superlemmas als auch Einzelwort-Lemmas eines Idioms sind mit den jeweiligen Konzeptebenen, auf der die Bedeutungen repräsentiert sind, verknüpft. Daher werden die Konzepte der Einzelwort-Lemmas bei der Aktivierung der Idiome ko-aktiviert. Diese Einzelwort-Lemmas innerhalb von Idiomen sind identisch zu denen innerhalb wörtlicher Ausdrücke (der *Sack* innerhalb des Idioms *die Katze aus dem Sack lassen* ist das gleiche Lemma wie der *Sack* in einem wörtlichen Kontext *den Müll in einen Sack schmeißen*). Insgesamt ähnelt die Verarbeitung und Repräsentation von MWE derer einzelner Wörter.

Ein weiteres hybrides Modell ist die *Konfigurations-Hypothese* (Cacciari & Tabossi, 1988). Diese postuliert, dass Idiome als Konfiguration der bestehenden Wörter abgespeichert sind. Ein Ausdruck wird so lange wörtlich interpretiert, bis genügend Informationen verfügbar sind, um diese Wort-Konfiguration als Idiom zu erkennen. Der sogenannte *Idiom Key* definiert den Punkt innerhalb des Idioms, an dem dieses eindeutig als solches erkannt wird (vertikaler Strich kennzeichnet Idiom Key innerhalb des Beispiels: ital. *era al settimo | cielo*, dt. *er war im siebten | Himmel*). Ab diesem Zeitpunkt wird die figurative Phrasen-Bedeutung aktiviert und eine wörtliche Interpretation und dementsprechend auch die Aktivierung der Einzelwort-Bedeutung des Ausdrucks ist weniger wahrscheinlich. Laut der Konfigurations-Hypothese hängt der Zeitpunkt des Erkennens von Faktoren wie der Vorhersagbarkeit der Idiome oder dem Kontext ab.

Das *multideterminierte Modell* geht in Anlehnung an die Konfigurations-Hypothese davon aus, dass nicht nur die Vorhersagbarkeit, sondern noch weitere Merkmale der jeweiligen Idiome dem direkten Abruf der figurativen Bedeutung aus dem mentalen Lexikon zugrunde liegen (Libben & Titone, 2008). Merkmale wie die Bekanntheit des jeweiligen Ausdrucks und die Plausibilität der wörtlichen Interpretation bestimmen den Zeitpunkt, ab wann der jeweilige Ausdruck als Idiom erkannt und die figurative Bedeutung aktiviert wird. Die Bedeutung von Ausdrücken, die früh erkannt werden, sind nicht kompositionell abgespeichert und werden direkt aus dem mentalen Lexikon abgerufen. Andere Ausdrücke werden hingegen kompositionell verarbeitet.

Häufigkeits-basierte Annahmen postulieren, dass der wiederholte Gebrauch einer MWE deren Repräsentation stärkt und dadurch eine effizientere Verarbeitung ermöglicht (z.B. Arnon & Cohen Priva, 2014; Arnon & Snider, 2010; Tremblay & Baayen, 2010; Tremblay, Derwing, Libben & Westbury, 2011). Die internale Struktur von MWE kann mitverarbeitet und deren Einzelbestandteile aktiviert werden. Diese Modelle sehen die Häufigkeit des Gebrauchs als grundlegend für den besonderen Status, der nicht einer rein holistischen Repräsentation entspricht.

Im Gegensatz dazu postuliert das *stamm-basierten Modell* (Smolka & Eulitz, 2020; Smolka et al., 2007), dass Idiome vergleichbar zu morphologisch komplexen Wörtern (wie zum Beispiel Komposita, *Schwachsinn*) verarbeitet werden. Die Kombination der einzelnen Morpheme bzw. Wörter aktiviert das damit verknüpfte Konzept, das im Fall von Idiomen deren figurativer Bedeutung entspricht. Die Häufigkeit, mit der bestimmte Morphem- oder Wort-Konfigurationen auftreten, bestimmt, wie stark die Verknüpfung zu dem jeweiligen Konzept ist. Da die figurative Bedeutung aus der Kombination dieser Wörter aktiviert wird, sind auch Einzelwort-Bedeutungen nach dem Erkennen eines Idioms verfügbar.

Ausgewählte Modelle zum Erwerb von Mehrworteinheiten

Reuterskiöld und Van Lancker Sittis (2013) nehmen an, dass Idiome auf Grund ihrer besonderen Merkmale (wie beispielsweise emotionale Valenz oder prosodische Merkmale) als holistische Einheit gelernt werden. In einer Studie konnten die Autor:innen zeigen, dass Kinder unbekannte Idiome besser erkennen und verstehen können, falls sie diese zuvor einmalig gehört haben. Dieser rapide Erwerb von Idiomen lässt sich nicht durch die Häufigkeit der Exposition erklären. Daher nehmen die Autor:innen an, dass Idiome aufgrund ihrer Besonderheiten qualitativ anders als neuartige Ausdrücke erworben werden (vgl. Dual-Route-Modell).

Sprenger und Kolleg:innen (2019) gehen ebenfalls davon aus, dass Idiome als Mehrwort-Repräsentationen erworben werden. Die Autor:innen postulieren, dass die Erwerbskurve von Idiomen der Erwerbskurve von Wörtern ähnelt: Im Kindesalter steigt das Idiomwissen zunächst steil an, flacht sich aber im frühen Erwachsenenalter ab und steigt nur noch moderat weiter an.

Die *Language-Experience-Hypothese* (Nippold & Rudzinski, 1993; Ortony, Turner & Larson-Shapiro, 1985) postuliert, dass die Häufigkeit der wiederholten, bedeutsamen Exposition und daher die Stärke der Bekanntheit mit den jeweiligen Idiomen eine Rolle beim Erwerb von Idiomen spielt. Unter bedeutsamer Exposition kann beispielsweise der Gebrauch von Idiomen in unterstützenden Kontexten verstanden werden. Positive Evidenz für diese Hypothese kommt von Studien, die zeigen, dass Kinder bekanntere Idiome besser verstehen als weniger bekannte Idiome (Nippold & Rudzinski, 1993; Nippold & Taylor, 2002, 1995). Ähnlich kommt Ackerman (1982) zu dem Schluss, dass eine häufigere Exposition erklärt, warum ältere Kinder und Erwachsene im Vergleich zu jüngeren Kindern die figurative Bedeutung von Idiomen auch ohne unterstützenden Kontext verstehen. Crutchley (2007) zeigte, dass Kinder Idiome mit steigender Korpushäufigkeit immer besser korrekt, figurativ verstehen. Daher nehmen diese Autor:innen an, dass Kinder Idiome nach häufiger Exposition als holistische Einheit erwerben.

Andere Ansichten gehen davon aus, dass Kinder im Allgemeinen Sprache als un-analyisierte Chunks erwerben (Cruttenden, 1981; Wray, 2005), da sie Wortgrenzen noch nicht erkennen können und daher diese nicht für die Segmentierung des sprachlichen Inputs miteinbeziehen. Daher extrahieren sie größere Einheiten aus der sprachlichen Umgebung.

Häufigkeits-basierte Modelle sehen ebenfalls eine „Untersegmentierung“ des sprachlichen Inputs und die Expositionshäufigkeit als grundlegend für den Erwerb von MWE (Arnon & Christiansen, 2017; Grimm, Cassani, Gillis & Daelemans, 2017; Lin, 2018). Kinder extrahieren MWE als Einheiten aus dem sprachlichen Input, und mit wiederholter Exposition verstärkt sich die Repräsentation dieser Wortgruppe. Gleichzeitig gehen sie davon aus, dass einzelne Bestandteile trotzdem verfügbar sind. Deshalb macht dieses Modell hybride Annahmen über die Repräsentation von MWE.

Im Gegensatz zu holistischen und hybriden Modellen geht das *Global-Elaboration*-Modell (Levorato, 1999; Levorato & Cacciari, 1992, 1995) davon aus, dass die gleichen Mechanismen für das Verstehen von Idiomen und wörtlicher Sprache gelten. Die Entwicklung allgemeiner kognitive Fähigkeiten (allgemeine Sprachfähigkeiten, Fähigkeit zur semantischen Analyse und die Fähigkeit kontextuelle Informationen zu nutzen) ermöglicht es Kindern, Idiome zu verstehen und zu erwerben. Beispielsweise konnten Cain, Oakhill und Lemmon (2005) und Cain, Towse und Knight (2009) zeigen, dass die korrekte Interpretation von Idiomen in Zusammenhang mit der allgemeinen Fähigkeit steht, Kontexte zu verstehen. Das *Global-Elaboration*-Modell beschreibt jedoch hauptsächlich, wie sich die Fähigkeit entwickelt bildliche Ausdrücke zu interpretieren, aber nicht, wie die mentale Repräsentation von Idiomen entsteht.

Die beschriebenen Modelle für die Repräsentation von MWE im Erwachsenenalter und im Erwerb unterscheiden sich hauptsächlich darin, wie groß die Elemente sind, mit denen die Phrasen-Bedeutung verknüpft wird. Zudem gehen einige Modelle davon aus, dass sich die mentale Repräsentation beispielsweise mit der Erfahrung verändern kann. Basierend auf empirischen Untersuchungen, die die Größe der sprachlichen Einheiten im mentalen Lexikon erforschen, könnte man für eines dieser Modelle argumentieren.

1.2 Fragestellung der vorliegenden Dissertation

Die vorliegende Dissertation soll die Architektur der Repräsentation und die Veränderung dieser mit dem Alter empirisch ergründen. Zum einen untersucht sie, ob MWE tatsächlich eine eigene Repräsentation im mentalen Lexikon aufweisen. Zum anderen soll die Natur dieser Repräsentation untersucht werden, indem die Rolle der einzelnen Bestandteile gemessen wird. Dies erfolgt exemplarisch an der Untersuchung von Idiomen, da die Aktivierung derer Einzelwort-Bedeutungen nicht relevant für das Verstehen der Phrasen-Bedeutung ist. Mit zunehmendem Alter steigt die Erfahrung und die Expositionshäufigkeit mit MWE (Sprenger et al., 2019). Daher wurden empirische Untersuchungen mit Versuchspersonen unterschiedlicher Altersgruppen (Erwachsene Muttersprachler:innen in **Studie I**, dreizehnjährige Kinder in **Studie III** und zehnjährige Kinder in **Studie II**) durchgeführt. Unterschiede zwischen den Altersgruppen könnten auf die erfahrungsbasierte Stärke dieser Repräsentation hinweisen.

Die Repräsentation von MWE im mentalen Lexikon kann indirekt über deren Verarbeitung gemessen werden. In Kapitel 2 werden relevante Methoden vorgestellt, die für die Messung der Verarbeitung von sprachlichen Einheiten im mentalen Lexikon geeignet sind. In Kapitel 3 werden Verarbeitungsmechanismen von MWE und Idiomen im Erwachsenenalter beschrieben. Wie schnell MWE verarbeitet werden und wie stark Wörter innerhalb der konventionalisierten Wort-Konfiguration erwartet werden, kann indirekte Hinweise darauf geben, ob diese direkt aus dem mentalen Lexikon abgerufen werden. Weiterhin ist bedeutsam, wie flexibel die Wort-Konfiguration ist und ob auf einzelne Bestandteile innerhalb von MWE zugegriffen wird. Kapitel 4 stellt empirische Befunde zur Verarbeitung von Idiomen im Erwachsenenalter dar und beinhaltet das Manuskript von Kessler, Weber und Friedrich (zur Veröffentlichung akzeptiert). Kapitel 5 beschreibt die bisherige Forschungslage zur Verarbeitung von MWE und Idiomen im Kindesalter und im Spracherwerb. Die Kapitel 6 und 7 beinhalten zwei empirische Arbeiten zur Verarbeitung von Idiomen im Spracherwerb (Kessler und Beck, eingereicht; Kessler und Friedrich, eingereicht). In Kapitel 8 erfolgt eine Zusammenfassung der einzelnen, empirischen Arbeiten und eine allgemeine Diskussion und Interpretation der Ergebnisse.

2 Ausgewählte Methoden zur Messung von Sprachverarbeitungsmechanismen

nicht auf den Kopf gefallen

Die Repräsentation sprachlicher Einheiten im mentalen Lexikon kann nicht direkt untersucht werden. Alternativ können Verarbeitungsmechanismen als Indiz für die Repräsentation dienen, weil die Aktivierung sprachlicher Einheiten von deren Repräsentation im mentalen Lexikon abhängt. Beispielsweise spiegelt sich die netzartige Organisation des mentalen Lexikons in der Ko-Aktivierung relatierter Wörter wider (z.B. *Hund* – *Katze*). Somit können indirekt Rückschlüsse auf die mentale Repräsentation sprachlicher Einheiten und deren Verknüpfungen gezogen werden. Die Aktivierung einzelner Wörtern und die Ko-Aktivierung relatierter Wörter kann über *Bottom-Up* und *Top-Down* Verarbeitungsmechanismen erfolgen.

Bei Bottom-Up Verarbeitungsmechanismen erfolgt die Aktivierung von sprachlichen Informationen aufgrund der tatsächlichen Präsentation derer im Input. Typischerweise wird dieser Verarbeitungsmechanismus innerhalb von Priming-Experimenten evoziert. Unter *Priming* (dt. Bahnung) versteht man die Bahnung bestimmter Gedächtnisinhalte aufgrund der Wahrnehmung eines relatierten Reizes. Beispielsweise kann die Wahrnehmung und Verarbeitung des Primewortes „*Hund*“ die Verarbeitung des nachfolgend präsentierten, semantisch relatierten Zielwortes „*Katze*“ erleichtern (z.B. Bentin, McCarthy & Wood, 1985, schematische Darstellung siehe Abbildung 2).



Abbildung 2: Schematische Darstellung für die Bottom-Up Aktivierung des Wortes „*Katze*“ nach der Präsentation des Wortes „*Hund*“.

Bei Top-Down Verarbeitungsmechanismen erfolgt die Aktivierung eines Wortes aufgrund der aus dem vorherigen Kontext aufgebauten Erwartung. Beispielsweise kann das Wort „Sand“ anhand des vorherigen Satzkontextes „Im Urlaub bauten sie eine Burg aus ...“ erwartet werden. Die *Cloze Probability* beschreibt die Wahrscheinlichkeit, mit der ein Satz mit einem bestimmten Wort vervollständigt wird (*cloze* = engl. Lückentext). Typischerweise erhalten Versuchspersonen Sätze, bei denen jeweils das letzte Wort oder auch die letzte Wortgruppe fehlt (z.B.: *Im Urlaub bauten sie eine Burg aus ...*), wie eine Art Lückentext (Taylor, 1953). Die Versuchspersonen sollen die Lücke mit einem Wort vervollständigen, von dem sie annehmen, dass es diesen Satzkontext sinnvoll vervollständigt (z.B. *Sand*). Anhand der Antworten der Versuchspersonen wird die Wahrscheinlichkeit berechnet, mit der bestimmtes Wort erwartet wird. Darüber hinaus kann ein Satzkontext weniger oder stärker die Erwartungen für ein bestimmtes Wort einschränken. Der oben genannte Beispielsatz schränkt die Erwartung eines bestimmten Wortes stark ein. Im Gegensatz dazu schränkt der Satz „Hannes rieb sich Käse über seine ...“ weniger stark ein. Obwohl ein essen-relatiertes Wort wahrscheinlich ist, könnte der Satz mit mehreren Wörtern sinnvoll vervollständigt werden, wie beispielsweise „Nudeln“, „Spaghetti“ oder „Pizza“. Die Erwartung eines Wortes innerhalb eines Satzkontextes (*Sand* in *Im Urlaub bauten sie eine Burg aus ...*) kann anhand von Verarbeitungsunterschieden zwischen erwarteten Wörtern und nicht-erwarteten Wörtern (*Kuchen*) gemessen werden. Innerhalb eines Satzes, das mit einem unerwarteten Wort vervollständigt wird, wird das erwartete Wort nie direkt präsentiert. Unerwartete Wörter können gleichzeitig überlappende Merkmale zu dem erwarteten Wort aufweisen (z.B. „Felsen“ ist semantisch relatiert zu „Sand“). Auf diese Weise kann man untersuchen, inwiefern diese Merkmale während der Erwartung ko-aktiviert werden. Inwiefern jedoch innerhalb wörtlicher, neuartiger Satzkontexte tatsächlich spezifische Wörter vorhergesagt werden oder eher zugrundeliegende Konzepte, ist bisher noch umstritten (Überblick siehe Van Petten & Luka, 2012).



Abbildung 3: Schematische Darstellung für die Top-Down Aktivierung des Wortes „Felsen“ nach der Erwartung des Wortes „Sand“.

Die beschriebenen Verarbeitungsmechanismen lassen sich anhand von Verarbeitungsgeschwindigkeiten wie Reaktions-, Lese- oder Produktionszeiten, von Blickbewegungen im *Eye-Tracking* oder Amplitudenunterschieden in ereigniskorrelierten Potentialen (EKP) messen. Diese Messmethoden unterscheiden sich grundsätzlich darin, ob sie nach der Sprachverarbeitung (offline) oder während der laufenden Sprachverarbeitung (online) stattfinden können. Im Folgenden werden Messmethoden und empirische Befunde zur Verarbeitung einzelner Wörter und

deren semantischen Merkmale bei erwachsenen Muttersprachler:innen und Kindern beschrieben. Diese Befunde bieten die Grundlage für die empirische Untersuchung der Repräsentation und Verarbeitung von MWE und einzelner Wörter innerhalb dieser, mit der sich die vorliegende Dissertation befasst.

2.1 Offline-Methoden

In sogenannte Offline-Methoden erfolgt die Messung nach der automatischen Sprachverarbeitung, z.B. nachdem ein Wort oder ein Satz gehört oder gelesen wurde. Ein Beispiel für diese Methode ist die lexikalische Entscheidungsaufgabe (Meyer & Schvaneveldt, 1971). Bei dieser Aufgabe lesen oder hören Versuchspersonen Zielwörter, die entweder tatsächliche Wörter oder Pseudowörter sind. Pseudowörter haben keine Bedeutung, aber entsprechen den phonotaktischen Regeln der jeweiligen Sprache (z.B. *Bamse* im Deutschen). Die Versuchspersonen entscheiden für jedes Zielwort, ob es sich um ein Wort oder ein Pseudowort handelt. Reaktionszeiten auf diese Zielwörter zeigen an, wie schnell die jeweiligen Wörter verarbeitet werden. Beispielsweise können so Priming-Effekte gemessen werden. In diesem Fall werden sukzessiv Primewörter und Zielwörter präsentiert. Für die Zielwörter führten Versuchspersonen eine lexikalische Entscheidungsaufgabe durch. Durch die Geschwindigkeit und Genauigkeit der Reaktionen auf die Zielwörter kann man messen, welche Informationen der Primewörter aktiviert werden. So zeigen sich zum Beispiel semantische Priming-Effekte. Erwachsene und Kinder können schneller auf das letzte Wort innerhalb semantisch relatierter Wortpaare (z.B. engl. *dog – cat*, dt. *Hund – Katze*) reagieren als auf das letzte Wort innerhalb unrelatierter Wortpaaren (z.B. engl. *car – pen*, dt. *Auto – Stift*) (z.B. Bentin et al., 1985; Radeau, 1983).

Anhand einer lexikalischen Entscheidungsaufgabe lassen sich auch Erwartungsprozesse messen. In diesem Fall treffen Versuchspersonen eine lexikalische Entscheidung zu Zielwörtern innerhalb von Sätzen (vgl. Rommers, Dijkstra & Bastiaansen, 2013). Typischerweise sind Reaktionen kürzer für Wörter, die sehr stark erwartet wurden (ndrl. *Na de Lunch draaide de electricien het nieuwe peertje in de lamp gisteren.*, dt. *Nach dem Mittagessen drehte der Elektriker gestern die neue Glühbirne in die Lampe.*) als für unerwartete Wörter (ndrl. *Na de Lunch draaide de electricien het nieuwe peertje in de kaars/vis gisteren.*, dt. *Nach dem Mittagessen drehte der Elektriker gestern die neue Glühbirne in die/den Kerze/Fisch.*) (Rommers, Dijkstra & Bastiaansen, 2013). Falls das unerwartete Wort semantisch relativiert (*kaars*, dt. *Kerze*) zu dem erwarteten Wort war, reagierten erwachsene Versuchspersonen schneller auf dieses als auf unrelatierte, unerwartete Wörter (*vis*, dt. *Fisch*). Dies zeigt, dass bei der Erwartung eines Wortes auch dessen semantischen Merkmale ko-aktiviert werden. Für Kinder im Alter zwischen sechs und 14 Jahren konnten ähnliche semantische Erwartungs-Effekte innerhalb einer auditiven Wortwiederholungsaufgabe gefunden werden (Mahler & Chenery, 2018).

2.2 Online-Methoden

Im Gegensatz zu Offline-Methode messen sogenannte Online-Methoden während der automatischen Sprachverarbeitung, z.B. während ein Wort oder ein Satz gehört oder gelesen wird. Zu diesen Methoden zählen beispielsweise die Messung von Blickbewegungen (Eye-Tracking) und EKP. Gegenüber Offline-Methoden bieten Eye-Tracking- und EKP-Methoden den Vorteil, dass die Sprachverarbeitung mit einer hohen zeitlichen Auflösung gemessen werden kann. Außerdem stellen sie eine Möglichkeit dar, implizite Sprachverarbeitungsmechanismen zu messen, ohne dass sie eine willkürliche Reaktion der Versuchspersonen erfordern. Aus diesem Grund sind beide Methoden sehr gut geeignet, um Untersuchungen mit Kindern durchzuführen.

Eye-Tracking-Methoden

In Eye-Tracking-Experimenten werden Blickbewegungen auf Objekte oder Wörter während der laufenden Sprachverarbeitung gemessen. Die Orientierung des Blickes spiegelt den Aufmerksamkeitsfokus von Personen wider. Beim Eye-Tracking sendet eine Kamera Infrarotlicht aus, welches von der Pupille reflektiert wird. Diese Reflektion wird von der Kamera aufgezeichnet. Anhand dieser Aufzeichnung kann nachvollzogen werden, wo die Personen auf dem Bildschirm hinschauten (methodologische Übersicht siehe Duchowski, 2017).

Zur Erforschung von Sprachverarbeitungsmechanismen wird die Eye-Tracking-Methode in verschiedenen Paradigmen verwendet (Conklin, Pellicer-Sánchez & Carrol, 2018). In Eye-Tracking-Lesestudien geben Sakkaden (schnelle, ruckartige Augenbewegungen) und Fixationen auf Wörtern Aufschluss über Verarbeitungsschwierigkeiten. Weiterhin wird Eye-Tracking im sogenannten *Visual-World-Paradigma* verwendet (Abbildung 4). In diesem Paradigma sehen Versuchspersonen visuelle Stimuli (z.B. Szenen, Objekte oder geschriebene Wörter) auf einem Bildschirm. Gleichzeitig hören sie linguistische Stimuli (z.B. Wörter oder Sätze). Es wird gemessen, inwiefern die Präsentation eines auditiven Stimulus beeinflusst, auf welche visuellen Stimuli Versuchspersonen schauen.

Mithilfe des Visual-World-Pradigmas kann man Bottom-Up Verarbeitungsmechanismen während der laufenden Sprachverarbeitung messen. In einer Studie von Huettig und Altmann (2005) hörten erwachsene Versuchspersonen englische Sätze wie „*Eventually, the man agreed hesitantly, but then he looked at the piano and appreciated that it was beautiful.*“ (dt. *Schließlich stimmte der Mann zögernd zu, aber dann schaute er auf das Piano und würdigte, dass es wunderschön war.*). Dabei wurde untersucht, auf welche Objekte die Versuchspersonen bei der Präsentation des kritischen Wortes (*piano*, dt. *Piano*) schauten. Versuchspersonen fixierten signifikant häufiger auf das genannte Objekt (*Piano*) und auch auf ein Objekt, das semantisch relatiert zu dem genannten Wort war (*Trompete*), im Vergleich zu unrelatierten Objekten (*Möhre*). Ähnliche semantische Effekte zeigten sich, wenn geschriebene Wörter auf dem Bildschirm präsentiert wurden (vgl. Abbildung 4, Huettig, Rommers & Meyer, 2011;

Primativo, Reilly & Crutch, 2017; Shen, Qu & Li, 2016). Des Weiteren zeigen auch Kinder eine Sensitivität gegenüber semantischen Relationen im Visual-World-Paradigma (Borovsky, Elman & Fernald, 2012; Cortés-Monter, Angulo-Chavira & Arias-Trejo, 2017; Huang & Snedeker, 2011). Beispielsweise fanden Cortés-Monter und Kolleg:innen (2017), dass Kinder im Alter von acht Jahren sowohl häufiger auf geschriebene Wörter oder Abbildungen dieser Wörter schauten, die semantisch related zu dem gesprochenen, kritischen Wort waren, als auf unrelatierte Wörter.

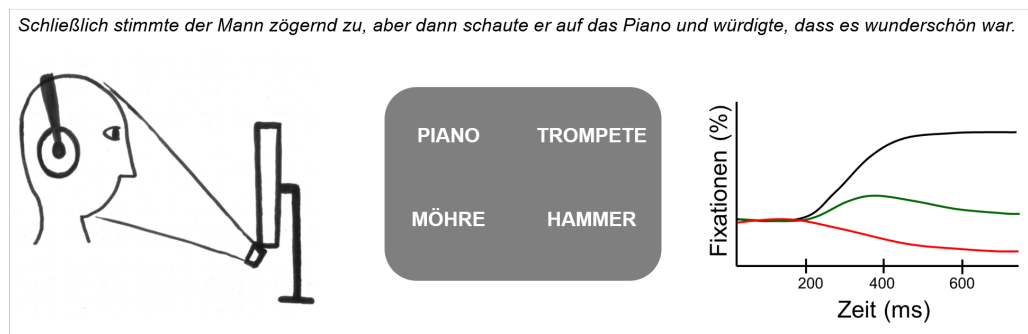


Abbildung 4: Schematische Darstellung des Eye-Tracking im Visual-World-Paradigma anhand eines Beispiels aus Huettig und Altmann (2005). **Links:** Person hört über Kopfhörer einen Satz, während ihre Fixationen auf dem Bildschirm über eine Kamera aufgezeichnet werden. **Mitte:** Geschriebene Wörter auf dem Bildschirm. **Rechts:** Zeitverlauf der Fixationen auf die Wörter auf dem Bildschirm (schwarz = kritisches Wort (*Piano*); grün = semantisch relatierter Kompetitor (*Trompete*); rot = unrelatierter Distraktor (*Möhre*)).

Fixationsunterschiede auf Objekte im Visual-World-Paradigma kennzeichnen Erwartungsprozesse während der laufenden Sprachverarbeitung (Altmann & Kamide, 1999, 2007; Kamide, Altmann & Haywood, 2003). Während erwachsene Versuchspersonen englische Sätze wie „*The boy will eat/move the cake.*“ (dt., *Der Junge wird den Kuchen essen/bewegen.*) hören, schauen sie häufiger auf das Bild eines Kuchens als auf andere nicht-essbare Objekte (Altmann & Kamide, 1999). Erwartungseffekte zeigen sich im Vergleich zwischen der Präsentation der unterschiedlichen Verben (*eat/move*, dt. *essen/bewegen*). Für Sätze, die das Verb „*essen*“ enthalten, treten Fixationsunterschiede bereits nach der Präsentation dieses Verbs und somit vor der Präsentation des kritischen Wortes „*Kuchen*“ auf. Im Gegensatz dazu war der Effekt für Sätze, die das Verb „*bewegen*“ enthielten, erst nach der Präsentation des Wortes „*Kuchen*“ erkennbar. Daher kann man schlussfolgern, dass Menschen Informationen aus dem sprachlichen Input nutzen, um Vorhersagen darüber zu treffen auf welche Objekte referiert wird. Ähnliche Erwartungseffekte zeigen Kinder bereits im Alter von zwei Jahren (z.B. Mani & Huettig, 2012, 2014).

Weiterhin kann man in dem Visual-World-Paradigma messen, ob bei der Erwartung eines Wortes ebenfalls dessen semantische Informationen ko-aktiviert werden (Ito & Husband, 2017; Rommers, Meyer, Praamstra & Huettig, 2013). Beispielsweise fixierten Versuchspersonen während des Hörens des Satzes „*In 1969 Neil Armstrong was the first man to set foot on the*

moon.“ (dt. *Neil Armstrong war 1969 der erste Mensch, der den Mond betrat.*) häufiger das Bild einer Tomate, deren Objekt-Form relatiert zu der eines Mondes ist, als unrelatierte Objekte noch bevor das Wort „*Mond*“ präsentiert wurde (Rommers, Meyer et al., 2013).

Zusammenfassend kann man schlussfolgern, dass Fixationen im Visual-World-Paradigma sensitiv gegenüber der Aktivierung tatsächlich präsentierter und erwarteter Wörter und daraus resultierender semantischen Ko-Aktivierung ist. Ein Vorteil des Visual-World-Paradigma ist es, dass der zeitliche Verlauf der Erwartung während der laufenden Sprachverarbeitung gemessen werden kann. Nachteilig hingegen ist, dass man auf die Präsentation auditiver, linguistischer Stimuli und eine Auswahl an visuellen Stimuli beschränkt ist (vgl. Kamide, 2008).

Ereigniskorrelierte Potentiale

Eine weitere Möglichkeit, implizite Sprachverarbeitung zu untersuchen ist die Messung von Hirnpotentialen mittels Elektroenzephalographie (EEG) (Luck, 2014). Ein EEG wird über Elektroden, die an der Kopfoberfläche angebracht sind, aufgezeichnet. Diese Elektroden messen summierte post-synaptische Spannungspotentiale. Für die Messung der laufenden Sprachverarbeitung sind insbesondere ereignis-korrelierte Potentiale (EKP) relevant. EKP werden durch die zu einem Ereignis – zum Beispiel die Darbietung eines Stimulus – zeitlich-gebundene Mittelung von EEG-Epochen abgeleitet (Abbildung 5). EKP setzen sich aus verschiedenen Komponenten zusammen. Diese Komponenten sind mit bestimmten Verarbeitungsschritten des Stimulus assoziiert und haben eine charakteristische Polarisierung, Latenz und Lokalisierung. Amplitudenunterschiede in diesen Komponenten indizieren Verarbeitungsunterschiede zwischen systematisch variierenden Stimuli. Mit dieser Methode kann die Aktivierung einzelner Wörter und verknüpfter Informationen gemessen werden.

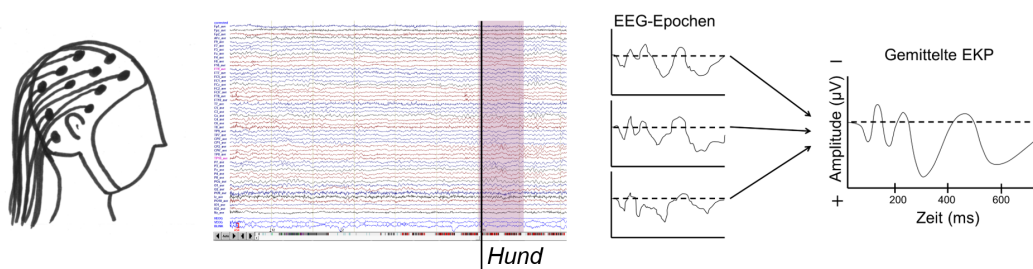


Abbildung 5: Schematische Darstellung der EEG und EKP. **Links:** Elektroden, die an der Kopfoberfläche über eine spezielle Mütze angebracht sind. **Mitte:** Roh-EEG-Daten mit markiertem Beginn eines Stimulus („*Hund*“) und markierter Epoche. **Rechts:** Mehrere EEG-Epochen gemittelt ergeben EKP, in denen unterschiedliche Komponenten erkennbar sind.

Die N400-Komponente ist eine EKP-Komponente, die mit der Aktivierung semantischer Informationen verknüpft ist (Überblick siehe Kutas & Federmeier, 2000, 2011; Lau, Phillips & Poeppel, 2008). Semantisch unpassende Stimuli (z.B. engl. *He spread the warm bread with*

socks., dt. *Er bestrich sein warmes Brot mit Socken.*) weisen eine negativere Amplitude in der N400 auf als semantisch passende Stimuli (z.B. engl. *He spread the warm bread with butter.*, dt. *Er bestrich sein warmes Brot mit Butter.*) (Kutas & Hillyard, 1980). Wie im Folgenden näher beschrieben, bildet die N400 beispielsweise semantische Priming- oder Erwartungseffekte ab. Die semantische N400 beginnt 200 bis 300 ms nach Präsentationsbeginn des kritischen Stimulus und erreicht um 400 ms ihr Maximum. Typischerweise ist sie über posterioren Elektroden lokalisiert. Bei Kindern weist die semantische N400-Komponente im Vergleich zu Erwachsenen eine breitere Verteilung (Friederici & Hahne, 2001; Holcomb, Coffey & Neville, 1992) und stärkere Amplitudenunterschiede auf (z.B. Atchley et al., 2006; Holcomb et al., 1992; Juottonen, Revonsuo & Lang, 1996).

Die N400 ist sensitiv gegenüber semantischem Priming (z.B. Bentin et al., 1985; Brown & Hagoort, 1993; Holcomb & Neville, 1990). Die aufeinanderfolgende Präsentation von semantisch unrelatierten Wortpaaren (z.B. engl. *car – pen*, dt. *Auto - Stift*) löst eine stärkere N400 auf dem letzten Wort aus als die Präsentation semantisch relatierter Wortpaare (z.B. engl. *dog – cat*, dt. *Hund - Katze*). In diesem Fall vereinfacht das Hören oder Lesen eines Wortes die Verarbeitung semantisch relatierter Wörter. EKP-Studien zeigen semantische Priming-Effekte ebenfalls für Kinder spätestens im Alter von 24 Monaten (Rämä, Sirri & Serres, 2013; von Koss Torkildsen, Syversen, Simonsen, Moen & Lindgren, 2007; Wojcik, 2018).

Weiterhin wird der N400-Effekt durch die Erwartung einzelner Wörter innerhalb eines Satzes moduliert. Die N400-Amplitude variiert systematisch mit der Cloze Probability eines Wortes (z.B. Kutas & Hillyard, 1984). Typischerweise ist die N400-Amplitude für Wörter, die nicht erwartet wurden (niedrige Cloze Probability: engl. *He mailed the letter without a thought.*, dt. *Er verschickte den Brief ohne einen Gedanken.*), negativer als für Wörter, die stark erwartet wurden (hohe Cloze Probability: engl. *He mailed the letter without a stamp.*, dt. *Er verschickte den Brief ohne eine Briefmarke.*) (Diaz & Swaab, 2007). Erste Befunde bezüglich Erwartungseffekten auf die N400 bei Kindern, zeigen dass Zehn- bis Zwölfjährige sensitiv gegenüber der Erwartung für satz-finale Wörter sind (Meng, Tian, Jian & Zhou, 2007).

Die N400 ist ebenfalls sensitiv gegenüber semantischen Erwartungen (Abbildung 6). Dies äußert sich in einem N400-Effekt, der durch die semantische Beziehung zwischen einem erwarteten und einem unerwarteten Wort moduliert wird (DeLong, Chan & Kutas, 2018; Federmeier & Kutas, 1999; Ito, Corley, Pickering, Martin & Nieuwland, 2016). Beispielsweise ist das letzte Wort innerhalb des Satzkontextes „*At the dinner party, I wondered why my mother wasn't eating her soup. Then I noticed that she didn't have a spoon.*“ (dt. *Auf der Dinnerparty fragte ich mich, warum meine Mutter ihre Suppe nicht aß. Dann bemerkte ich, dass sie keinen Löffel hatte.*) stark erwartet (Federmeier & Kutas, 1999). Ersetzt man dieses letzte Wort durch ein unerwartetes, aber zu dem erwarteten Wort semantisch relatives Wort (*knife*, dt. *Messer*), löst dies typischerweise eine geringere N400-Amplitude aus als unrelatierte

Wörter (*bowl*, dt. *Schüssel*). Diese semantische Modulation der N400 ist besonders stark in Sätzen, die die Erwartung auf ein bestimmtes Wort sehr stark einschränken (Federmeier & Kutas, 1999; Ito et al., 2016). Diese Modulation ist nicht auf semantische Relationen zwischen den Wörtern innerhalb des Satzkontextes zurückzuführen (Ito et al., 2016). Daher wird davon ausgegangen, dass semantische Informationen während der Erwartung eines Wortes voraktiviert werden. Umstritten ist allerdings, ob die N400 innerhalb neuartiger Sätze die Erwartung eines spezifischen Wortes oder eines Konzeptes widerspiegelt (DeLong, Urbach & Kutas, 2005; Van Petten & Luka, 2012).

[...] *Dann bemerkte ich, dass sie kein/e/n Löffel/Messer/Schüssel hatte.*

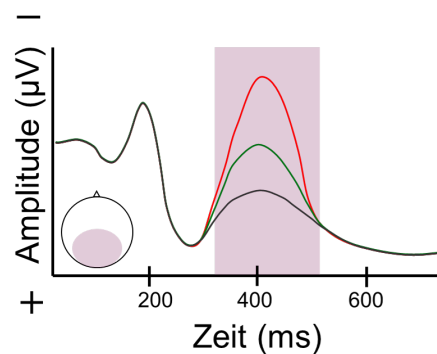


Abbildung 6: Schematische Darstellung der semantischen N400 in einem Erwartungs-Paradigma. Graduelle Abstufung der N400-Amplitude in Abhängigkeit von dem satz-finalen Wort (schwarz = erwartetes Wort; grün = semantisch relatives Wort; rot = unrelatives Wort). Schattierung des N400-Zeitfensters und der Lokalisierung über posteriore Bereiche.

In der Literatur wird weiterhin diskutiert, ob die N400 den Zugriff auf semantische Merkmale im mentalen Lexikon (z.B. Kutas & Federmeier, 2000) oder die Integration semantischer Informationen in den Kontext reflektiert (z.B. Hagoort, Baggio & Willems, 2009). Semantische Erwartungseffekte auf die N400 treten aber auch dann auf, wenn semantisch relatierte Wörter innerhalb des Satzkontextes nicht plausibel sind (z.B. engl. *The student is going to the library to borrow a page.*, dt. *Die Studentin geht zur Bücherei um eine Seite auszuleihen.*, *page/Seite* semantisch relatiert zu dem erwarteten Wort *book/Buch*). Deshalb ist es wahrscheinlich, dass die N400 den Zugriff auf und die Voraktivierung von semantischen Merkmalen reflektiert (Ito et al., 2016). Generell sind N400-Effekte vergleichbar zwischen Priming- und Erwartungsexperimenten, weshalb es unwahrscheinlich ist, dass diese Prozesse sich grundlegend unterscheiden (Kutas, 1993).

Während N400-Studien häufig eine visuelle Präsentation verwenden, tritt für die auditive Präsentation der Stimuli bereits eine frühe Negativierung auf (Connolly & Phillips, 1994). In EKP-Studien ist man für die Messung der visuellen Sprachverarbeitung auf eine

Wort-für-Wort-Präsentation auf einem Bildschirm beschränkt. Diese verursacht die wenigsten Augenbewegungsartefakte im EEG. In solch einem Design ist das jeweilige geschriebene Wort sofort komplett verfügbar und mit einer einzelnen Fixierung erfassbar. Im Gegensatz dazu wird gesprochene Sprache typischerweise inkrementell in EKP-Experimenten präsentiert. In diesem Fall ist das sprachliche Signal zeitlich gedehnt und komplette Wörter sind nicht gleichzeitig im Input verfügbar. In Bezug auf die semantische Erwartung wurden vergleichbare N400-Effekte für auditive im Vergleich zu visueller Präsentation gefunden (Federmeier, McLennan, De Ochoa & Kutas, 2002). Einige Studien berichten, dass die N400 bei der Verarbeitung von gesprochener Sprache zeitlich früher nach Stimulusbeginn auftritt als für geschriebene Sprache (z.B. Connolly & Phillips, 1994; McCallum, Farmer & Pockock, 1984; Van Petten, Coulson, Rubin, Plante & Parks, 1999). Es ist bisher umstritten, ob diese frühe Negativierung lediglich einen frühen Start der N400 während der auditiven Verarbeitung darstellt oder funktional unterschiedlich von der N400 ist (Überblick siehe Nieuwland, 2019).

Einige Autor:innen nehmen an, dass die frühe Negativierung bei der Verarbeitung gesprochener Sprache mit der Erwartung phonologischer Merkmale in Zusammenhang stehen könnte. Somit würde sie eine abgrenzbare, funktional unterschiedliche Komponente zur N400 darstellen (*Phonological Mismatch Negativity*, PMN: Boudewyn, Long & Swaab, 2015; Connolly, D'Arcy, Kujala & Alho, 2001; Connolly & Phillips, 1994; Newman & Connolly, 2009). Eine erste systematische Untersuchung führten Connolly und Phillips (1994) durch. Sie verwendeten auditives Satzmaterial mit einer hohen Cloze Probability des letzten Wortes. Die Sätze wurden mit Wörtern vervollständigt, die sich systematisch in phonologischer bzw. semantischer Relation zu dem erwarteten Wort unterschieden. Die Wörter waren entweder phonologisch relatiert am Wortbeginn aber nicht semantisch zu dem erwarteten Wort relatiert (engl. *icicles – eyes*, dt. *Eiszapfen – Augen*), semantisch aber nicht phonologisch zu dem erwarteten Wort relatiert (engl. *kitchen – sink*, dt. *Küche – Spülbecken*) oder weder phonologisch noch semantisch relatiert (engl. *nose – milk*, dt. *Nase – Milch*). Für alle Bedingungen mit semantischer Erwartungsverletzung wurden vergleichbare N400-Effekte gefunden. Für alle Bedingungen mit phonologischer Erwartungsverletzung hingegen wurde eine erhöhte, frühe Negativierung gefunden. Für geschriebene Sprache konnten die Autor:innen diese Ergebnisse in einer Studie mit dem gleichen Material und Design nicht replizieren (Connolly, Phillips & Forbes, 1995). Deshalb postulieren sie, dass dieser Effekt spezifisch für die Verarbeitung gesprochener Sprache ist und auftritt, wenn das präsentierte Wort phonologisch nicht mit dem erwarteten Wort übereinstimmt.

Eine alternative Interpretation für die frühe Negativierung bieten Hagoort und Kollegen. Sie argumentieren, dass der frühe Start der N400 bei der Verarbeitung gesprochener Sprache nicht nur den Abgleich der phonologischen Form, sondern die lexikalische Auswahl widerspiegelt, weil dieser Effekt abhängig vom Kontext sei (Hagoort & Brown, 2000; Van Den Brink, Brown & Hagoort, 2001). Im Gegensatz dazu besteht auch die Möglichkeit, dass diese frühe Negativierung sich nicht funktional von der N400 unterscheidet, sondern lediglich einen frühen

Beginn der N400-Komponente während der Verarbeitung gesprochener Sprache darstellt (Diaz & Swaab, 2007; Van Petten et al., 1999). Unabhängig der Interpretation dieses Effektes, zeigen die beschriebenen Studien, dass Amplitudenunterschiede zwischen erwarteten und unerwarteten Wörtern zeitlich früher für die Verarbeitung gesprochener als geschriebener Sprache auftreten können.

Aktuellere Forschung bezüglich semantischer Erwartungseffekte fokussiert sich weiterhin auf eine späte Positivierung, die zeitlich nach der N400 auftritt (*post-N400 positivity*, PNP: DeLong, Quante & Kutas, 2014; Kuperberg & Wlotko, 2020; Thornhill & Van Petten, 2012; Van Petten & Luka, 2012). Nach bisherigem Forschungsstand liegt die Vermutung nahe, dass zwei unterschiedliche Typen dieser späten Positivierung auftreten können, die sich in ihrer Lokalisierung und Funktion unterscheiden (Überblick siehe Van Petten & Luka, 2012). Die anteriore PNP spiegelt Effekte der lexikalischen Vorhersage wider. Diese Komponente ist unabhängig von der semantischen Beziehung zwischen dem unerwarteten und dem erwarteten Wort. Außerdem tritt sie für plausible als auch nicht plausible Interpretationen auf. Dies deutet darauf hin, dass die anteriore PNP eine Erwartung der lexikalischen Form reflektiert und jede Verletzung dieser Erwartung in größeren Amplitudenunterschieden erkennbar ist. Der zweite Typ der PNP, der eine posteriore Lokalisierung hat, steht in Zusammenhang mit der Plausibilität der Erwartungsverletzungen. Daher reflektiert sie Schwierigkeiten, unerwartete, unplausible Satzendungen in den bisherigen Satzkontext zu integrieren. Beide Typen der PNP werden nicht direkt durch die semantische Beziehung zwischen erwartetem Wort und wahrgenommenem Wort moduliert.

Bisher liegen wenige EKP-Studien vor, die semantische Erwartungen in Kindern gemessen haben. Kessler und Friedrich (eingereicht) testeten zehnjährige Kinder in einem semantischen Erwartungsparadigma mit gesprochener Sprache (vgl. Federmeier et al., 2002). Die Kinder hörten Satzanfänge wie *“Im Urlaub bauten sie eine Burg aus...“*. Diese wurden entweder mit einem erwarteten Wort (*Sand*), einem semantisch relatierten Wort (*Felsen*) oder einem unrelatierten Wort (*Kuchen*) vervollständigt. Kinder zeigten N400-Effekte, die mit der semantischen Erwartung systematisch variierten. Vergleichbar zu der frühen Negativierung, die typisch für die Verarbeitung gesprochener Sprache ist (Überblick siehe Nieuwland, 2019), zeigten auch Kinder Amplitudenunterschiede zwischen 100 und 300 ms nach Beginn der satz-finalen Wörter. In diesem Zeitfenster zeigten sich ebenfalls Amplitudenunterschiede zwischen relatierten und unrelatierten Wörtern. Daher bietet dieser Effekt Evidenz für eine Voraktivierung der Bedeutung des erwarteten Wortes. Weiterhin fanden die Autor:innen eine späte Positivierung, deren Amplituden sich zwischen erwarteten und unrelatierten und zwischen relatierten und unrelatierten Vervollständigungen unterschieden. Dieser Effekt könnte daher resultieren, dass relatiertere Vervollständigungen einfacher in den Satzkontext integrierbar sind. Insgesamt deuten die Ergebnisse dieser Studie darauf hin, dass Kinder vergleichbar zu Erwachsenen schnelle semantische Erwartungen aufbauen können.

Zusammenfassend kann anhand von Offline- und Online-Methoden sowohl die Aktivierung sprachlicher Einheiten als auch die Ko-Aktivierung der damit verknüpften semantischen Informationen gemessen werden. Nichtsdestotrotz gibt es Hinweise darauf, dass beispielsweise EKP-Methoden eine automatische semantische Aktivierung detektieren, obwohl diese in Reaktionszeit-Daten nicht messbar ist (Heil, Rolke & Pecchinenda, 2004). Insgesamt bieten Online-Methoden daher den Vorteil, dass sie sensitiver gegenüber einer automatischen semantischen Aktivierung und den zeitlichen Verlauf dieser Aktivierung abbilden können.

Verarbeitungsmechanismen können indirekt Rückschlüsse auf die Repräsentation sprachlicher Einheiten im mentalen Lexikon sowohl bei erwachsenen Muttersprachler:innen als auch bei sprachlernenden Kindern gezogen werden. Bisher wurde beschrieben, wie diese Methoden verwendet wurden, um die Verarbeitung und Repräsentation einzelner Wörter innerhalb wörtlicher, nicht formelhafter Ausdrücke zu überprüfen. Die gleichen Methoden können angewendet werden, um die mentale Repräsentation von MWE oder die Verarbeitung einzelner Bestandteile innerhalb dieser zu untersuchen. Daher sind diese Methoden relevant, um zu untersuchen, ob MWE eine Mehrwort-Repräsentation besitzen und ob einzelne Wörter mit dieser Repräsentation verknüpft sind.

3 Verarbeitung und Repräsentation von Mehrworteinheiten im Erwachsenenalter

sich in die Wolle kriegen

Die in Kapitel 2 beschriebenen Methoden können verwendet werden, um zu untersuchen, wie MWE verarbeitet werden. Dadurch können indirekte Hinweise auf deren mentale Repräsentation gezogen werden. Diese Methoden können angewendet werden, um zu messen, wie schnell MWE verarbeitet werden und ob einzelne Bestandteile innerhalb dieser erwartet werden. Solche Befunde bieten Hinweise darauf, ob MWE eine Mehrwort-Repräsentation im mentalen Lexikon besitzen. Außerdem kann untersucht werden, wie veränderte Varianten von Idiomen verarbeitet werden. Anhand von solchen Befunden kann geschlussfolgert werden, wie flexibel die Mehrwort-Repräsentation ist, beziehungsweise, ob auf einzelne Bestandteile dieser zugegriffen werden kann. Im Folgenden werden empirische Befunde zur Verarbeitung von MWE für erwachsenen Muttersprachler:innen beschrieben.

3.1 Verarbeitungsvorteile für Mehrworteinheiten

Verarbeitungsvorteile von MWE gegenüber neuartigen Ausdrücken mit vergleichbarer Länge, die aber keinen formelhaften Charakter aufweisen, zeigen sich über verschiedene Messmethoden hinweg (Überblick siehe Conklin & Schmitt, 2012; Siyanova-Chanturia & Martinez, 2015). In Offline-Methoden zeigte sich der Verarbeitungsvorteil in schnelleren Reaktionen, mit denen Menschen Urteile über entsprechende Ausdrücke treffen (Arnon & Snider, 2010; Jiang & Nekrasova, 2007; Tabossi et al., 2009), und besseren Erinnerungsleistung (Tremblay & Baayen, 2010; Tremblay et al., 2011). Weiterhin manifestierte sich der Verarbeitungsvorteil in einer kürzeren Produktionsdauer von MWE gegenüber nicht MWE (Arnon & Cohen Priva, 2013; Siyanova-Chanturia & Lin, 2018).

Der Verarbeitungsvorteil von MWE gegenüber Nicht-MWE zeigt sich ebenfalls in Online-Lesestudien: Versuchspersonen lesen MWE schneller als Nicht-MWE (z.B. Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin & Schmitt, 2011; Siyanova-Chanturia, Conklin & van Heuven, 2011; Tremblay et al., 2011; Underwood, Schmitt & Galpin, 2004). Underwood und Kolleg:innen (2004) führten eine Eye-Tracking-Lesestudie durch, bei der sie die Fixationen während des Lesens verschiedener Sätze aufzeichneten. In dieser Studie fixierten Versuchspersonen Wörtern in finaler Position weniger häufig, wenn diese in formelhaften Ausdrücken auftraten

(engl. *put all your eggs in one basket*, dt. *alle Eier in einen Korb legen*), als in nicht-formelhaften Ausdrücken (engl. [...] *who dropped my flower basket*, dt. [...] *der meinen Blumenkorb fallen ließ*). Schnelleres Lesen wurde ebenfalls in *Self-Paced-Reading*-Studien gefunden (Conklin & Schmitt, 2008; Tremblay et al., 2011). In *Self-Paced-Reading*-Studien lesen Versuchspersonen Sätze bzw. Satzfragmente auf einem Bildschirm und entscheiden selbstständig, wie schnell sie weiterlesen. In diesem Paradigma wurden Verarbeitungsvorteile von MWE gegenüber Kontrollausdrücken allerdings nur für größere sprachliche Präsentationseinheiten (Satz-für-Satz: engl. *I sat in the middle of the bullet train.*, dt. *Ich saß in der Mitte des Hochgeschwindigkeitszuges.*; Phrase-für-Phrase: engl. *I sat – in the middle of the – bullet train.*), nicht aber für eine Wort-für-Wort-Präsentation (engl. *I – sat – in – the – middle – of – the – bullet – train.*) gefunden (Conklin & Schmitt, 2008; Schmitt & Underwood, 2004; Tremblay et al., 2011). Da der Fokus während der Wort-für-Wort-Präsentation auf jedem einzelnen Wort liegt, könnte der Verarbeitungsvorteil erst für größere, dargestellte Einheiten erkennbar sein. Außerdem verdeutlichen diese Unterschiede, dass die Präsentationsart für die Messung des Verarbeitungsvorteils ausschlaggebend sein kann.

Über behaviorale Daten hinaus spiegelt sich der Verarbeitungsvorteil von MWE gegenüber Nicht-MWE auch auf neurophysiologischer Ebene wider (la Roi, Sprenger & Hendriks, 2019; Laurent, Denhieres, Passerieux, Iakimova & Hardy-Bayle, 2006; Rommers, Dijkstra & Bastiaansen, 2013; Siyanova-Chanturia, Conklin, Caffarra, Kaan & van Heuven, 2017; Strandburg et al., 1993). Typischerweise evozieren Idiome im Vergleich zu wörtlichen oder sinnlosen Ausdrücken eine reduzierte N400-Komponente (Strandburg et al., 1993). Diese Ergebnisse verdeutlichen, dass idiomatische Ausdrücke einfacher verarbeitet werden können als wörtliche, neuartige Ausdrücke.

Die Befunde des Verarbeitungsvorteils von MWE gegenüber Nicht-MWE legen nahe, dass MWE schnell aus dem Gedächtnis abgerufen werden können und einen besonderen Status im mentalen Lexikon besitzen. Dennoch können unterschiedliche Begründungen dafür gefunden werden.

Eine mögliche Erklärung für den Verarbeitungsvorteil von MWE ist, dass sie als holistische Einheit im mentalen Lexikon abgespeichert sind und als solche direkt abgerufen werden können (z.B. Swinney & Cutler, 1979). Entsprechend dieser Interpretation haben MWE einen eigenen Eintrag im mentalen Lexikon und können vergleichbar zu „langen Wörtern“ abgerufen werden.

Alternativ könnte der Verarbeitungsvorteil auch daher resultieren, dass die jeweiligen Ausdrücke häufig im Sprachgebrauch auftreten (z.B. Arnon & Snider, 2010; Siyanova-Chanturia, Conklin & van Heuven, 2011; Tremblay et al., 2011). In der Tat verstärkt sich der Verarbeitungsvorteil von MWE gegenüber neuartigen Ausdrücken mit steigender Phrasenhäufigkeit in Korpora (Arnon & Cohen Priva, 2013; Arnon & Snider, 2010). Daraus lässt sich schließen: Je

häufiger ein bestimmter Ausdruck auftritt, desto stärker ist dessen Repräsentation im mentalen Lexikon, desto schneller kann dieser Ausdruck verarbeitet werden.

Für den Verarbeitungsvorteil von nicht-kompositionellen MWE wie Idiomen, die im Einzelnen nicht besonders häufig im Sprachgebrauch auftreten (Moon, 1988), ist weiterhin entscheidend, wie bekannt deren konventionalisierte Wort-Konfiguration ist (Carrol & Littlemore, 2020; Tabossi et al., 2009). Für die Verarbeitungsgeschwindigkeit sehr bekannter Idiome scheinen semantische Merkmale von Idiomen wie Zerlegbarkeit und wörtliche Plausibilität nicht ausschlaggebend zu sein (Carrol & Conklin, 2020; Carrol & Littlemore, 2020; Libben & Titone, 2008; Tabossi et al., 2009; Titone & Connine, 1999; Titone et al., 2019), obwohl diese die figurative Interpretation der Ausdrücke beeinflussen können (Findlay & Carrol, 2018; Titone & Libben, 2014). Ebenso ist der Verarbeitungsvorteil bekannter Idiome unabhängig von dem jeweiligen Kontext, in dem sie auftreten (Carrol & Littlemore, 2020; Conklin & Schmitt, 2008; Geeraert, Baayen & Newman, 2018; Siyanova-Chanturia, Conklin & Schmitt, 2011). Die Einbettung der Idiome in einen kontextuellen Rahmen kann deren Interpretation modulieren, wodurch die figurative Bedeutung früher erkannt wird (Fanari, Cacciari & Tabossi, 2010; Holsinger, 2013). Dennoch werden Idiome in ihrer konventionalisierten Wort-Konfiguration sowohl innerhalb von figurativen (z.B. engl. *Although these were new stocks, they suddenly went through the roof.*, dt. *Obwohl dies neue Aktien waren, gingen sie plötzlich durch die Decke.*, Bedeutung: die Aktien stiegen stark an) als auch innerhalb von wörtlichen Kontexten (z.B. engl. *While the guys were reshingling, they suddenly went through the roof.*, dt. *Während die Jungs das Dach neu deckten, gingen sie plötzlich durch die Decke.*) schneller gelesen als Kontrollsätze (Conklin & Schmitt, 2008; Geeraert et al., 2018; Siyanova-Chanturia, Conklin & Schmitt, 2011). Dies lässt darauf schließen, dass die Bekanntheit der konventionalisierten Wort-Konfiguration von Idiomen für den Verarbeitungsvorteil ausschlaggebend ist.

Aufgrund der Bekanntheit mit der konventionalisierten Wort-Konfiguration einer MWE steigt ebenfalls die Wahrscheinlichkeit, einzelne Bestandteile dieser korrekt vorherzusagen (Tabossi, Fanari & Wolf, 2005; Titone & Connine, 1994). Beispielsweise kann das Idiom „jemand ließ die Katze aus dem Sack“ aufgrund seiner Bekanntheit schon vor dem letzten Wort erkannt werden, weshalb für dieses eine starke Erwartung aufgebaut werden kann. Innerhalb von Cloze-Probability-Aufgaben ergänzten Versuchspersonen idiomatische Ausdrücke mit größerer Wahrscheinlichkeit mit dem fehlenden, idiomatisch korrekten Bestandteil als mit einem anderen Wort (Van Lancker Sidtis, Cameron, Bridges & Sidtis, 2015). Lesestudien belegen sowohl, dass Versuchspersonen häufig das letzte Wort innerhalb einer bekannten MWE überspringen (Carrol & Conklin, 2020; Underwood et al., 2004), als auch, dass nachdem formelhafte Ausdrücke als solche erkannt wurden, diese schneller gelesen werden (Cacciari & Corradini, 2015; Vespignani, Canal, Molinaro, Fonda & Cacciari, 2010). Daher kann der Verarbeitungsvorteil von MWE auch dadurch begründet werden, dass die letzten Wörter innerhalb bekannter Wort-Konfigurationen beim Lesen übersprungen werden können, weil diese bereits erwartet wurden.

Weitere Evidenz für die Vorhersage finaler Wörter innerhalb bekannter MWE stammt aus EKP-Studien (Molinaro & Carreiras, 2010; Rommers, Dijkstra & Bastiaansen, 2013; Vespignani et al., 2010). Rommers, Dijkstra und Bastiaansen (2013) verglichen die Verarbeitung von bekannten Idiomen, deren letztes Wort die idiomatisch korrekte Vervollständigung war und eine hohe Cloze Probability aufwies (ndrl. [...] *liep tegen de lamp*, dt. [...] *lief gegen die Lampe*, Bedeutung: gefasst werden), mit Varianten, in denen sie das letzte Wort durch ein Wort ersetzten, das nicht erwartet wurde (ndrl. [...] *liep tegen de vis*, dt. [...] *lief gegen den Fisch*). Während der Verarbeitung des letzten Wortes fanden sie reduzierte N400-Amplituden für idiomatisch korrekte Vervollständigungen im Vergleich zu inkorrekten, unerwarteten Vervollständigungen. Da die N400-Effekte der Cloze Probability reflektiert (siehe Kapitel 2.2), bestätigen diese Ergebnisse die Erwartung für spezifische Wörter innerhalb einer bekannten MWE auf neurophysiologischer Ebene. Zusätzlich fanden die Autor:innen eine späte Positivierung (vergleichbar zur PNP-Komponente, siehe Kapitel 2.2), die ebenfalls stärker für inkorrekte als für korrekte Vervollständigungen der Idiome war. Gemäß den Autor:innen ist dies ein Indiz für eine vereinheitlichte Repräsentation der Idiom-Konfiguration im mentalen Lexikon. Vespignani und Kolleg:innen (2010) fanden zusätzlich zu Unterschieden in der N400- auch Unterschiede in der P300-Komponente zwischen korrekten und inkorrekten Idiom-Vervollständigungen. Die Autor:innen postulieren, dass nach dem Erkennen des Idioms dessen Konfiguration aktiviert wird und jede Abweichung von dieser Konfiguration in einem P300-Effekt resultiert. Insgesamt deuten die empirischen Befunde darauf hin, dass die Bekanntheit mit den jeweiligen Ausdrücken die Erwartung für spezifische Wörter innerhalb dieses Ausdrucks steigert.

Unabhängig der zugrundeliegenden Erklärungsansätze bieten die beschriebenen empirischen Befunde eine Grundlage für die Annahme, dass die konventionalisierte Wort-Konfiguration von MWE in irgendeiner Art und Weise Spuren im Gedächtnis hinterlässt, weshalb sie schnell verarbeitet werden können. Aufgrund von konsistenten Befunden zu Verarbeitungsvorteilen von MWE gegenüber neuartigen Ausdrücken, scheint der besondere Status von MWE relativ gesichert, obwohl die genauen Gründe dieses Verarbeitungsvorteils nicht ausreichend ergründet sind. Daher ist unklar, ob sie als eine holistische Einheit abgespeichert sind oder generell aufgrund ihrer Bekanntheit und Häufigkeit einen besonderen Status im mentalen Lexikon besitzen.

3.2 Flexibilität der Wort-Konfiguration innerhalb von Mehrworteinheiten

Für die zugrundeliegende Form der Repräsentation von MWE im mentalen Lexikon ist außerdem entscheidend, wie flexibel die Wort-Konfiguration ist. Zu diesem Zweck untersuchten einige Studien, wie modifizierte Varianten der konventionalisierten Konfiguration verarbeitet werden (z.B. Geeraert et al., 2018; Holsinger, 2013; Kyriacou, Conklin & Thompson, 2020). Modifikationen beinhalten beispielsweise syntaktische (z.B. Passivierung: *Die Katze wurde aus dem Sack*

gelassen), lexikalische (z.B. Ersetzen eines Wortes durch ein nahes Synonym: *Die Katze aus dem Sack freilassen*), formale (z.B. Adjektiv Einschub: *Die große Katze aus dem Sack lassen*) oder morpholexikalische (Diminutiv: *Das Kätzchen aus dem Sack lassen*) Modifikationen. Korpusdaten legen nahe, dass Modifikationen von Idiomen im tatsächlichen Sprachgebrauch auftreten (z.B. Langlotz, 2006). Außerdem verdeutlichen empirische Befunde, dass selbst modifizierte Idiome oder Kollokationen ihre figurative Bedeutung nicht komplett verlieren (Holsinger, 2013; Mancuso et al., 2020; Molinaro, Canal, Vespignani, Pesciarelli & Cacciari, 2013; Smolka & Eulitz, 2020). Einige syntaktische oder lexikalische Modifikationen von MWE wurden von Muttersprachler:innen auch als akzeptable Ausdrücke bewertet (Geeraert, Baayen & Newman, 2017; Tabossi et al., 2009). Außerdem wurden modifizierte Varianten schneller gelesen als Kontrollsätze (Geeraert et al., 2018; Kyriacou et al., 2020). Die vorliegenden Befunde belegen somit auf verschiedenen Ebenen, dass die Repräsentation von MWE kein strikter lexikalischer Eintrag im mentalen Lexikon ist, sondern die Mehrwort-Repräsentation durchaus veränderbar ist.

Darüber hinaus gibt es direkte Belege dafür, dass eine syntaktische Analyse von MWE durchgeführt wird (Konopka & Bock, 2009; Peterson, Burgess, Dell & Eberhard, 2001; Snider & Arnon, 2012). Beispielsweise bahnte die Verwendung von Dativobjekten innerhalb von Idiomen (z.B. engl. *The lifeguard gave the surfer the creeps.*, dt. *Der Rettungsschwimmer machte dem Surfer eine Gänsehaut.*) die Produktion entsprechender, syntaktischer Strukturen ähnlich stark wie nicht-idiomatische Ausdrücke (z.B. engl. *The mother gave the hungry baby some food.*, dt. *Die Mutter gab dem hungrigen Säugling etwas Nahrung.*) (Snider & Arnon, 2012). Dies deutet darauf hin, dass die syntaktische Struktur tatsächlich analysiert wird und verändert werden kann.

3.3 Verarbeitung der Einzel-Bestandteile innerhalb von Mehrworteinheiten

Falls MWE einen besonderen Status im mentalen Lexikon aufweisen, ist ebenfalls bedeutsam, inwiefern diese Mehrwort-Repräsentationen in die einzelnen Wörter zerlegt werden (vgl. Kapitel 1.1). Dies kann beispielsweise anhand der Verarbeitung der Einzelwort-Häufigkeit oder der Einzelwort-Bedeutung gemessen werden.

Falls sich Häufigkeitseffekte von MWE nicht nur auf Phrasen- sondern auch auf Einzelwort-Ebene zeigen, würde dies eine Aktivierung der Einzel-Bestandteile innerhalb dieser Wort-Konfigurationen implizieren. In der Tat sind Versuchspersonen sensitiv gegenüber Einzelwort-Häufigkeiten innerhalb von MWE in Produktionsdauer (Arnon & Cohen Priva, 2014) und Erinnerungsleistung (Tremblay & Baayen, 2010). Beispielsweise konnten Arnon und Cohen Priva (2014) zeigen, dass die Einzelwort-Häufigkeit auch bei sehr häufigen Phrasen in Zusammenhang mit der Produktionsdauer steht.

Um die Aktivierung der Einzelwort-Bedeutung innerhalb von MWE zu untersuchen, sind Idiome besonders gut geeignet. Da die Phrasen-Bedeutung von Idiomen sich nicht aus der Interpretation der einzelnen Wörter ableiten lässt (siehe Kapitel 1), ist eine Aktivierung und Interpretation der Einzelwort-Bedeutungen nicht notwendig, nachdem Idiome als solche erkannt wurden. Falls dennoch die Einzelwort-Bedeutungen aktiviert werden, könnte dies in Verarbeitungsunterschiede zwischen Wörtern, die semantisch relatiert (*Korb*), und Wörtern, die semantisch unrelatiert (*Bauch*) zu einem Idiom-Bestandteil (*Sack* in *die Katze aus dem Sack lassen*) sind, gemessen werden. Studien, die diesen Ansatz generell verfolgten, gelangten allerdings zu unterschiedlichen Ergebnissen. Die vorliegende Dissertation fokussiert sich auf die Aktivierung der Einzelwort-Bedeutungen und die bisherigen, abweichenden Ergebnisse. Daher erfolgt eine detaillierte Beschreibung der empirischen Untersuchungen – insbesondere hinsichtlich der verwendeten Paradigmen und daraus resultierenden Verarbeitungsmechanismen, der Modalität und der Vorhersagbarkeit der idiom-finalen Wörter.

Die Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomen kann anhand der Bottom-Up Verarbeitung eines Idiom-Bestandteiles gemessen werden (vgl. Kapitel 2, siehe Abbildung 7). Bei der Messung der Ko-Aktivierung semantisch relatierter Wörter wird der kritische Idiom-Bestandteil direkt innerhalb des Idioms präsentiert. Studien, die diesen Ansatz verfolgten, finden konsistent positive Befunde für die Aktivierung der Einzelwort-Bedeutung (Beck & Weber, 2016a; Holsinger, 2013; Rabanus, Smolka, Streb & Rösler, 2008; Smolka et al., 2007; van Ginkel & Dijkstra, 2019).

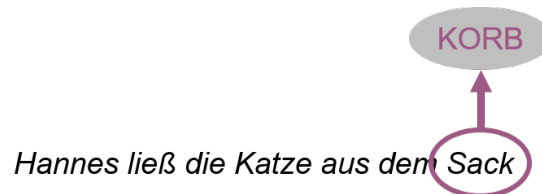


Abbildung 7: Schematische Darstellung für die Bottom-Up Aktivierung des Wortes „Korb“ nach der Präsentation des Wortes „Sack“ innerhalb des Idioms „jemand ließ die Katze aus dem Sack“.

Beck und Weber (2016a) führten eine *cross*-modale Satzpriming-Studie durch und fanden Evidenz für eine Aktivierung der Einzelwort-Bedeutung. In dieser Studie hörten die Versuchspersonen Idiome innerhalb eines minimalen, neutralen Kontextes (z.B. engl. *His uncle kicked the bucket.*, dt. *Sein Onkel trat den Eimer.*, Bedeutung: Sein Onkel starb) und führten eine lexikalische Entscheidungsaufgabe für ein darauffolgendes, geschriebenes Zielwort durch. Versuchspersonen reagierten signifikant schneller auf Zielwörter, die semantisch relatiert zu dem idiom-finalen Wort waren (*pail* Synonym zu *bucket*, dt. *Eimer*), als auf unrelatierte Zielwörter (*boat*, dt. *Boot*). Vergleichbare Ergebnisse fanden van Ginkel und Dijkstra (2019) in einer uni-modalen Satzpriming-Studie mit geschriebenen Idiomen und Zielwörtern.

Unter der Verwendung von Online-Messmethoden wurde ebenfalls eine Aktivierung der Einzelwort-Bedeutung während der direkten Verarbeitung des Idiom-Bestandteiles gefunden. Holsinger (2013) führte ein Visual-World-Eye-Tracking-Experiment (siehe Kapitel 2.2) durch, in dem Versuchspersonen neutrale Sätze hörten, die Idiome enthielten (z.B. engl. *John kicked the bucket last Thursday*, dt. *John trat den Eimer letzten Donnerstag*). Bedeutung: John starb letzten Donnerstag). Währenddessen sahen sie vier geschriebene Wörter auf dem Bildschirm, die entweder semantisch relativiert zu einem Idiom-Bestandteil (*foot* relativiert zu *kick*; dt. *Fuß* relativiert zu *treten*), zu der figurativen Bedeutung assoziiert (*death*, dt. *Tod*) oder unrelativiert (*triangle*, *animal*, dt. *Dreieck*, *Tier*) waren. Nachdem die Versuchspersonen die jeweiligen Idiom-Bestandteile hörten, fixierten sie häufiger auf semantisch relativierte Wörter als auf unrelativierte Wörter.

Weiterhin fanden Smolka und Kolleg:innen (2007) und Rabanus und Kolleg:innen (2008), dass die Einzelwort-Bedeutung ebenfalls von stark erwarteten, idiom-finalen Wörtern aktiviert wird (*Mir ist ein Stein vom Herzen gefallen*). Sie führten ein uni-modales Satzpriming-Experiment durch, in dem Versuchspersonen Sätze Wort-für-Wort und anschließend ein Zielwort lasen. Es zeigten sich kürzere Reaktionszeiten auf Zielwörter, die semantisch zum dem idiom-finalen Verb relativiert (*Leiter* relativiert zu *gefallen*) waren, im Vergleich zu unrelativierten Zielwörtern (*Indien*). Stark vorhersagbare Idiome können bereits vor deren Offset erkannt werden (Cacciari & Corradini, 2015). Einige Autor:innen nehmen deshalb an, dass nach dem Erkennen des Idioms die Aktivierung der Einzelwort-Bedeutung weniger wahrscheinlich ist (Cacciari & Tabossi, 1988). Im Gegensatz dazu zeigt die Studie von Smolka und Kolleg:innen (2007), dass selbst nach dem Erkennen des Idioms Einzelwort-Bedeutungen aktiviert werden können.

Aufgrund der starken Vorhersagbarkeit einiger Idiome (Diskussion siehe Kapitel 3.1), kann die Aktivierung der Einzelwort-Bedeutung auch anhand der Top-Down Erwartung des letzten Wortes gemessen werden (vgl. Kapitel 2, siehe Abbildung 8). Bei der Messung der Ko-Aktivierung semantisch relatierter Wörter wird der jeweilige Idiom-Bestandteil nicht direkt innerhalb des Idioms präsentiert. Studien, die den kritischen Idiom-Bestandteil nicht präsentierten, finden widersprüchliche Ergebnisse hinsichtlich der Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomen (Rommers, Dijkstra & Bastiaansen, 2013; Sprenger et al., 2006).

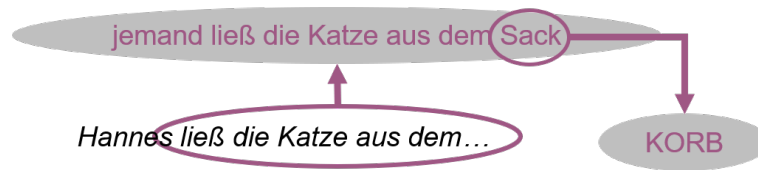


Abbildung 8: Schematische Darstellung für die Top-Down Aktivierung des Wortes „Korb“ nach der Erwartung des Wortes „Sack“ innerhalb des Idioms „jemand ließ die Katze aus dem Sack“.

Sprenger und Kolleg:innen (2006) führten eine Produktionsstudie durch und fanden positive Evidenz für die Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomen. In dieser Studie untersuchten die Autor:innen Priming-Effekte zwischen bekannten Idiomfragmenten (ndrl. *Jan liep tegen de [lamp]*, dt. *Jan lief gegen die [Lampe]*, Bedeutung: Jan wurde gefasst) und Wörtern, die semantisch related zu den idiom-finalen Wörtern waren (*kaars* related zu *lamp*, dt. *Kerze* related zu *Lampe*), untersuchten. In diesen Experimenten wurden idiom-finale Wörter nicht direkt präsentiert. In einem Experiment lasen Versuchspersonen Idiomfragmente und sollten das fehlende idiom-finale Wort so schnell wie möglich verbal produzieren. Währenddessen hörten sie ein Primewort, das zu dem idiom-finalen Wort entweder semantisch related oder unrelated war. Die Zeit zwischen dem Präsentationsstart dieses Primewortes und dem Präsentationsstart des Idiomfragments (*Stimulus Onset Asynchrony*, kurz SOA) betrug entweder -150 ms, 0 ms, 100 ms oder 200 ms. Versuchspersonen produzierten Idiom-Bestandteile schneller nach der Präsentation eines semantisch relateden Primes als nach der Präsentation eines unrelateden Primes. Dieser Priming-Effekt war am stärksten für kurze SOAs. Dieser Befund deutet darauf hin, dass die Aktivierung der Einzelwort-Bedeutung nur kurzzeitig auftritt.

Ähnliche Ergebnisse einer frühen semantischen Aktivierung konnten Sprenger und Kolleg:innen (2006) in einem anderen Experiment mit dem gleichen Stimulusmaterial finden. In diesem Experiment untersuchten sie Effekte der Planung des idiom-finalen Wortes auf die Produktion semantisch relateder Wörter. Versuchspersonen lasen Idiomfragmente (ndrl. *Jan liep tegen de*, dt. *Jan lief gegen die*) und sollten das letzte Wort verbal produzieren (*lamp*, dt. *Lampe*), sobald ein Fragezeichen auf dem Bildschirm erschien. In der Hälfte der Durchgänge wurde anstelle des Fragezeichens ein Zielwort präsentiert, welches die Versuchspersonen anstelle des idiom-finalen Wortes produzieren sollten. Diese Zielwörter wurden ebenfalls zu variierenden SOAs (100 ms, 200 ms, 300 ms, 400 ms) relativ zu dem Idiomfragment präsentiert. Versuchspersonen produzierten semantisch relatede Zielwörter (*kaars*, dt. *Kerze*) schneller als unrelated Wörter, aber nur für SOAs von 100 und 200 ms. Dies zeigt, dass während der Planung des finalen Idiom-Bestandteiles dessen Bedeutung ko-aktiviert wird. Dieser Effekt tritt aber früh und nur kurzzeitig auf.

Rommers, Dijkstra und Bastiaansen (2013) untersuchten semantische Erwartungs-Effekte während der Verarbeitung von Idiomen mittels Offline- und Online-Methoden und fanden keine Evidenz für die Aktivierung der Einzelwort-Bedeutung. Die Autor:innen führten eine

semantische Erwartungsaufgabe durch (siehe Kapitel 2.2), bei der die Versuchspersonen Idiome innerhalb eines Satzkontextes, der eine figurative Interpretation des Ausdrucks begünstigte, lasen Wort für Wort (ndrl. *Na vele transacties liep de onvoorzichtige fraudeur uiteindelijk tegen de lamp gisteren.*, dt. *Nach vielen Transaktionen lief der unvorsichtige Betrüger gestern gegen die Lampe.*, Bedeutung: gefasst werden). Das Idiom wurde mit drei verschiedenen Varianten vervollständigt: (i) mit den stark erwarteten, idiomatisch korrekten Wörtern (*lamp*, dt. *Lampe*) oder mit inkorrekten Wörtern, die entweder (ii) semantisch related zu dem erwarteten Wort (*kaars*, dt. *Kerze*) oder (iii) unrelatiert (*vis*, dt. *Fisch*) waren. In einem ersten Experiment führten sie mit diesem Material eine lexikalische Entscheidungsaufgabe zu dem Zielwort durch. Versuchspersonen reagierten schneller auf idiomatisch korrekte Zielwörter als auf inkorrekte Zielwörter. Die Reaktionszeiten zwischen relateden und unrelatierten Zielwörtern unterschied sich nicht.

In einem zweiten Experiment, in dem sie EKP auf dem Zielwort maßen, ohne dass die Versuchspersonen eine Aufgabe bearbeiteten, fanden Rommers, Dijkstra und Bastiaansen (2013) ebenfalls keine Evidenz für eine Aktivierung der Einzelwort-Bedeutung. Sie fanden stärkere N400-Amplituden für beide inkorrekten Zielwörter als für korrekte Zielwörter. Inkorrekte Zielwörter unterschieden sich nicht hinsichtlich ihrer N400-Amplitude. Die Autor:innen interpretieren diesen Effekt insofern, dass während des lexikalischen Zugriffs auf einen Idiom-Bestandteil dessen Bedeutung nicht ko-aktiviert wird. Zusätzlich fanden sie eine späte Positivierung über posteriore Elektroden (vergleichbar zur PNP-Komponente, siehe Kapitel 2.2), die nicht durch semantische Relationen der Zielwörter manipuliert wurde. Den Autor:innen zufolge zeigt diese Komponente die zugrundeliegende, kohärente Repräsentation der Idiom-Konfiguration im mentalen Lexikon auf.

Insgesamt finden Rommers, Dijkstra und Bastiaansen (2013) in beiden Aufgaben keine Hinweise auf die Aktivierung der Einzelwort-Bedeutung ausgehend von der Erwartung des idiom-finalen Wortes. Im Vergleich zu den anderen beschriebenen Studien, die die Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomem untersuchten, sind Rommers und Kollegen somit die Einzigen, die keine positiven Befunde in Hinsicht auf eine Einzelwortaktivierung fanden. Sie erklären die Abwesenheit dieses Effektes dadurch, dass im Vergleich zu Bottom-Up-Verarbeitungsmechanismen eine Top-Down-Erwartung nicht ausreichend ist, um die Einzelwort-Bedeutung zu aktivieren. Allerdings wurden in der Studie von Sprenger und Kolleg:innen (2006) die idiom-finalen Wörter ebenfalls nicht präsentiert. Somit konnte die Aktivierung der Einzelwort-Bedeutung nicht aufgrund des sensorischen Inputs, sondern nur durch die Top-Down-Erwartung des idiom-finalen Wortes erfolgen. Alternativ könnte der figurative Kontext darin resultieren, dass die nicht-wörtliche Verwendung des Idioms früher erkannt wird und deshalb die Einzelwort-Bedeutung nicht aktiviert wird (Erklärung siehe van Ginkel & Dijkstra, 2019). Nichtsdestotrotz war die Aktivierung der Einzelwort-Bedeutung auch innerhalb von figurativen Kontexten während der Bottom-Up-Verarbeitung möglich (Holsinger, 2013,

Experiment 2). Daher liegt noch keine eindeutige Erklärung zu den abweichenden Befunden vor.

Widersprüchliche, empirische Befunde hinsichtlich der Rolle einzelner Bestandteile innerhalb MWE erschweren die Konstitution eines Modells, wie MWE im mentalen Lexikon repräsentiert sind (vgl. 1.1). Die Studie von Rommers, Dijkstra und Bastiaansen (2013) unterschied sich in einigen Faktoren von den anderen Studien, die eine Aktivierung der Einzelwort-Bedeutung nachweisen konnten (Zusammenfassung Tabelle 1): in der Modalität (Lesen vs. Hören) und der daraus resultierender Präsentationsart (auditiver Wortstrom, Wort-für-Wort-Lesen, Fragment-Lesen), im Paradigma (Priming, Visual-World, Erwartung, Produktion) und der daraus resultierenden Verarbeitungsmechanismen (Bottom-Up-Priming vs. Top-Down-Erwartung/Planung), in der Vorhersagbarkeit der Idiome und im Kontext (minimaler, neutraler Kontext, figurativer Kontext). Hier stellt sich die Frage, inwiefern die methodologischen Unterschiede in Bezug zu den abweichenden Befunden stehen. Daher geht **Studie I** (Kapitel 4) insbesondere auf diese Fragestellung ein.

Tabelle 1: Zusammenfassung über Studien, die die Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomem gemessen haben.

Autoren	Idiom: Modalität (Präsentationsart)	Zielwort: Modalität	Paradigma	Abhängige Variable	Verarbeitungsmechanismus	Kontext	Vorhersagbarkeit des letzten Bestandteils	Einzelwort-Bedeutung
Beck und Weber (2016a)	auditiv (inkrementell)	visuell	Cross-modales Priming + Lexikalische Entscheidungsaufgabe	Reaktionszeiten	Bottom-Up	minimal/neutral	nicht kontrolliert	aktiviert
van Ginkel und Dijkstra (2010)	visuell (Word-für-Wort)	visuell	Uni-modales Priming + Lexikalische Entscheidungsaufgabe	Reaktionszeiten	Bottom-Up	minimal/neutral	nicht kontrolliert	aktiviert
Holsinger (2013, Experiment 1)	auditiv (inkrementell)	visuell	Visual-World-Paradigma	Fixationen	Bottom-Up	minimal/neutral	nicht kontrolliert	aktiviert
Holsinger (2013, Experiment 2)	auditiv (inkrementell)	visuell	Visual-World-Paradigma	Fixationen	Bottom-Up	figurativ	nicht kontrolliert	aktiviert (nur früh)
Smolka et al. (2007)	visuell (Word-für-Wort)	visuell	Uni-modales Priming + Lexikalische Entscheidungsaufgabe	Reaktionszeiten	Bottom-Up	minimal/neutral	hoch	aktiviert
Sprenger et al. (2006, Experiment 2)	visuell (Fragment)	auditiv	Produktion	Produktionszeiten	Top-Down	minimal/neutral	hoch	aktiviert (früher stärker als später)
Sprenger et al. (2006, Experiment 3)	visuell (Fragment)	visuell	Produktion	Produktionszeiten	Top-Down	minimal/neutral	hoch	aktiviert (nur früh)
Rommers, Dijkstra und Bastiaansen (2013, Experiment 1)	visuell (Word-für-Wort)	visuell	Erwartungsparadigma + Lexikalische Entscheidungsaufgabe	Reaktionszeiten	Top-Down	figurativ	hoch	nicht aktiviert
Rommers, Dijkstra und Bastiaansen (2013, Experiment 2)	visuell (Word-für-Wort)	visuell	Erwartungsparadigma	ERP	Top-Down	figurativ	hoch	nicht aktiviert

4 Studie I

wie ein Strohfeder

4.1 Forschungsfrage

Studie I der vorliegenden Dissertation untersuchte die Fragestellung, ob und unter welchen Umständen Einzelwort-Bedeutungen innerhalb von Idiomen verarbeitet werden. Da sich die Modelle zur Repräsentation von MWE unter anderem in der Funktion der Einzelwort-Bedeutung unterscheiden (Kapitel 1.1), können die Ergebnisse von **Studie I** Rückschlüsse über entsprechende Modelle bieten. Während die meisten Studien positive Belege für die Aktivierung der Einzelwort-Bedeutung finden (Beck & Weber, 2016a; Holsinger, 2013; Smolka et al., 2007; Sprenger et al., 2006), widerspricht insbesondere die Studie von Rommers, Dijkstra und Bastiaansen (2013) diesen Befunden. **Studie I** knüpft an diesen Widersprüchlichkeiten zwischen bisherigen empirischen Befunden an. Das Ziel von **Studie I** ist zu untersuchen, ob Unterschiede im Studiendesign auch die widersprüchlichen Ergebnisse erklären können. Da Rommers, Dijkstra und Bastiaansen (2013) als einzige in der bisherigen Literatur keine semantische Aktivierung für Idiom-Bestandteile finden, baut das vorliegende Studien-Design grundlegend auf dem von Rommers und Kollegen auf. Das verwendete Stimulusmaterial ist im Anhang aufgelistet.

In den Experimenten in **Studie I** präsentierten wir Idiome nur innerhalb eines minimalen, neutralen Kontextes (z.B. *Hannes ließ die Katze aus dem Sack.*). Alle Studien die eine Aktivierung der Einzelwort-Bedeutung gefunden haben, präsentierten Idiome in einem minimalen, neutralen Kontext (Beck & Weber, 2016a; Holsinger, 2013; Smolka et al., 2007; Sprenger et al., 2006; van Ginkel & Dijkstra, 2019). Ausschließlich die Studie von Rommers, Dijkstra und Bastiaansen (2013) bettete die Idiome in einen figurativen Kontext ein. Innerhalb eines figurativen Kontextes kann die figurative Bedeutung eines Idioms früher erkannt werden (z.B. Fanari et al., 2010). Ein figurativer Kontext, könnte eine Ablehnung der wörtlichen Bedeutung begünstigen. Weiterhin wurden Wörter innerhalb der Idiome in der Studie von Rommers, Dijkstra und Bastiaansen (2013) aufgrund des Kontextsatzes getrennt (Bestandteile des wörtlich übersetzten Idioms unterstrichen: *Nach vielen Transaktionen lief der unvorsichtige Betrüger gestern gegen die Lampe.*), wodurch die konventionalisierte Wort-Konfiguration unterbrochen wird. Innerhalb eines minimalen, neutralen Kontextes kann das Idiom in zusammenhängender Form präsentiert werden. Wir wollen daher untersuchen, ob wir in einem minimalen, neutralen

Kontext mit einem ähnlichen Paradigma wie Rommers, Dijkstra und Bastiaansen (2013) deren Ergebnisse (keine Aktivierung der Einzelwort-Bedeutung) replizieren können.

Im Hinblick auf den Verarbeitungsmechanismus fokussierten wir uns auf Top-Down-Erwartungsprozesse, weil es hierbei widersprüchliche Ergebnisse zwischen Produktions- und Verarbeitungsstudien gab (Rommers, Dijkstra & Bastiaansen, 2013; Sprenger et al., 2006). Da aufgrund der Bekanntheit mit der konventionalisierten Wort-Konfiguration eines Idioms eine starke Erwartung für den letzten Bestandteil aufgebaut werden kann (z.B. Carrol & Conklin, 2020; Tabossi et al., 2005), kann dessen Verarbeitung in einem Erwartungsparadigma gemessen werden. In allen Experimenten verwendeten wir bekannte Idiome, innerhalb derer das letzte Wort stark vorhersagbar war (z.B. *Hannes ließ die Katze aus dem Sack.*). Ähnlich wie Rommers und Kollegen haben wir in **Studie I** unter Verwendung eines semantischen Erwartungsparadigmas gemessen, ob in stark vorhersagbaren Idiomen tatsächlich eine Erwartung für das finale Wort aufgebaut wird und ob diese Erwartung auch dessen semantische Merkmale ko-aktiviert. Dies kann man anhand von Verarbeitungsunterschieden zwischen Wörtern messen, die entweder zu dem idiom-finalen Wort semantisch relatiert (*Korb*) oder unrelatiert (*Bauch*) sind (siehe Kapitel 3.3).

Das semantische Erwartungsparadigma implementierten wir in Eye-Tracking- und EKP-Experimenten (siehe Kapitel 2.2), da Online-Methoden möglicherweise sensitiver gegenüber der Messung automatischer semantischer Aktivierung sind als Offline-Methoden (Heil et al., 2004). Beide Methoden weisen gewisse Vor- und Nachteile auf (z.B. die Messung des zeitlichen Verarbeitungsverlaufs oder die verwendbare Modalität), ergänzen sich aber in dieser Hinsicht sehr gut.

In einem Visual-World-Eye-Tracking-Experiment (Abbildung 9) hörten Versuchspersonen Idiomfragmente bis zum letzten Wort (z.B. *Hannes ließ die Katze aus dem...*). Auf dem Bildschirm wurden vier geschriebene Wörter präsentiert, wovon (i) eines erwartet und idiomatisch korrekt war (*Sack*), (ii) eines unerwartet und semantisch relatiert zu der idiomatisch korrekten Vervollständigung war (*Korb*) und (iii) zwei unrelatierte Wörter waren (*Arm*, *Bauch*). Während die Versuchspersonen das Idiomfragment hörten, wurde die Anzahl und der Zeitverlauf der Fixationen auf die jeweiligen Wörter gemessen. Anhand des Zeitverlaufs der Fixationen kann gemessen werden, welche Vorhersagen während der laufenden Sprachverarbeitung getroffen werden. Nachteilig an dieser Methode ist jedoch, dass die Wörter jeweils auf dem Bildschirm präsentiert werden. Dadurch ist man zum einen auf eine Auswahl an Wörtern beschränkt und könnte durch die simultane Präsentation dieser Wörter deren Aktivierung evozieren.

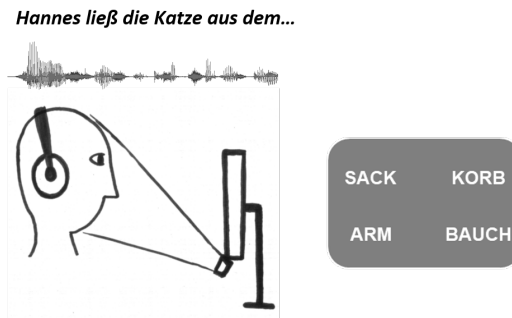


Abbildung 9: Eye-Tracking-Experiment: Versuchsperson hört Idiomfragment als auditiven Wortstrom während sie vier Wörter auf dem Bildschirm sieht.

Wir führten weiterhin zwei EKP-Experimente durch, einmal mit gesprochener Sprache und einmal mit Wort für Wort geschriebener Sprache (Abbildung 10). In diesen Experimenten präsentierten wir Idiome entweder in (i) korrekter Variante (*Hannes ließ die Katze aus dem Sack.*) oder in inkorrekt Variante. In den inkorrekten Varianten wurde das idiom-finale Wort entweder durch (ii) ein semantisch relatedes Wort (*Hannes ließ die Katze aus dem Korb.*) oder durch (iii) ein unrelatiertes Wort ersetzt (*Hannes ließ die Katze aus dem Arm.*). Im Vergleich zu der Realisierung in einem Visual-World-Paradigma, werden korrekte Zielwörter nicht innerhalb der Bedingung mit relateden oder unrelatierten Wörtern präsentiert.

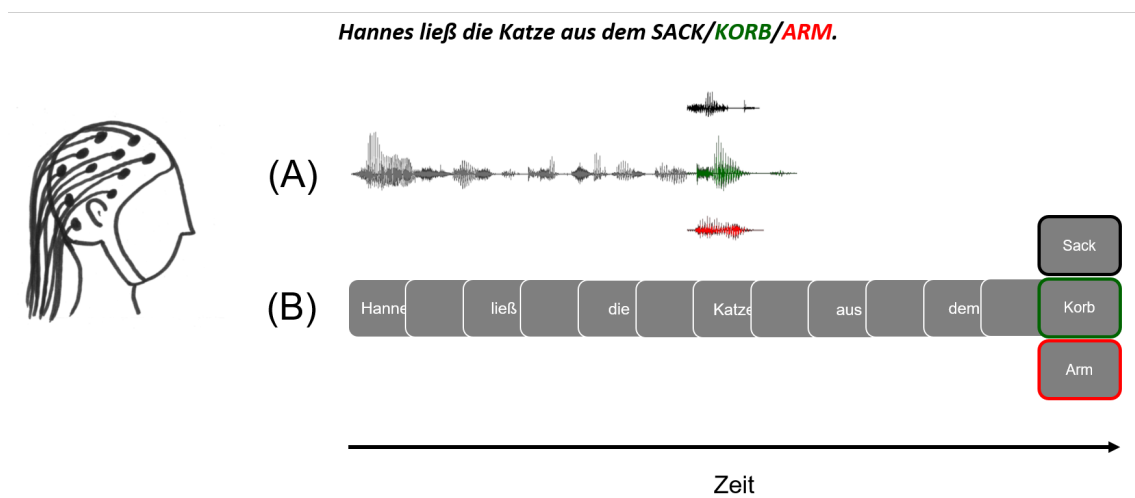


Abbildung 10: EKP-Experimente in den drei Bedingungen (schwarz = korrekt; grün = related; rot = unrelatiert): Versuchsperson (A) hört Varianten eines Idioms als auditiven Wortstrom bzw. (B) liest Varianten eines Idioms Wort-für-Wort auf dem Bildschirm geschrieben (zwischen den einzelnen Wörtern erscheint jeweils ein leerer Bildschirm).

Innerhalb von EKP-Experimenten untersuchten wir Verarbeitungsmechanismen sowohl für die auditive als auch die visuelle Modalität. Die Verwendung einer bestimmten Modalität und der damit verbundenen Präsentationsart bewirken, dass unterschiedlich viele linguistische Informationen zeitgleich verfügbar sind (vgl. Kapitel 2.2). Die Präsentation gesprochener

Sprache ist durch einen zeitlich inkrementellen Verlauf gekennzeichnet, in dem nur einzelne linguistische Informationen zum selben Zeitpunkt verfügbar sind. Geschriebene Sprache kann innerhalb eines experimentellen Settings in verschiedenen Größeneinheiten präsentiert werden (Satz-für-Satz, Phrase-für-Phrase, Wort-für-Wort). Die Verwendung geschriebener Sprache innerhalb von EKP-Studien beschränkt die Präsentationsart auf die Wort-für-Wort-Ebene, da diese am wenigsten Augenbewegungsartefakte beim Lesen verursacht. Beim Wort-für-Wort-Lesen sind immer komplette Wortinformationen zur gleichen Zeit verfügbar. Um auf diese modalitäts-abhängigen Unterschiede einzugehen, führten wir die EKP-Studie sowohl für die auditiv als auch für die Wort-für-Wort-Präsentationsart durch.

Zusammenfassend adressierte **Studie I** folgende Fragen (Abbildung 11):

- (1) Bauen erwachsene Muttersprachler:innen eine Erwartung für das idiom-finale Wort auf? Innerhalb des Visual-World-Experimentes würde diese Erwartung in mehr Fixationen auf der korrekten als auf den übrigen Wörtern auf dem Bildschirm resultieren. Falls das idiom-finale Wort bereits vor dem Ende des Idiomfragments erwartet wird (Tabossi et al., 2005; Vespignani et al., 2010), wäre dies in zeitlich frühen Fixationen auf das korrekte Wort sichtbar. Hinsichtlich der EKP-Experimente wäre ein Erwartungseffekt in Amplitudenunterschieden zwischen korrekten und inkorrekten Bedingungen in der N400- und der PNP-Komponente messbar. Die Vorhersage des idiom-finalen Wortes würde Modelle unterstützen, die annehmen, dass Idiome eine gewisse Form von Kohärenz im mentalen Lexikon aufweisen.
- (2) Aktivieren erwachsene Muttersprachler:innen die Einzelwort-Bedeutung der idiom-finalen Wörter, während sie diese vorhersagen? Falls Einzelwort-Bedeutungen während der Vorhersageprozesse ko-aktiviert werden, wäre dies im Visual-World-Experiment in Fixationsunterschieden zwischen relatierten und unrelatierten Wörtern erkennbar. Ähnlich erwarteten wir in diesem Fall Amplitudenunterschieden in der N400-Komponente zwischen relatierten und unrelatierten Vervollständigungen.



Abbildung 11: Bauen erwachsene Muttersprachler:innen eine Erwartung für das idiom-finale Wort auf und aktivieren sie währenddessen dessen Einzelwort-Bedeutung?

4.2 Manuskript: Activation of literal word meanings in idioms: Evidence from eye tracking and ERP experiments

Abstract

How the language processing system handles formulaic language like idioms is a matter of debate. We investigated the activation of constituent meanings by means of predictive processing in an eye tracking experiment and in two ERP experiments (auditory and visual). In the eye tracking experiment, German-speaking participants listened to idioms in which the final word was excised (*Hannes let the cat out of the...*). Well before the offset of these idiom fragments, participants fixated on the correct idiom completion (*bag*) more often than on unrelated distractors (*stomach*). Moreover, there was an early fixation bias towards semantic associates (*basket*) of the correct completion, which ended shortly after the offset of the fragment. In the ERP experiments, sentences (spoken or written) either contained complete idioms, or the final word of the idiom was replaced with a semantic associate or with an unrelated word. Across both modalities, ERPs reflected facilitated processing of correct completions across several Regions of Interest (ROIs) and time windows. Facilitation of semantic associates was only reliably evident in early components for auditory idiom processing. The ERP findings for spoken idioms compliment the eye tracking data by pointing to early decompositional processing of idioms. It seems that in spoken idiom processing, holistic representations do not solely determine lexical processing.

Introduction

There is an open debate in psycholinguistic research on whether and how formulaic sequences or multi-word expressions, as for example in collocations (*black coffee*), phrasal verbs (*dig into something*), or idioms (*kick the bucket*), are stored in the mental lexicon (for a review, see Conklin & Schmitt, 2012). In some accounts, the linguistic system is assumed to store formulaic sequences as larger units and to process them holistically (e.g., Jackendoff, 2002; Swinney & Cutler, 1979; Wray, 2005). According to this account, formulaic sequences have their own lexical entry comparable to ‘long words’. More recently, other accounts emphasize the internal syntactic and semantic structure of these multi-word expressions (e.g., Kyriacou, Conklin, & Thompson, 2020; Mancuso, Elia, Laudanna, & Vietri, 2020; Marantz, 2005; Snider & Arnon, 2012; Sprenger, Levelt, & Kempen, 2006; Tremblay & Baayen, 2010). While parsers are indeed sensitive to phrase frequencies, they access representations of all individual constituents in a phrase simultaneously (Arnon & Christiansen, 2017). According to these accounts, single constituents within multi-word units can be accessed separately.

In order to capture the hybrid nature of multi-word sequences, accounts of idiom processing have been proposed in which the structural properties of an idiom are preserved, while its meaning and form are also stored holistically. For example, the Configuration Hypothesis by Cacciari and Tabossi (1988) assumes that idioms are processed like novel, literal language,

but only until the parser recognizes a phrase as an idiom. After this ‘idiom key’, the parser directly retrieves the idiom configuration and associated meaning from the mental lexicon. According to a multidetermined view of idiom processing, factors such as familiarity or literal plausibility in addition to predictability determine the time point of recognition (Libben & Titone, 2008; Titone, Lovseth, Kasparian, & Tiv, 2019). Thus it is not surprising that in highly predictable idioms the recognition of the phrase can occur prior to the final word (Cacciari & Corradini, 2015). The Superlemma Hypothesis for speech production states that single word meanings within idiomatic expressions are necessarily activated (Sprenger et al., 2006). According to this view, idioms are accessible as both individual words (*simple lemmas*) and lexical units (*superlemmas*). In the present study, we investigated the neurocognitive reality of the representation of formulaic language in the mental lexicon by tracing the temporal dynamics of online activation of idiom constituent meaning.

The assumption of holistic processing is typically backed up by empirical evidence of greater processing ease for formulaic than for comparable non-formulaic language (e.g., Conklin & Schmitt, 2008; Gibbs, 1980; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Strandburg et al., 1993; Swinney & Cutler, 1979; Tabossi, Fanari, & Wolf, 2009; Tremblay, Derwing, Libben, & Westbury, 2011; Underwood, Schmitt, & Galpin, 2004). Several studies have found, for example, that participants read fixed multi-word expressions faster than novel phrases (e.g., Conklin & Schmitt, 2008; Tremblay et al., 2011), and that they fixate on words in idioms less extensively than on words in control sentences (Siyanova-Chanturia et al., 2011; Underwood et al., 2004). Based on holistic processing accounts, it has been argued that formulaic sequences are retrieved faster from the semantic memory than novel controls because there is no need for the parser to access single word meanings.

However, processing advantages might not originate exclusively from a purely holistic representation of formulaic phrases. They might also emerge from phrase frequency, predictive mechanisms for frequently co-occurring constituents, or phrase familiarity (e.g., Carrol & Conklin, 2020). For example, Canal, Vespignani, Molinaro, and Cacciari (2010) propose that predictive mechanisms within idioms are based on the knowledge of their specific lexical form in the mental lexicon and these predictive mechanisms might differ qualitatively from predictions within literal, non-formulaic expressions. Arguably, more direct evidence for holistic representations would show that the linguistic system does not process single words and their meanings separately, but receives multi-word expressions un-analyzed from the mental lexicon (see the discussion in, e.g., Siyanova-Chanturia, 2015).

Idioms are well suited for investigating holistic processing vs. decomposition into single constituents of formulaic expressions. In many idioms, the figurative meaning cannot be inferred from the compositional meaning of the constituent words. For example, the figurative meaning of *to let the cat out of the bag* (to reveal a secret unintentionally) is not derived from the

meaning of the single noun constituents (*bag* and *cat*) or from their combination with the verb (*to let*). Therefore, evidence that such multi-word idioms are obligatorily decomposed into their single constituents would strongly speak against a model assuming solely holistic processing of idioms not allowing access to single words. One way to test whether single constituents within idioms are processed individually is to measure the activation of semantic associates (*basket*) of these constituents (*bag*). Because, in general, activation of a word in the mental lexicon will spread to semantically related words (Collins & Loftus, 1975), activation of semantic associates within idioms would indicate that the parser processes individual constituents.

Following the approach of spreading semantic activation, priming and word production studies have indeed shown that parsers have single word meanings available quickly during idiom processing (e.g., Beck & Weber, 2016; Smolka, Rabanus, & Rösler, 2007; Sprenger et al., 2006; van Ginkel & Dijkstra, 2019). In these studies, participants typically first read idioms (Rabanus, Smolka, Streb, & Rösler, 2008; Smolka et al., 2007; van Ginkel & Dijkstra, 2019) or listened to idioms (Beck & Weber, 2016), like *to pull someone's leg* (meaning 'to spoof someone'), and subsequently performed a lexical decision task on immediately following written target words. Across these studies, participants responded faster to targets that were semantically related to the literal meaning of a constituent word (e.g., *walk*) compared to unrelated targets. In two experiments conducted by Sprenger and colleagues (2006, Experiment 2 and Experiment 3), participants read idiom fragments (e.g., *Jan liep tegen de [lamp]*, literally translated: *Jan walked against the [lamp]*, meaning 'to get caught' in Dutch) and were asked to complete the idiom by speaking aloud the final, missing noun (e.g., *lamp*). Both experiments tested whether participants have semantic associates of idiom-final words (e.g., *candle*) available while they prepare their responses. In Experiment 2, participants received a spoken prime while they prepared their response. Semantic associates facilitated participants' responses compared to unrelated primes. In Experiment 3, participants were prompted to produce the idiom-final word when a question mark appeared on the screen. However, when another word appeared on the screen instead of the question mark, they had to switch the task and produce that word. In this production task, participants responded faster to semantic associates of the idiom-final constituent compared to unrelated probes.

Evidence for spreading semantic activation originating from single idiom constituents was also found in an eye tracking study by Holsinger (2013). Participants listened to idiomatic phrases (*hit the hay*) while they saw four printed words on the screen, including an associate of a constituent word (*barn*). Fixations showed that participants considered the semantic associate more often than they considered unrelated distractors. Together, priming and eye-tracking results are in line with accounts assuming that the parser has idiom internal structures available (e.g., Marantz, 2005; Snider & Arnon, 2012; Sprenger et al., 2006; Tremblay & Baayen, 2010).

In contrast to priming and eye tracking work, data from an Event-Related Potentials (ERPs) study found no apparent involvement of single word meanings during idiom processing (Rommers, Dijkstra, & Bastiaansen, 2013). In this experiment participants read highly predictable Dutch idioms (e.g., literally translated *to walk against the lamp*). In a related condition, a semantic associate replaced the idiom’s final noun (*candle*), and in an unrelated condition, an unrelated word replaced the final noun (*fish*). Semantic associates of idiom-final nouns did not elicit different ERPs than completely unrelated words did (see Experiment 2, for further discussion of the specific ERP effects elicited in this study). Rommers and colleagues argued that participants did not form semantic predictions of idiom-final constituents. Results rather indicated holistic processing of idioms, as would be suggested by representational accounts viewing idioms as ‘large words’ (Jackendoff, 2002) or ‘lexical items’ (Swinney & Cutler, 1979), which are processed as a whole.

Design-related differences (such as modality, paradigm, and idiom characteristics) in previous studies might account for the mixed results regarding the processing of idiom constituents. For example, the modality in which idioms were presented differed between experiments and this comes with different amounts of linguistic information available to participants at any given point in time. While connected spoken language makes single words only sequentially available (as they are unfolding over time), written language makes complete words or phrases available at once. Using spoken idioms combined with written probes, Beck and Weber (2016) and Holsinger (2013) found semantic activation of single idiom constituents. Other studies presented idioms and probes visually, either phrase-wise (Sprenger et al., 2006, Experiment 3) or word-by-word (Rabanus et al., 2008; Rommers et al., 2013; Smolka et al., 2007). The experiments by Sprenger and colleagues (2006, Experiments 2 and 3) using phrase-wise presentation, where the whole idiom fragment was available at once, revealed semantic activation of the idiom constituent. Results were mixed for experiments using word by word presentation in a rapid serial sequence (Rabanus et al., 2008; Rommers et al., 2013; Smolka et al., 2007). Clearly, the time course of word recognition and semantic activation might differ depending on the amount of linguistic information available at a certain point in time (e.g., Anderson & Holcomb, 1995; Van Petten, Coulson, Rubin, Plante, & Parks, 1999) and this might play a role in processing differences found across different studies.

Different experimental paradigms could also relate to different results. In most studies that support decomposition of idioms (Beck & Weber, 2016; Holsinger, 2013; Rabanus et al., 2008; Smolka et al., 2007), activation of semantically related words might have resulted from bottom up spreading activation, due to the critical idiom constituent being actually presented. For example, the eye tracking study by Holsinger (2013) reported biased eye movements towards semantic associates (*barn*) shortly after the participants heard the critical idiom constituent (*hay*) as part of the idiom. Similarly, the critical constituent was part of the primes in priming studies showing semantic activation (Rabanus et al., 2008; Smolka et al., 2007; van Ginkel &

Dijkstra, 2019). In these studies, the critical idiom constituent might have briefly activated semantic associates in a bottom-up fashion without the idiom representation being involved. In contrast, participants were not presented with the critical idiom constituent (*lamp*) in the ERP study by Rommers and colleagues (2013), which did not find evidence for activation of semantic associates (*candle*). According to the authors of the latter study, the prediction of the correct idiom-final word might not be sufficient to activate single word meanings within idioms and, thus, no processing benefit for semantically related words was found. However, while critical idiom constituents were also not presented in the production study by Sprenger and colleagues (2006, Experiments 2 and 3), these authors did find that facilitation of semantically related words was induced merely by planning to produce the idiom-final constituent.

Finally, experiments differed in terms of idiom characteristics such as predictability. Depending on the amount of given linguistic constraints, individual idioms can be recognized prior to their last constituent (Libben & Titone, 2008). Earlier vs. later activation of the idiomatic form might result in higher vs. lower predictability of the idiom-final word (Canal et al., 2010). According to the Configuration Hypothesis (Cacciari & Tabossi, 1988), predictability might affect the activation of literal constituent meanings. Since in highly predictable idioms the idiom key should be well before the final constituent, literal activation of the latter would be less likely. Nevertheless, Rabanus et al. (2008), Rommers et al. (2013), Smolka et al. (2007), and Sprenger et al. (2006) measured lexical activation of highly predicted, idiom-final constituents and came to different conclusions. Taken together, different idioms used across different studies render comparisons of results obtained with different paradigms and presentation modalities difficult.

In the present study, we targeted the previously obtained inconsistencies regarding literal meaning activation of single idiom constituents. To this end, (i) we varied presentation modality by presenting idioms and probes cross-modally (Experiment 1), auditorily (Experiment 2), and visually (Experiment 3), (ii) we focused on top down prediction mechanisms, for example by not presenting the critical constituent in the input in order to discourage pure bottom up spreading of semantic information (Experiments 2 and 3), and (iii) we kept the idiom characteristics constant by using the same highly predictable idioms across experiments. Furthermore, we employed different implicit methods by relying on eye tracking (Experiment 1) and ERPs (Experiments 2 and 3) measures. Implicit online measures might be more sensitive in detecting spreading semantic activation (Heil, Rolke, & Pecchinenda, 2004).

Experiment 1

In Experiment 1, we addressed the question of semantic activation of idiom constituent meanings through predictive processing by conducting an eye tracking study. We exploited the tendency of gaze behavior (e.g., time course and amount of fixations) to be biased towards implicit linguistic aspects of displayed words (for a review, see Huettig, Rommers, & Meyer,

2011). Fixation biases include semantic associates of target words as reflected, for example, in more fixations towards the printed word *shark* while the word *turtle* is mentioned (Huettig & McQueen, 2011). These results imply that eye movements are a powerful tool to investigate bottom up spreading semantic activation exerted by spoken input.

In the eye tracking study on idiom processing by Holsinger (2013), participants' eye movements were attracted by semantic associates of idiom constituents while they listened to the idiom containing the respective constituent. For example, while listening to *hay* in *hit the hay*, participants fixated the printed word *BARN* more often than unrelated control words. That is, the design of this former study does not allow disentangling rapid bottom-up semantic spread exerted by the presentation of the single word and decomposition of the idiom during processing. In order to study the latter, we have to rely on a paradigm that does not present the critical idiom constituent in the input.

In order to avoid presentation of the critical idiom constituent, we exploited predictive processing in online comprehension. Numerous eye tracking studies have shown that participants use sentence contexts to predict upcoming words and their semantic properties (Altmann & Kamide, 1999, 2007; Kamide, Altmann, & Haywood, 2003). For instance, when participants listened to a sentence like *the boy will eat the cake*, they fixated on the picture of a cake in a visual scene at the offset of the verb *eat* (Altmann & Kamide, 1999). That is, eye fixations reflect predictions built during online processing before the critical word can exert bottom-up semantic spread. Therefore, prediction of semantic features for idiom constituents that are not part of the input, can indicate decomposable memory traces for idioms.

In order to investigate prediction within idioms, we measured predictive fixations to displayed words before the full idiom has been heard and processed. In Experiment 1, we used highly predictive German idiomatic phrases. Participants listened to incomplete idioms, missing the final critical word, without any biasing context (e.g., *Hannes ließ die Katze aus dem ...*, 'Hannes let the cat out of the ...'). Visual displays included four printed words: the correct idiomatic completion (*SACK*, 'BAG'), a semantic associate of the correct completion (*KORB*, 'BASKET'), and two unrelated distractors, with a semantic relation to each other (*ARM*, 'ARM' and *BAUCH*, 'STOMACH'). Participants had to choose which of the displayed words was the correct completion of the idiomatic phrase. In order to fixate the correct item, participants had to anticipate the complete idiom. This should result in fixations of the correct completions. Fixation biases to correct completion will be informative about idiom recognition. If semantic associates of single idiom constituents are available for predictive processing, this would be indicated in fixations to semantic associates of the correct completion as soon as the idiom is recognized. This would support word-by-word predictions based on decomposable memory traces for idioms. If semantic associates of single idiom constituents do not attract more fixations than unrelated distractors, this would speak for holistic idiom representations, not allowing a

word-by-word analysis.

Methods

Participants Thirty-one adults (mean age = 20.97 years, range = 18–30, 22 female, 9 male) participated in the experiment. Participants were recruited at the University of Tübingen and received subject credits as compensation. Prior to the experiment, participants gave written informed consent. All participants were native, monolingual speakers of German. Participants had no hearing impairments and normal or corrected-to-normal vision. Experiment 1 was approved by the Ethical Committee for Psychological Research at the University of Tübingen (reference number: 2016/1027/22).

Stimuli and Design We selected 20 well-known German idioms (see Appendix)¹. The idioms were embedded in sentences with a comparable structure (see Table 2): (i) a person carrying out the action of the sentence, (ii) a sentence body that originated from a German idiom, and (iii) the final target word of the idiom (which was not presented auditorily in Experiment 1). All idiomatic sentences were spoken in their complete form by a native speaker of German and digitally recorded. For Experiment 1, we removed the final target word from the recording. Participants heard each idiomatic sentence fragment once, while seeing four visual words on a computer screen. The four words represented one of these four types: (1) Correct Completion: correct completion of the idiomatic phrase, (2) Related Distractor: semantic associate of the correct completion, and (3&4) Distractors Unrelated 1 and Unrelated 2: semantically unrelated to the correct completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Correct Completions and Related Distractors used with other sentence fragments in the experiment (avoiding phonological overlap). All words on the screen had the same grammatical gender fitting the preceding sentence context.

We ensured semantic relatedness between correct and related words by comparing pairwise semantic spaces using the R package *LSAfun* (Günther, Dudschig, & Kaup, 2015) and testing these similarity values with a Wilcoxon signed rank test. On average, semantic similarity between correct-related word pairs was significantly higher than between correct-unrelated1 ($Z = 189, p < .001$) and correct-unrelated2 ($Z = 185, p = .002$) word pairs. Semantic similarity between correct-unrelated1 and correct-unrelated2 word pairs did not differ ($Z = 75, p = .28$). Furthermore, semantic similarity between correct-related word pairs was significantly higher than between related-unrelated1 ($Z = 180, p = .004$) and related-unrelated2 ($Z = 182, p = .003$) word pairs.

¹Since we had planned to test children with the same material and paradigm in the future, we only selected highly familiar short idioms that German children would already be expected to know.

Tabelle 2: German Example Sentence for Types 1-4 with English Equivalent

(i) Person	(ii) Sentence Body	(iii) Target Words			
		(1) Correct	(2) Related	(3) Unrelated 1	(4) Unrelated 2
Hannes	ließ die Katze aus dem	Sack	Korb	Bauch	Arm
<i>Hannes</i>	<i>let the cat out of the</i>	<i>bag</i>	<i>basket</i>	<i>stomach</i>	<i>arm</i>

Displayed words were presented in white font (Arial, font size 28) on a grey background. The position of the displayed words was counterbalanced across items and participants. The order of the trials was randomized.

Procedure Participants completed the experiment in a single session. For the experimental task, participants received both written and oral instructions. Prior to the experimental task, each participant received a 5 point grid for calibration and a practice block consisting of five trials.

An exemplary trial scheme is displayed in Figure 12. Each trial began with a 1500 ms inter-trial interval followed by a 500 ms presentation of a fixation cross. Then the visual display of the set of four words appeared on the screen and remained until the end of the trial. The presentation of the audio stimuli started after a total of 2150 ms and was presented via headphones. After they heard the auditory stimuli, the task of the participants was to decide for each item which of the visually presented words was the best completion for the idiom by saying their choice out loud.² The experimenter noted the participants' responses. Participants were instructed to press a button after their oral response in order to continue on to the next trial.

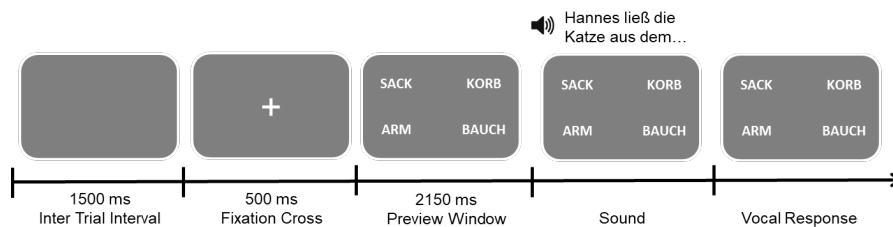


Abbildung 12: Example with times indicating the duration of the respective displays.

²Since we planned testing children with the same material and paradigm in the future, we had to adapt the paradigm. Therefore, we chose this long preview window of the four printed words so that there was enough time to read all four words before the onset of the auditory stimuli. Since the specific idiom knowledge of children is very different, we only wanted to include idioms that are known to the individual children. Therefore, we chose an overt task where participants had to find the correct idiomatic completion.

We recorded fixations using a portable Tobii eye-tracker with a sampling size of 60 Hz. In total, the eye tracking experiment took around 20 minutes including instructions, calibration and the experimental task, which took around 10 minutes.

Results

For the analysis, we divided the screen into four areas of interest. The analysis time window was aligned to the offset of each audio stimulus (offset = 0 ms). For the statistical analysis, we only included items responded to correctly, that is in which the participants completed the sentence aloud with the correct final word of the idiom (these were 98.87% of all trials). Figure 13 (Panel A) shows fixations proportions towards correct, related, and aggregated unrelated words as fixation proportions from 800 ms before to 1000 ms after the offset of the spoken stimuli. Running t-tests comparing fixations towards correct completions and unrelated distractors at succeeding measurement points (every 16.67 ms) showed that participants' fixations were biased towards the correct idiomatic completion 464 ms prior to the offset of the audio stimuli ($p < .01$). This can be interpreted as the recognition point of the idiom. To compare the amount and time course of fixations towards related and unrelated distractors, we conducted a Growth Curve Analysis (GCA) with orthogonal polynomials (Mirman, Dixon, & Magnuson, 2008). As the starting point of the GCA time window, we chose the start of observable anticipation (464 ms prior to the offset) for a duration of 1200 ms.

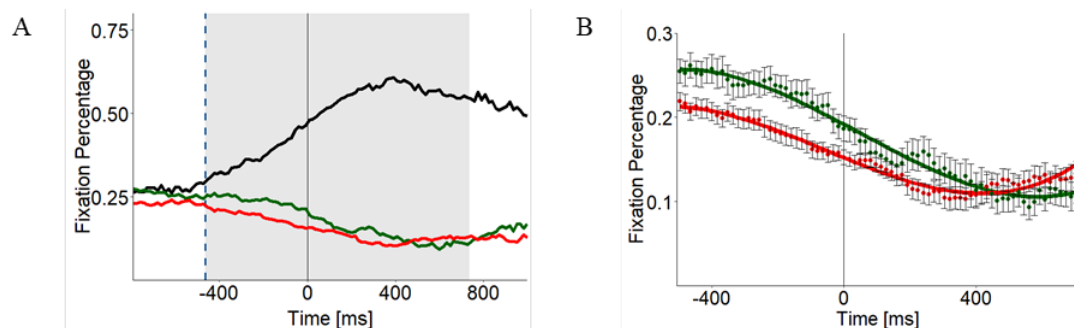


Abbildung 13: Panel (A) Fixation percentage for correct completions (black), related distractors (green) and mean of unrelated distractors (red); black vertical line = offset of spoken stimuli (0 ms); blue vertical, dashed line = start of the anticipation (-464 ms); grey background = time window for GCA. Panel (B) Fixation percentage for semantically related and unrelated distractors (points = mean; error bars = standard error) with fit of the growth curve model (line).

Fixation proportions were modeled with third-order orthogonal polynomials, because visual inspection of the time course bent at two points. To test the effect of Distractor Type (related vs. unrelated), we compared models using the $-2LL$ deviance statistic. Including the effect of Distractor Type significantly improved the model fit ($\chi^2 = 42.98$, $p < .0001$). Estimated parameter terms of Distractor Type are summarized in Table 3. The intercept term reflects

the average magnitude of the curve. Thus, the significant effect on the intercept term indicates that participants fixated more on related than on unrelated distractors across the complete time window. The linear term is comparable to the overall slope of the curve. In this case, the significant effect on the linear term implies variation across time with larger differences between the distractor types at the beginning of the time-window. The quadratic term reflects symmetric inflection of the curve around the center meaning. Thus, the curve of the related distractor is shallower than the curve of the unrelated distractor, and towards the end of the time window the proportions of looks to related and unrelated distractors converge. The cubic term reflects inflections of the curve at the ends of the analysis time window. We found no significant effect on this term.

Table 3: Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).

Term	Estimate	Standard Error	t	$p <$
Intercept	-0.0080	0.0031	-2.6224	.012
Linear	0.0655	0.0298	2.1950	.029
Quadratic	-0.0249	0.0049	-5.0924	.001
Cubic	-0.0066	0.0049	-1.3428	.180

Discussion

The predictive eye-movements recorded in an eye tracking paradigm in Experiment 1 were in line with previous priming and eye tracking research (Beck & Weber, 2016; Holsinger, 2013; Smolka et al., 2007; Sprenger et al., 2006) in showing that single word meanings are available in online processing of idioms. In Experiment 1, participants looked more often at distractor words that were related to the idiom’s final word than at unrelated distractor words. Moreover, this fixation bias emerged anticipatorily, meaning well before the point in time at which the idiom’s final word would have become evident in the speech signal. In fact, participants started to anticipate, i.e., look at, correct idiomatic completions around 460 ms prior to the offset of the phrase fragment. Because programming of saccades after onset of the critical word typically takes around 200 ms (Saslow, 1967), we can assume that recognition of the idiom occurred even before 460 ms. Simultaneously with the increase in fixations on correct idiomatic completions, the fixation bias for semantic associations emerged. The fixation bias towards semantic associates diminished over time and ended 400 ms after the offset of the phrase fragment. In sum, our eye tracking data suggest rapid prediction of upcoming idiom completions revealing that listeners represented these ordered strings in their mental lexicon.

In addition, predictive eye movements to semantic associates of idiom completions demonstrate that listeners not only pre-activate and predict words within idioms in a holistic fashion, but they also appear to pre-activate single constituents together with their respective meanings.

For some of the used idiom fragments, the related distractor provided a literally plausible interpretations which might have compromised the fixation towards this distractor. In a post-hoc visual inspection, we plotted fixation data for items that allow a literal interpretation of the related completion (*Sarah band sich einen Klotz ans Knie.*, literally translated: *Sarah tied herself a chunk to her knee.*) and implausible, related completions (*Hannah schlug sich die Zeit um die Augen.*, literally translated: *Hannah hit herself the time around the eyes.*) separately. We did not observe decreased semantic activation for literally implausible, related completions supporting the interpretation of pre-activation of idiom constituents together with their semantic features. This complements the results from a visual world experiment using literal, novel phrases that show anticipatory fixations towards predicted words and semantic competitors, although the latter were implausible completions of the phrase (Ito & Husband, 2017).

The relatively long preview window that we implemented in the present experiment might have biased participants towards predictive processing. For example, Ferreira, Foucart, and Engelhardt (2013), suggested that preview time is associated with the strength of expectations participants form. Accordingly, longer preview of words or objects on the display is associated with stronger expectations participants form with regard to which word on the display is likely to be referred to. In this case, we would expect participants of Experiment 1 to build up stronger expectations for the correct idiom completion as part of the conventionalized phrase, and therefore weaken any tendency to look at related words. As a result of these stronger expectations, we might have overestimated the timing of the anticipation onset.

Another aspect of the eye tracking design in Experiment 1 potentially limits its straightforward interpretation: the visually presented probes might have induced spreading semantic activation in a bottom-up fashion. Although we did not present a spoken version of the idiom-final constituent, a written version of it was available on the visual display, simultaneously with a written version of its semantic associate. Thus, fast fixations towards the correct idiomatic completion might have induced fast visual word processing and spreading semantic activation, which might have rapidly biased fixations towards the semantic associate. However, similar onsets of the fixation biases towards correct completions, on the one hand, and semantic associates, on the other hand, somewhat restrict an interpretation in terms of spreading activation exerted by the visual versions of the correct completions, because this mechanism might need some extra processing time (i.e., visual word recognition of the correct completion, spreading activation, and elicitation of eye movements towards its semantic associate). Nevertheless, similar to results of other studies (Beck & Weber, 2016; Holsinger, 2013; Smolka et al., 2007), the present eye

tracking data might overestimate decomposition because an instance of the critical constituent was visually included in each trial. In Experiment 2 and Experiment 3, we attempted to further rule out this alternative interpretation by avoiding any presentation of the critical idiom constituent for which we attempt to measure prediction effects in ERP experiments.

Experiment 2

In the following two experiments, we exploited semantic expectancy in spoken (Experiment 2) and written idioms (Experiment 3) in an ERP paradigm comparable to that of Rommers and colleagues (2013). As in the former study, we focused on N400 effects. Typically, reduction of the N400 ERP component is related to facilitated semantic processing, including semantic expectancy mechanisms (e.g., Federmeier & Kutas, 1999; Kutas & Federmeier, 2011; Laszlo & Federmeier, 2009). The N400 is a centro-posterior negative-going ERP component peaking around 400 ms after word onset. In N400 experiments, semantic expectations are usually determined via the cloze probability of a critical word within a given context. This measure reflects how often participants complete a phrase or sentence with a specific word. The N400 amplitude inversely correlates with this index: the higher the cloze probability of a word, the smaller the N400 amplitude it elicits (Kutas & Hillyard, 1984). Respective predictive mechanisms are so strong that even the processing of an unexpected word (with a low cloze probability) might reduce N400 amplitude if it shares semantic features with the expected stimulus (e.g., Federmeier & Kutas, 1999; Federmeier, McLennan, De Ochoa, & Kutas, 2002).

Evidence for the sensitivity of the N400 to the prediction of semantic features originally came from Federmeier and Kutas (1999), who presented participants with written versions of highly predictive sentences, such as ‘*They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of...*’. Sentences ended with either a highly expected word (*palms*), an unexpected word from the same semantic category (*pin*) or an unexpected word from a different semantic category (*tulips*). In this experiment, the N400 amplitude for unexpected words from both categories clearly differed from the N400 amplitude for expected words. Moreover, N400 amplitudes were graded: words from the same semantic category as the expected word elicited a significantly smaller N400 amplitude than words from a different semantic category. Therefore, the N400 effect shows that semantic features of expected words are co-activated during online comprehension and words sharing these features benefit from predictive processing.

In the context of written idioms, Rommers and colleagues (2013) did not replicate the N400 prediction effect for semantic associates of final words. Participants read Dutch idioms embedded in figuratively biasing contexts (*After many transactions the careless scammer eventually walked against the lamp yesterday.*) in which the final word of the embedded idiom was either correct (*lamp*), not expected but from the same semantic category as the correct completion (*candle*), or not expected and from a different semantic category (*fish*). An N400

reduction for correct idiom-final words was found. This effect emerged with the typical topography (posterior) and within the typical time window of the N400 (300–400 ms). Yet semantic associates of correct idiom completions did not elicit an N400 reduction. That is, ERPs did not indicate facilitated processing of semantic associates of single idiom constituents. In addition to the N400 effect, Rommers and colleagues (2013) found a reduced late positivity ranging between 500 and 800 ms for correct idiom completions compared to related and unrelated completions. Again, the related and the unrelated condition did not differ. Rommers and colleagues interpreted this positivity as an instance of the P600 component reflecting a violation of the idiom representation as a linguistic unit.

In Experiment 2 and Experiment 3, we adopted the semantic expectancy ERP paradigm by Rommers and colleagues (2013) to preclude possible bottom up spreading semantic activation (as in the eye tracking paradigm in Experiment 1). In Experiment 2, we examined spoken versions of idioms in a unimodal design in order to be able to relate the results to previous cross-modal designs with spoken idioms that found activation of semantic associates of idiom constituents (eye-tracking paradigm in Experiment 1, Beck & Weber, 2016; Holsinger, 2013). In the literature, results for the semantic N400 effect in sentences is fairly comparable for visual and auditory processing (Connolly, Phillips, Stewart, & Brake, 1992; Federmeier et al., 2002; Hagoort & Brown, 2000). This includes semantic expectancy effects (Federmeier et al., 2002). Only the onset of N400 might differ, in that it starts earlier for auditory than for visual processing. It is still a matter of debate whether this early onset is functionally different from the N400 or not (Connolly & Phillips, 1994; Diaz & Swaab, 2007; Nieuwland, 2019; Van Den Brink, Brown, & Hagoort, 2001).

We again presented German idioms in short sentences without further context, including the ones we used in our eye tracking study (Experiment 1) as well as additional items. Participants listened to highly predictive idiomatic phrase onsets (e.g., *Hannes ließ die Katze aus dem . . .*, ‘Hannes let the cat out of the . . .’). Phrase onsets were completed either (1) with the expected and correct final idiom word (*Sack*, ‘bag’), (2) with an unexpected but semantically related completion (*Korb*, ‘basket’), or (3&4) with an unexpected and semantically unrelated completion (*Arm*, ‘arm’, *Bauch*, ‘stomach’). If processing is solely holistic, the words in related and unrelated conditions should show comparable ERP amplitudes as was shown by Rommers and colleagues (2013). Such a finding would suggest that fixations towards semantic associates of correct completions in Experiment 1 were merely an epiphenomenon of bottom-up spreading activation exerted by the visual probe being presented together with the correct completion within the same display. If literal meanings of expected words are accessed, the processing of semantically related words should benefit more from this expectation when compared to unrelated words. This would yield graded ERP amplitudes for related and unrelated completions.

Methods

Participants Forty-two healthy participants volunteered for Experiment 2. None of the participants had taken part in Experiment 1. We excluded the data of one bilingual participant and of one participant for whom we had technical problems with the ERP recording. Participants whose data were included in the analysis ($N = 40$, mean age = 22.9 years, range = 18–32, 20 female and 20 male) were right-handed as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971), monolingual native speakers of German, and had no history of a neurological, psychiatric, or hearing disorder. As compensation, subjects were paid for the experiment or provided with subject credits. Experiment 2 was approved by the Ethical Committee of the German Psychological Society (reference number: RK 112015).

Stimuli In order to arrive at a sufficient number of trials for an ERP study, we extended the experimental materials from Experiment 1 from 20 to 40 phrases using the same criteria of familiarity and predictability (see Appendix). Linguistic stimuli resulted from the combination of the sentence body with the four sentence final target words in four conditions with a combination logic following that of Experiment 1 (see Table 2). The conditions were the following: (1) Correct Condition: the target word was the correct completion of the idiomatic phrase, (2) Related Condition: the target word was semantically related to the correct completion, and (3&4) Conditions Unrelated 1 and Unrelated 2: the target word was semantically unrelated to the correct completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Correct and Related Conditions used with other sentence bodies in the experiment (no phonological or semantic overlap). Each sentence body was repeated four times, once in all four conditions. This resulted in 160 different combinations of sentence bodies and target words. The same native speaker of German as in Experiment 1 spoke all linguistic stimuli. The linguistic stimuli that were repeated across conditions (sentence body and final words) were realized as the same recordings.

We conducted rating studies to determine some characteristics of the materials essential for ERP research. In a cloze probability task, 17 participants read the 40 sentence bodies and filled in the word that they considered to be the most likely completion. The mean cloze probability of the correct idiom-final word was 93.82% ($SD = 9.69$).

Furthermore, we controlled for the semantic relatedness between critical words by means of a second rating study. Fifteen participants received lists of word pairings of the target words and judged their relatedness on a scale from one to seven. The association strength between words presented as critical words in the Correct Condition (i.e., between the correct idiom completion) and words presented in the Related Condition (see Table 1) was significantly higher than the association strength of critical words presented in the Correct Condition and both Unrelated Conditions (Wilcoxon signed rank test: Unrelated 1 $Z = 120$, $p < .001$; Unrelated 2 $Z = 120$, $p < .001$). The association strength between critical words presented in the Correct Conditions and those presented in both Unrelated Conditions did not differ ($Z = 78$, $p = .32$).

Procedure Participants completed the experimental task in a single session. After signing an informed consent form, participants sat in a comfortable chair facing a computer screen in a dimly lit room. During the experimental task, they were instructed to sit still and avoid eye-movements including blinking. Later, participants took part in a calibration task at the beginning and the end of the session. In this task, eye movements were systematically evoked for offline ocular correction. Before the experimental task, participants received both written and oral instructions. The participants received a practice block consisting of eight trials to ensure that they were familiar with the procedure and the task.

For each experimental trial, a sentence was presented auditorily via loudspeakers on both sides of the computer screen. During the presentation of the sentences, participants viewed a fixation cross at the center of the screen. After the auditory presentation, the task of the participants was to decide for each sentence whether it was a correct idiomatic phrase or not by pressing buttons with the index fingers of the right or the left hand.³ The side for yes- and no-buttons was counterbalanced across participants. The response type was a delayed response; 1200 ms after onset of the target stimulus a question mark appeared at the center of the screen to signal the start of the response window for the participants. If they responded before the start of the response window, participants were given feedback (*too fast*). The interval between succeeding trials was 1500 ms.

The experiment consisted of 8 blocks of 20 trials, 160 trials in total, with five trials in each condition in each block. The order of trials was pseudorandomized in such a way that the same sentence body or target word never occurred in the same block. After each block, participants had the opportunity to take a self-timed break. The order of blocks was randomized using the Latin Square method. In total, the EEG-experiment took around 1.5 h including electrode application, instruction, calibration and the experimental task; the experimental task itself took around 15–20 minutes.

Electrophysiological Recordings Electrophysiological brain potentials were recorded with 46 active electrodes (Ag/AgCl) mounted in an elastic cap (Easycap GmbH, Herrsching, Germany) according to the 10-20 system (see Figure 14), online referenced to the nose. The ground electrode was positioned at the location of the AF3. In order to record eye movements, we attached two ocular electrodes below both eyes. The raw data were sampled at 500 Hz (bandpass filter 0.01–100 Hz, BrainAmpStandard, Brain Products, Gilching, Germany).

EEG Analysis For the ERP analysis, the raw data were re-referenced offline to the average reference and filtered with a 0.3 Hz Low-Cut-Off filter. Using surrogate MultipleSource

³As in Experiment 1, we chose an overt idiom recognition task because we wanted to conduct the same experiment with children. In order to account for differing idiom knowledge between children, we wanted to include only idioms that participants recognized correctly. A similar idiom recognition task was used in Qualls, Treaster, Blood, and Hammer (2003).

EyeCorrection (MSEC) by Berg and Scherg (1994), we removed horizontal and vertical eye movements as well as blinks from the continuous EEG signal. The EEG data were segmented into trials in epochs from 100 ms before and 1000 ms after the stimulus onset with a 100 ms pre-stimulus baseline subtraction. We excluded trials contaminated with artifacts and in which participants responded before the onset of the response time window 1200 ms after stimulus onset. Further, we only included individual items that participants responded to correctly in the Correct Condition in the analysis in all conditions, because we assumed that when participants recognized the idiom correctly in the Correct Condition (94.31 %), they had established memory traces of the correct idiom form. These inclusion criteria resulted in the following percentage of trials per condition: Correct: 79.4 %; Related: 81.3 %; Unrelated 1: 78 %; Unrelated 2: 78.7 %. For further analyses, we aggregated the conditions Unrelated 1 and Unrelated 2 into one condition Unrelated by averaging the mean voltages of the two conditions for each participant. Following this process, the final three conditions discussed in the analyses were: Correct, Related, and Unrelated.

Based on visual inspection of ERP results, we chose six Regions of Interest (ROIs), covering lateral and midline anterior and posterior sites (see Figure 14). Both lateral anterior ROIs included six electrode positions over both temporal cortices (left: F9, F7, FT9, FT7, FC5, T7; right: F10, F8, FT10, FT8, FC6, T8). The anterior midline ROI covered six fronto-central electrodes (F3, Fz, F4, FC1, FCz, FC2). Both lateral posterior ROIs included six temporo-parietal electrode positions (left: TP9, TP7, CP5, P7, PO9, O1; right: CP6, TP8, TP10, P8, PO10, O2). The posterior midline ROI covered six centro-parietal electrode positions (CP1, CP2, P3, Pz, P4, POz).

For statistical analysis, we conducted a $3 \times 3 \times 2$ repeated-measures ANOVA (RM-ANOVA) with the within-participant factors Condition (Correct, Related, Unrelated), Hemisphere (Left, Central, Right), and Region (Anterior, Posterior). First, we conducted RM-ANOVAs for each 100 ms time window. We identified three relevant time windows, which showed three-way interactions for Condition, Region, and Hemisphere (see Table 4): 100–200 ms, 300–500 ms, and 700–1000 ms. Both later time windows approximately align with the effects obtained in Rommers and colleagues (2013), with the 300–500 ms time window reflecting an N400 effect, and the 700–1000 ms time window reflecting a late positivity. The early time window does not find a parallel in previous ERP work on idiom processing. We label it as ‘pre-N400’ throughout the results section. For further analysis, we aggregated amplitudes across these time windows.

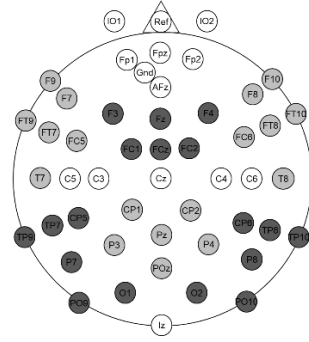


Abbildung 14: Electrode configuration used in the experiment. Anterior-Left, Anterior-Right, and Posterior-Central ROIs are highlighted in light gray. Anterior-Central, Posterior-Left, and Posterior-Right ROIs are highlighted in dark grey.

Tabelle 4: RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions.

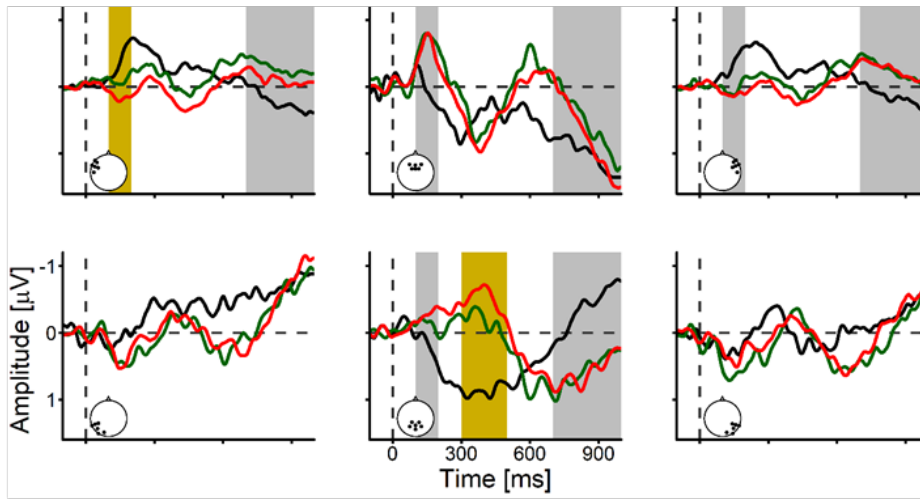
	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms	800–900 ms	900–1000 ms
C	*								
CxR			*	*		*	*	*	*
CxH	*	*	*	*	*	*		*	*
CxRxH	*		*	*			*	*	*

Results

Figure 15 (Panel A) depicts Grand-Average ERPs aggregated over Regions of Interest. Visual inspection of grand-averaged ERPs justified the selected time windows. As shown in the difference topographies (Figure 15, Panel B), the effect is most prominent over posterior sites. Moreover, a late positivity was observable over posterior sites.

RM-ANOVAs revealed significant three-way interactions for 100–200 ms ($F(4, 156) = 3.04$, $p = .03$), 300–500 ms ($F(4, 156) = 12.62$, $p < .0001$), and 700–1000 ms ($F(4, 156) = 4.27$, $p = .004$). All reported p-values are Greenhouse-Geisser or Bonferroni (for post-hoc t-tests) corrected.

Panel A



— Correct
 — Related
 — Unrelated
 no difference between Related and Unrelated
 difference between Related and Unrelated

Panel B

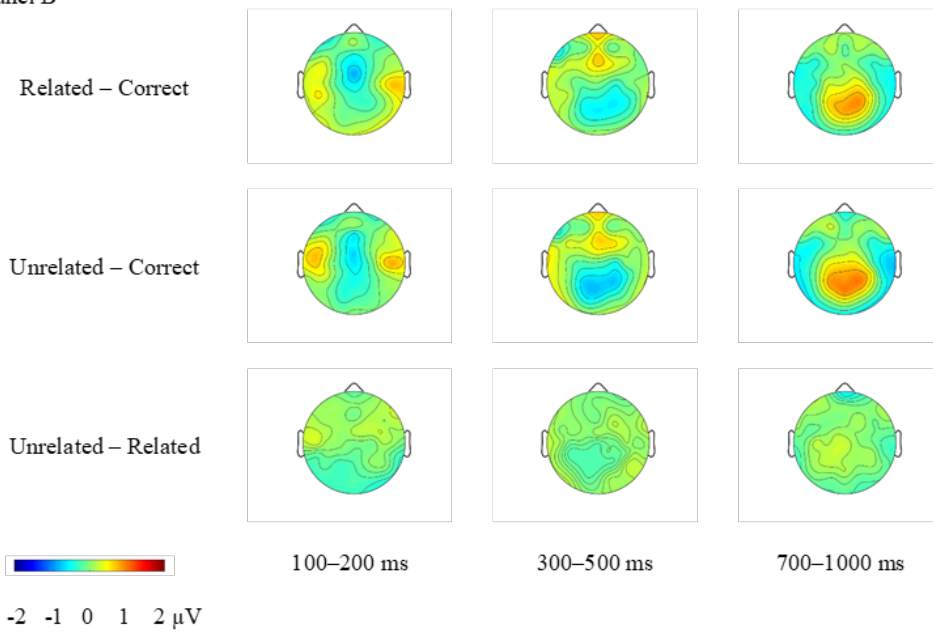


Abbildung 15: Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right. (B) Difference topographies for the time windows 100–200 ms, 300–500 ms, and 700–1000 ms.

100–200 ms time window (pre-N400) Post-hoc analyses of the three way interaction revealed a significant Condition effect for the Anterior-Left, Anterior-Central, Anterior-Right, and Posterior-Central ROIs (all $F(2, 78) \geq 7.18$, $p \leq .002$). Over the Anterior-Left sites only, all three conditions differed from each other: Correct vs. Related ($t1(39) = 2.93$, $p < .018$), Correct vs. Unrelated ($t2(39) = 6.21$, $p < .001$), and Related vs. Unrelated ($t3(39) = 3.09$, $p < .018$). Over the remaining three sites, we found differences between the Correct Condition vs. the Related Condition (all $t2(39) \geq |3.71|$, all $p \leq .002$), and for the Correct Condition vs. the Unrelated Condition (all $t3(39) \leq |1.04|$, all $p \geq .91$). In sum, we found parallel effects of semantic activation and no semantic activation.

300–500 ms time window (N400) For the 300–500 ms time window, a Condition effect was only evident over Posterior-Central sites ($F(2, 78) = 38.17$, $p < .0001$). Bonferroni-corrected post-hoc tests revealed significant differences between all three conditions: Correct vs. Related ($t1(39) = 5.91$, $p < .001$), Correct vs. Unrelated ($t2(39) = 7.43$, $p < .001$), and Related vs. Unrelated ($t3(39) = 2.71$, $p < .03$). Across Posterior-Central electrodes, amplitudes for the Unrelated Condition were more negative than those for the Related Condition, and amplitudes for the Correct Condition were most positive. Together, we found graded condition effects for a Posterior-Central electrode cluster typically associated with the N400.

700–1000 ms time window (late positivity) For the 700–1000 ms time window, we report those ROIs where a condition effect was significant ($F(2, 78) > 8.12$, $p \leq .002$). Post-hoc tests for these regions revealed differences of Related and Unrelated Conditions with Correct Conditions, but not between Related vs. Unrelated. Over Left-Anterior sites, amplitudes for the Correct Condition were more positive than for the Related Condition ($t1(39) = 3.94$, $p < .001$) and the Unrelated Condition ($t2(39) = 3.35$, $p < .006$), but amplitudes for the Related Condition and the Unrelated Condition did not differ significantly ($t3(39) = 1.7$, $p = .294$). Similarly, over Right-Anterior sites, amplitudes for the Correct Condition were more positive than for the Related Condition ($t1(39) = 2.87$, $p < .020$) and the Unrelated Condition ($t2(39) = 3.58$, $p < .003$), while amplitudes for the Related Completion and the Unrelated Condition did not differ significantly ($t3(39) = 0.05$, $p = 1$). Over Posterior-Central sites, amplitudes for the Correct Condition were more negative than for the Related Condition ($t1(39) = 5.55$, $p < .001$) and the Unrelated Condition ($t2(39) = 5.89$, $p < .001$), but amplitudes for the Related Condition and Unrelated Condition did not differ significantly ($t3(39) = 0.28$, $p = 1$). In sum, late ERPs show that related and unrelated violations of the idiom yield comparable amplitudes of a late positivity with posterior distribution (and reversed amplitudes over anterior regions).

Discussion

Using a semantic expectancy ERP paradigm in Experiment 2, we investigated processing mechanisms in highly predictive spoken idiomatic phrases. In contrast to Experiment 1, the critical idiom constituent itself did not appear in trials in which we probed the activation of

semantic associates of this idiom constituent. This way, we aimed to rule out potential bottom up spread from sensory input, which could have biased results in the visual world eye tracking design exploited in Experiment 1.

Across ERP amplitudes, there was a clear effect of expectancy of the correct idiom: both related and unrelated violations showed significantly higher ERP amplitudes than correct completions. This indicates that correct completions of an idiom were highly expected and easier to access than both related and unrelated substitutes. Because idioms were presented without biasing context, this broadly supports the notion that predictability within idioms mainly stems from the knowledge of the idiom form (Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010).

Using spoken idioms, N400 amplitudes reflected semantic expectancy within violation trials. That is, we not only obtained N400 reductions for correct completions, but also for semantic associates of correct completions. Since N400 reductions are interpreted in terms of facilitated semantic processing, including semantic expectancy mechanisms (for a review, see Kutas & Federmeier, 2011), it seems that the anticipation of the correct completion activated semantic associates, for which semantic processing was facilitated. In this sense, the N400 effect observed here is compatible with the eye tracking data from Experiment 1. It appears that single constituents and their individual meanings are available when these are predicted. These results in the auditory modality do not replicate those obtained for visually presented idioms obtained by Rommers and colleagues (2013), and challenge the conclusions drawn by these authors, who concluded that the top down prediction of idiom completions does not lead to beneficial processing of substitutes that are semantically related to idiom constituents.

The ERPs obtained in Experiment 2 mainly reflect an N400 effect followed by a late positivity. Recently, it has been discussed whether during the processing of idioms or other formulaic sequences the N400 is preceded by a P300 effect (Molinaro & Carreiras, 2010; Siyanova-Chanturia, Conklin, Caffarra, Kaan, & van Heuven, 2017; Vespignani et al., 2010). The authors of those studies found an enhanced P300 amplitude for correct and expected idiomatic forms compared to violations of those forms. They concluded that the P300 reflects a template matching process. Although we cannot rule out that the present N400 effect might also include an instance of the P300, we hypothesize that an activation of semantic information as found in Experiment 2 would only be detectable in the N400 component. We therefore conclude that the graded ERP effect between 300 and 500 ms in Experiment 2 are indeed an instance of the semantic N400 effect.

A late positivity between 700 and 1000 ms across posterior sites was independent of semantic relatedness, that is it did not show amplitude differences between related and unrelated violations. This effect converges with findings by Rommers and colleagues (2013), who

interpreted this late effect as a violation of the idiom as a lexical item. More recently, the late positivity following the semantic N400 (post-N400 positivity, PNP) in prediction paradigms has been interpreted as revision of a predicted sentence representation (Brothers, Swaab, & Traxler, 2015; Kuperberg & Wlotko, 2020) irrespective of the semantic relations between presented sentence-final words (Thornhill & Van Petten, 2012). Thus, in idiom processing the late positivity might also reflect that listeners revise the activated representation of the idiom string when hearing related or unrelated violations. Together with the N400 effect suggesting decomposition, the late positivity effect could be interpreted as evidence for a dual representation of idioms in the mental lexicon as both individual words and chunked items (Sprenger et al., 2006).

Similarly, early ERP effects obtained in the present study suggest that decomposition is not the only strategy followed by the parser. In contrast to the results from the written presentation of idioms by Rommers and colleagues (2013), we obtained ERP effects preceding the N400 in our study with spoken materials. This is comparable with previous findings (for a review, see Nieuwland, 2019). Already early on (between 100 and 200 ms), we see evidence for parallel processing. Across anterior-left electrodes, amplitudes for related conditions significantly differed from amplitudes for unrelated conditions. Across central and anterior-right sites, amplitudes for related and unrelated conditions did not differ. These early ERP effects might relate to parallel pre-activation of lexical representations (e.g., Friedrich & Kotz, 2007). If so, the present ERP results dissociate two types of lexical idiom representations: a form and a meaning representation of the single constituents. The former is indicated by the mid to right-anterior ERP deflections, while the latter is indicated by the left-lateralized ERP deflection. Thus, within familiar and highly predictable idioms, final constituents including their semantic properties can be pre-activated before they are fully processed (Smolka & Eulitz, 2020).

In general, the results of Experiment 2 corroborate other studies presenting idioms auditorily (Beck & Weber, 2016; Holsinger, 2013) by showing that listeners activate idiom constituents and have semantic associates of these constituents available. Possibly, the pre-N400 ERP effects and the graded N400 effect that we found might be due to modality-related differences compared to the study by Rommers and colleagues (2013). In contrast to the written and serial presentation (word-by-word) in that study, we presented idioms and violated idioms auditorily in Experiment 2. Semantic information might be accessible earlier in spoken language processing compared to written language processing. For example, preceding information speeds up spoken word identification even before enough acoustic information has accumulated (Van Petten et al., 1999). Therefore, we conducted a third experiment in which we used the same task and material as in Experiment 2, but presented them in the written modality.

Experiment 3

In Experiment 3, we conducted a semantic expectancy ERP experiment using the same material as in Experiment 2, but with written instead of spoken idioms. While experiments on

spoken idiom processing clearly point to decomposition within idioms (Experiment 1 & 2; Beck & Weber, 2016; Holsinger, 2013), the evidence from word-by-word presentations of idioms is mixed (Rabanus et al., 2008; Rommers et al., 2013; Smolka et al., 2007). Therefore, we aimed to address the question of processing differences across modalities in Experiment 3 by using the same idiomatic expressions and violations of these forms as in Experiment 2. If there are any prediction effects inherent to the idioms we used, we should not replicate results by Rommers and colleagues (2013).

Methods

Participants Thirty adults participated in Experiment 3, of whom we had to exclude data of five participants, due to incorrect instructions (3), a psychiatric disorder (1), and insufficient eye movement correction (1). This resulted in a sample of 25 participants for statistical analysis ($N = 25$, mean age = 21.4 years, range = 18–27, 18 female and 7 male). Participants were recruited at the University of Tübingen and received subject credits or payment as compensation. All participants included in the analysis were native, monolingual speakers of German, right-handed as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971), and had no history of a neurological, psychiatric or hearing disorder and normal or corrected-to-normal vision. None of the participants took part in Experiments 1 or 2. Prior to the experiment, participants gave written informed consent.

Stimuli In Experiment 3, we used the same stimuli as in Experiment 2, but these were presented visually at the center of a computer screen.

Procedure The procedure was the same as in Experiment 2 except for the presentation modality of the stimuli. We used the same timing of presentation as in the EEG study by Rommers and colleagues (2013). Each trial started with a fixation cross (+) for 1500 ms. Sentences were presented word-by-word with 300 ms presentation duration of a word and 300 ms blank screen. 900 ms after the presentation of the sentence-final word, a question mark (?) appeared on the screen, resulting in a 1200 ms delayed response after the onset of the target word. When the question mark appeared, participants were asked to decide whether the presented sentence was a correct idiom or not. They gave their answers via button press. The response hand was counterbalanced across participants.

Electrophysiological Recordings Same as in Experiment 2.

EEG Analysis As in Experiment 2, we included items that did not contain artifacts and to which participants responded correctly in the Correct Condition (93.8%) and after the onset of the response time window (1200 ms after stimulus onset). This resulted in the following percentage of included trials per condition for the analysis: Correct: 69.5%, Related: 72.3%, Unrelated 1: 72.4%, Unrelated 2: 71.5%. Compared to Experiment 2, the number of trials was

lower, because the EEG recordings were more artifactual.

Conducting RM-ANOVAs for 100 ms time window steps, we identified two relevant time windows, which showed three-way interactions for Condition, Region, and Hemisphere (see Table 5): 300–400 ms, and 500–700 ms. The first time window aligns with the early N400 time window found in Rommers et al. (2013). The later time window partly aligns with the time window for the late positivity in Rommers et al. (2013, 500–800 ms). For further analysis, we aggregated amplitudes across these time windows.

Table 5: RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions.

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms	800–900 ms	900–1000 ms
C		*	*	*	*	*			*
CxR				*	*	*	*		
CxH	*	*	*						*
CxRxH			*		*	*			

Results

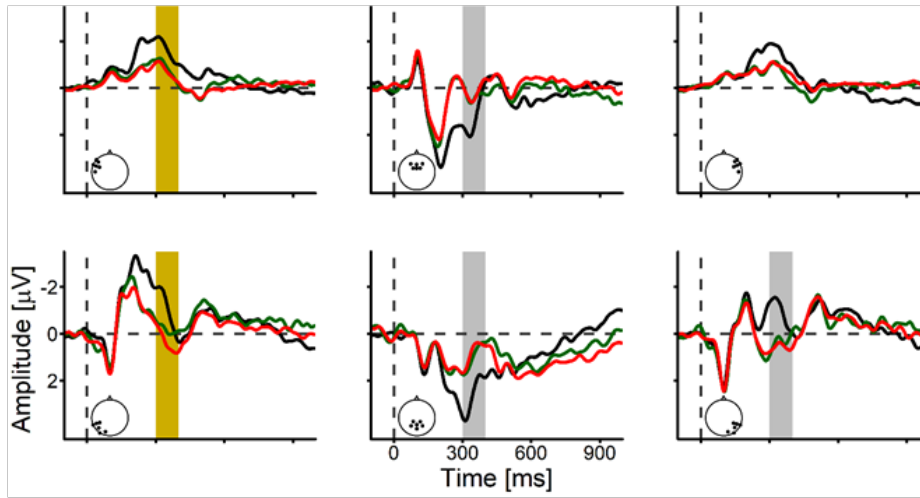
Figure 16 (Panel A) depicts Grand-Average ERPs aggregated over Regions of Interest. As shown in the difference topographies (Figure 16, Panel B), the N400 effect is most prominent over posterior sites. RM-ANOVAs revealed significant three-way interactions for 300–400 ms ($F(4, 96) = 5.31, p = .002$) and 500–700 ms ($F(4, 96) = 5.82, p < .001$). All reported p-values are Greenhouse-Geisser or Bonferroni (for post-hoc t-tests) corrected.

300–400 ms time window (N400) Across Left sites, we found a main effect for Condition ($F(2, 48) = 21.21, p < .001$), which was due to significant amplitude differences between Correct vs. Related ($t_1(24) = 3.60, p = .004$), Correct vs. Unrelated ($t_2(24) = 6.11, p < .001$), and Related vs. Unrelated ($t_3(24) = 3.01, p = .018$). Across Central sites, we also found a main effect for Condition ($F(2, 48) = 9.93, p < .001$). Post-hoc t-tests revealed amplitude differences between Correct vs. Related ($t_1(24) = 4.91, p < .001$) and Correct vs. Unrelated ($t_2(24) = 6.02, p < .001$) to be significant. Amplitudes for the Related and Unrelated Conditions did not differ ($t_3(24) = 1.08, p = .869$). Across Right sites, we found an interaction of Condition with Region

($F(2, 48) = 3.64, p = .04$). For Right-Anterior electrodes there was no effect of Condition ($F(2, 48) = 2.51, p = .11$), but for Right-Posterior electrodes the main effect for Condition was significant ($F(2, 48) = 13.24, p < .001$). Across the latter region, amplitudes between Correct vs. Related ($t1(24) = 3.66, p = .004$), and Correct vs. Unrelated ($t2(24) = 4.75, p < .001$) differed significantly. There was no amplitude difference between Related vs. Unrelated ($t3(24) = 0.29, p = 1$). Altogether, in the typical N400 time window and region we did not find evidence for a graded pattern of semantic expectancy.

500–700 ms time window (late positivity) For the 500–700 ms time window, we found a main effect of Condition across Left sites ($F(2, 48) = 5.41, p = .008$) and an interaction of Region and Condition across Central sites ($F(2, 48) = 6.00, p = .006$). Over Left electrodes, we found no amplitude differences between Correct vs. Related ($t1(24) = 0.70, p = 1$), significant amplitude differences between Correct vs. Unrelated ($t2(24) = 3.02, p = .018$), and marginally significant amplitude differences between Related vs. Unrelated ($t3(24) = 2.53, p = .055$). Across Central electrodes, there was only a Condition effect for Central-Anterior electrodes ($F(2, 48) = 4.92, p = .02$) with a significant amplitude difference between Correct vs. Unrelated ($t2(24) = 2.65, p = .042$) and no amplitude differences between Correct vs. Related ($t1(24) = 1.92, p = .199$) and Related vs. Unrelated ($t3(24) = 1.44, p = .489$). In sum, for the late effect the amplitude differences show a mixed pattern.

Panel A



— Correct
 — Related
 — Unrelated
 no difference between Related and Unrelated
 difference between Related and Unrelated

Panel B

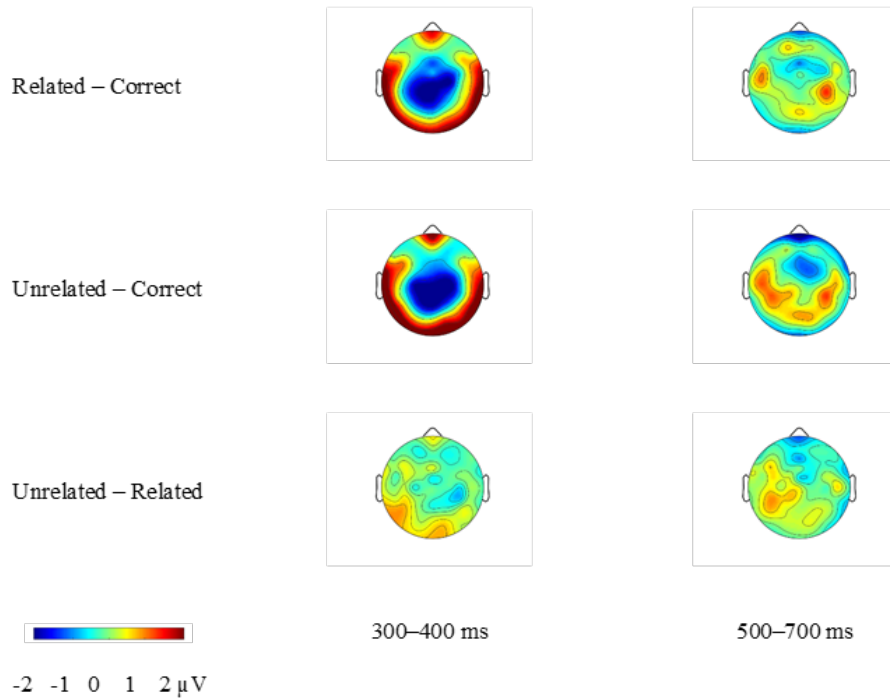


Abbildung 16: Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right. (B) Difference topographies for the time windows 300–400 ms and 500–700 ms.

Discussion

In Experiment 3, we again conducted a semantic expectancy ERP experiment to investigate top down spreading semantic activation within idioms. In contrast to Experiment 2, where we used the same material in a unimodal auditory paradigm, we presented idioms as written stimuli on the screen to further explore potential modality-related differences in processing. Therefore, the design was directly comparable to that of Rommers and colleagues (2013), who did not find activation of semantic associates of final constituents within written versions of idioms.

For written idioms, we found a clear expectancy effect on the N400: amplitudes for related and unrelated completions differed significantly from amplitudes for correct completions. This parallels findings for spoken idioms in Experiment 2. However, in contrast to Experiment 2, we did not find amplitude differences between related and unrelated targets in the typical semantic N400 region. Thus, we did not find an effect of semantic expectancy here. Based on the results by Rommers and colleagues (2013) and the present Experiment 2, these results might indicate that for online prediction of semantic features within idioms, the modality in which the idioms are presented might indeed play a role.

Nevertheless, we found differences between related and unrelated targets over left-hemispheric electrode leads in the N400 time window. Since there was no evidence of an N400 effect localized in this region either in Experiment 2, in the experiment by Rommers and colleagues (2013), or in the literature on the semantic N400 effect in non-idiomatic language an interpretation of this effect is difficult.

During a later time window, we did not replicate effects of a late positivity found in Experiment 2 and by Rommers and colleagues (2013). This effect was previously interpreted as indexing a violation of the holistic idiom representation. In Experiment 3, we only found consistent differences between correct and unrelated words. Amplitude differences between related and unrelated completions were mixed. Rommers and colleagues (2013) interpreted the late positive ERP effect as reflecting the difficulty to revise a predicted idiomatic multi-word representation. However, as we did not replicate such a late positive ERP effect with written idioms (Experiment 3), we are not confident about an interpretation at this point.

Like Rommers and colleagues (2013), we did not find a pre-N400 component for written idioms in an early time window. This suggests that the early component found in Experiment 2 was indeed specific to processing in the auditory modality (Connolly & Phillips, 1994; Connolly, Phillips, & Forbes, 1995).

Different ERP effects obtained for spoken idioms in Experiment 2 and written idioms in Experiment 3 challenge an alternative interpretation of activation effects for semantic associates

in our study. Even though some semantically related completions of the idioms we have presented might have a literally plausible interpretation, the results of Experiment 3 reveal that it is unlikely that the N400 is modulated by literal plausibility. If this were the case, we would also see an effect of semantic activation for written idioms in Experiment 3, because we used the same material for both modalities. Instead semantic activation was only observable for the processing of spoken idioms in Experiment 2. Moreover, Rommers and colleagues (2013) did not find N400 amplitude differences between related and unrelated completions although related completions were rated as more plausible than unrelated completions. For these reasons, we argue that literal plausibility does not account for the effects of semantic co-activation on the N400 component found in Experiment 2. Furthermore, research on plausibility and predictability in literal language suggests that rather than the N400 component, a post-N400-positivity is affected by the plausibility of the interpretation (DeLong, Quante, & Kutas, 2014; Quante, Bölte, & Zwitserlood, 2018). In the present study, we found no amplitude reduction of the late positivity for related completions indicating effects of plausibility. Furthermore, in the idiom literature amplitude differences in the N400 were not associated with semantic integration processes (Canal, Pesciarelli, Vespignani, Molinaro, & Cacciari, 2017). Altogether, we hypothesize that the reduction of the N400 amplitude for the spoken idioms that we obtained in Experiment 2 results from a short-lived semantic activation of the final constituent.

General Discussion

In the present study, we aimed to shed light on previous contradictory evidence on the extent to which idioms are processed holistically or decomposed into single items. Indirect evidence for holistic processing stems from studies showing faster processing for idioms compared to novel phrases (e.g., Conklin & Schmitt, 2008; Swinney & Cutler, 1979; Tabossi et al., 2009). Here, we tested idiom processing more directly by measuring their possible decomposition by means of semantic activation of individual idiom components (see e.g., Siyanova-Chanturia, 2015). Previous research demonstrated that semantic features of idiom constituents are available at least for priming processes in reading and listening (Beck & Weber, 2016; Holsinger, 2013; Smolka et al., 2007). However, when focusing on prediction mechanisms in reading, evidence for decomposition in idiom processing was lacking (Rommers et al., 2013). To rule out design and modality-related differences, we measured the level of semantic expectancy during online processing of highly predictive, spoken idioms in an eye tracking paradigm with written words (Experiment 1) as well as in a semantic expectancy ERP paradigm with spoken (Experiment 2) and written (Experiment 3) idioms.

Across all three experiments, we found evidence that participants built up an expectation of the idiom-final word. They fixated the correct idiom completion well before the idiom fragments presented in Experiment 1 ended, and they showed reduced N400 amplitudes for correct idiom completions compared to unrelated words in Experiments 2 and 3. Based on this, we conclude that idioms and their conventionalized forms can be recognized and activated

before their offset (Libben & Titone, 2008; Smolka & Eulitz, 2020; Vespignani et al., 2010). Together, these findings are evidence for multi-word representation of idioms. It appears that the mental lexicon stores information about the co-occurrence of specific words making up an individual idiom. Activation of respective multi-word representations triggers expectation of individual words that are part of these multi-word expressions. Here, we do not preclude a certain flexibility of these multi-word representations, but propose rather a strong coherence between the words of which they are composed (Cacciari & Tabossi, 1988; Geeraert, Baayen, & Newman, 2017; Kyriacou et al., 2020; Mancuso et al., 2020).

Across eye tracking and ERP methods with spoken idioms, we found evidence for early, short-lived semantic activation of individual idiom constituents. As soon as participants fixated correct idiom completions, they also fixated respective semantic associates (Experiment 1). In the ERPs, semantic associates of correct idiom completions elicited effects in the same early time window in which correct completions elicited effects (Experiment 2). Since we found anticipation of correct idiom completions in the fixation data, we conclude that early effects for semantic associates of idiom completions in the ERPs indeed relate to pre-activation of idiom constituents. Based on knowledge of conventionalized idiom forms, parsers seem to pre-activate a multi-word representation before the respective idiom is completely available in the auditory signal and this pre-activation includes single word representations that spread semantic activation within the network. It appears that even though literal constituent meanings typically do not contribute to the understanding of the idiomatic meaning, their processing is still automatically carried out. This conclusion is comparable to the notion that semantic processing cannot be switched off, as for example Connolly, Stewart, and Phillips (1990) showed for spoken language processing. We speculate that this is similar to a Stroop-like effect (Stroop, 1935) where the literal meaning of the word is not informative, but is nevertheless activated (cf. Glucksberg, 1993; McGlone, Glucksberg, & Cacciari, 1994).

It appears that semantic activation of constituent words within spoken idioms rapidly declines over time, as proposed for automatic spreading activation within the semantic network (e.g., Neely, O'Connor, & Calabrese, 2010). Neither fixation data nor ERPs gave evidence for long-lasting semantic activation of idiom constituents. Across Experiments 1 and 2, there was no longer a processing benefit for semantic associates compared to unrelated words after respective initial effects. Within spoken idioms, the present effect is comparable to that obtained by Sprenger and colleagues (2006), where semantic activation appeared to be strongest during early processing stages.

Here, we tentatively speculate that a rapid decay of semantic activation of constituent associates accounts for the presently and previously found mixed results for spoken and written idioms. Across paradigms using spoken idioms (Experiments 1 and 2), we consistently found evidence for activation of semantic associates of idiom-final words. Using written versions of

the same idioms as in Experiments 1 and 2, we did not find effects of semantic activation in Experiment 3 and this replicates results that Rommers and colleagues (2013) obtained for written idioms (word-by-word presentation). If automatic semantic activation of the idiom constituent decays rapidly, the time between idiom recognition and measurement of the semantic activation is crucial for observing respective effects. In general, it takes more time to present an idiom visually word-by-word (e.g., Experiment 3 of the present study or Rommers et al., 2013) than it takes to present a spoken version of the same idiom (e.g., Experiments 1 and 2). According to this timing difference, short-lived semantic activation might be still measurable at final constituents of spoken idioms, while it might have decayed already before the measurement in word-by-word reading (Rommers et al., 2013; Experiment 3 of the present study).

For priming experiments, where semantic spread presumably occurs in a bottom-up fashion, activation of semantic associates of idiom constituents was found for both modalities (Beck & Weber, 2016; Rabanus et al., 2008; Smolka et al., 2007). Since in those experiments the idiom constituent itself was always presented in the input, the recognition of the idiom and resulting pre-activation of its constituents is not the only source of spreading semantic activation. This led us to conclude that there is an interplay of the processing mechanism (top down vs. bottom up) and the modality-related rate of presentation. In addition, the results imply that even the top down prediction of idiom-final words is sufficient to activate single word meanings, but this is only measurable in the auditory modality in the present experiment. More research is needed to dissociate differences in these processes directly and to validate this explanation.

To account for individual idiom knowledge, we conducted an overt idiom recognition task in all experiments. In Experiment 1, participants had to choose the correct idiom completion among four alternatives. In Experiments 2 and 3, participants had to indicate whether the spoken or written strings were idioms. By performing these tasks, the participants might have been biased to activate canonical idiom forms only. However, if the participants would only have compared the incoming input with the activated idiom form, we should not have obtained a semantic activation of single word meanings in Experiments 1 and 2. In any case, general effects of the task cannot explain the differences between the results regarding activation of associates of idiom constituents of Experiments 2 and 3. Using the same task in both experiments, we show modality-related differences in online processing of idiomatic expressions.

The present results challenge the assumption that idioms are solely unanalyzed ‘long words’ (Jackendoff, 2002). In general, our results support hybrid models such as the Superlemma Hypothesis (Sprenger et al., 2006), in which idioms are represented as both multi-word representations (*superlemmas*) and simple lemmas of single constituents on a lexical level. The hybrid nature of idioms may allow the linguistic system to rely on single constituent and multi-word representations in parallel (Arnon & Christiansen, 2017; Tremblay & Baayen, 2010). We

hypothesize that the meanings of simple lemmas within idioms are available for only a short time after their activation.

Acknowledgement

We would like to thank Anne Bauch, Sara Beck, Stacie Boswell, Birte Herter, Babette Jakobi, Sören Koch, Tobias Kopp, Matteo Marks, Anne Rau, Ulrike Schild, and Charlotte Veil. We also warmly thank all participants. Furthermore, we would like to thank two anonymous reviewers for their helpful comments on a previous version of the manuscript. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 75650358 – SFB 833. The founding source had no involvement in the study.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502-518.
- Anderson, J. E., & Holcomb, P. J. (1995). Auditory and visual semantic priming using different stimulus onset asynchronies: An event-related brain potential study. *Psychophysiology*, *32*(2), 177-190.
- Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, *9*(3), 621-636.
- Beck, S. D., & Weber, A. (2016). Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, *7*, 1305.
- Berg, P., & Scherg, M. (1994). A multiple source approach to the correction of eye artifacts. *Electroencephalography and Clinical Neurophysiology*, *90*(3), 229-241.
- Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition*, *136*, 135-149.
- Cacciari, C., & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idioms processing: a reading-time study. *Journal of Cognitive Psychology*, *27*(7), 797-811.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*(6), 668-683.
- Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 928-943.
- Canal, P., Vespignani, F., Molinaro, N., & Cacciari, C. (2010). Anticipatory mechanisms in idiom comprehension: Psycholinguistic and electrophysiological evidence. In *Neuropsychology of Communication* (pp. 131-144). Milano: Springer.

- Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, 63(1), 95-122.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61.
- Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6(3), 256-266.
- Connolly, J. F., Phillips, N. A., & Forbes, K. A. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 94(4), 276-287.
- Connolly, J. F., Phillips, N. A., Stewart, S. H., & Brake, W. G. (1992). Event-related potential sensitivity to acoustic and semantic properties of terminal words in sentences. *Brain and Language*, 43(1), 1-18.
- Connolly, J. F., Stewart, S. H., & Phillips, N. A. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and Language*, 39(2), 302-318.
- DeLong, K. A., Quante, L., & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61, 150-162.
- Diaz, M. T., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, 1146, 85-100.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41(4), 469-495.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology*, 39(2), 133-146.
- Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, 69(3), 165-182.
- Friedrich, C. K., & Kotz, S. A. (2007). ERP evidence of form and meaning coding during online speech recognition. *Journal of Cognitive Neuroscience*, 19(4), 594-604.
- Geeraert, K., Baayen, R. H., & Newman, J. (2017). Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science*, 9(3), 653-669.
- Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, 8(2), 149-156.
- Glucksberg, S. (1993). Idiom meanings and allusional content. In C. T. Cacciari, P. (Ed.), *Idioms: Processing, Structure, and Interpretation* (pp. 3-26). Hillsdale, NJ: Erlbaum.

- Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930-944.
- Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: the P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*(11), 1531-1549.
- Heil, M., Rolke, B., & Pecchinenda, A. (2004). Automatic semantic activation is no myth: Semantic context effects on the N400 in the letter-search task in the absence of response time effects. *Psychological Science*, *15*(12), 852-857.
- Holsinger, E. (2013). Representing idioms: syntactic and contextual effects on idiom processing. *Language and Speech*, *56*(3), 373-394.
- Huettig, F., & McQueen, J. M. (2011). The nature of the visual environment induces implicit biases during language-mediated visual search. *Memory & Cognition*, *39*(6), 1068-1084.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171.
- Ito, A., & Husband, E. M. (2017). How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study. doi:10.13140/RG.2.2.33577.49765
- Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133-156.
- Kuperberg, G., & Wlotko, E. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, *32*(1), 12-35.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*(1), 621-647.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163.
- Kyriacou, M., Conklin, K., & Thompson, D. (2020). Passivizability of Idioms: Has the Wrong Tree Been Barked Up? *Language and Speech*, *63*(2), 404 –435.
- Laszlo, S., & Federmeier, K. D. (2009). A beautiful day in the neighborhood: An event-related potential study of lexical relationships and prediction in context. *Journal of Memory and Language*, *61*(3), 326-338.
- Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, *36*(6), 1103-1121.
- Mancuso, A., Elia, A., Laudanna, A., & Vietri, S. (2020). The Role of Syntactic Variability and Literal Interpretation Plausibility in Idiom Comprehension. *Journal of Psycholinguistic Research*, *49*(1), 99-124.

- Marantz, A. (2005). Generative linguistics within the cognitive neuroscience of language. *The Linguistic Review*, 22(2-4), 429-445.
- McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, 17(2), 167-190.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, 59(4), 475-494.
- Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, 83(3), 176-190.
- Neely, J. H., O'Connor, P. A., & Calabrese, G. (2010). Fast trial pacing in a lexical decision task reveals a decay of automatic semantic activation. *Acta Psychologica*, 133(2), 127-136.
- Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience and Biobehavioral Reviews*, 96, 367-400.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, 9(1), 97-113.
- Qualls, C. D., Treaster, B., Blood, G. W., & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: a reaction time study. *Journal of Communication Disorders*, 36(4), 245-261.
- Quante, L., Bölte, J., & Zwitserlood, P. (2018). Dissociating predictability, plausibility and possibility of sentence continuations in reading: evidence from late-positivity ERPs. *PeerJ*, 6, e5717.
- Rabanus, S., Smolka, E., Streb, J., & Rösler, F. (2008). Die mentale Verarbeitung von Verben in idiomatischen Konstruktionen. *Zeitschrift für Germanistische Linguistik*, 36(1), 27-47.
- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762-776.
- Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, 57(8), 1024-1029.
- Siyanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11(2), 285-301.
- Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, 175, 111-122.
- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251-272.

- Smolka, E., & Eulitz, C. (2020). Can you reach for the planets or grasp at the stars?—Modified noun, verb, or preposition constituents in idiom processing. In S. Schulte im Walde & E. Smolka (Eds.), *The Role of Constituents in Multiword Expressions: An Interdisciplinary, Cross-lingual Perspective* (pp. 179-204). Berlin: Language Science Press.
- Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing verbs in German idioms: Evidence against the Configuration Hypothesis. *Metaphor and Symbol, 22*(3), 213-231.
- Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. *Frequency Effects in Language* (pp. 127-163).
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language, 54*(2), 161-184.
- Strandburg, R., Marsh, J., Brown, W., Asarnow, R., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics: linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia, 31*(5), 412-434.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology: Human Perception and Performance, 18*(6), 643-662.
- Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior, 18*(5), 523-534.
- Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition, 37*(4), 529-540.
- Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *International Journal of Psychophysiology, 83*(3), 382-392.
- Titone, D. A., Lovseth, K., Kasparian, K., & Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? *Canadian Journal of Experimental Psychology, 73*(4), 216–230.
- Tremblay, A., & Baayen, R. H. (2010). Holistic Processing of Regular Four-word Sequences: A Behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151-173). London: Continuum.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning, 61*(2), 569-613.
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 155-172). Amsterdam: John Benjamins.
- Van Den Brink, D., Brown, C. M., & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience, 13*(7), 967-985.

- van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, *23*(1), 131-147.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E., & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(2), 394-417.
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682-1700.
- Wray, A. (2005). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

4.3 Schlussfolgerungen

Die Zielsetzung von **Studie I** war es, die Online-Verarbeitungsmechanismen von Idiomen im Erwachsenenalter zu untersuchen, um daraus Rückschlüsse auf deren Repräsentation im mentalen Lexikon zu ziehen. Diesbezüglich untersuchten wir (1) ob Versuchspersonen während der laufenden Sprachverarbeitung Erwartungen über das idiom-finale Wort aufbauen und (2) ob für dieses erwartete Wort dessen Bedeutung aktiviert wird, obwohl dies für das Verstehen der figurativen Bedeutung nicht notwendig ist. Zu diesem Zweck führten wir eine semantische Erwartungsaufgabe in Eye-Tracking- und EKP-Experimenten durch.

Bezüglich der Frage(1) fanden wir in den drei durchgeführten Experimenten starke Belege für einen Erwartungsaufbau des idiom-finalen Wortes, die sich in Fixations-, N400- und PNP-Amplitudenunterschieden zwischen korrekten und inkorrekten Vervollständigungen des Idioms manifestierten. Die Ergebnisse des Eye-Tracking-Experimentes belegen, dass Fixationsunterschiede zwischen korrekten und inkorrekten Vervollständigungen bereits auftreten, bevor das Idiomfragment im auditiven Input komplett gehört wurde. Daher kann geschlossen werden, dass aufgrund der Bekanntheit mit dem Idiom dieses bereits innerhalb des Idiomfragmentes erkannt und eine starke Erwartung für das letzte Wort aufgebaut wird. Dies bestätigt bisherige Befunde, die darauf hindeuten, dass Idiome bereits vor dem Ende erkannt werden können (Cacciari & Tabossi, 1988; Molinaro & Carreiras, 2010; Siyanova-Chanturia et al., 2017; Tabossi et al., 2005; Titone et al., 2019; Vespignani et al., 2010). Zusätzlich fanden wir für die PNP für die Verarbeitung gesprochener Idiome – ähnlich zu den Ergebnissen von Rommers, Dijkstra und Bastiaansen (2013) mit geschriebener Sprache – keine Amplitudenunterschiede zwischen relativen und unrelativen Vervollständigungen. Dies könnte als Verletzung der abgespeicherten Mehrwort-Repräsentation interpretiert werden. Allerdings ließ sich ein konsistenter Effekt der PNP nicht für die Verarbeitung geschriebener Idiome replizieren. Daher ist eine eindeutige Interpretation zu diesem Zeitpunkt nicht möglich. Insgesamt verdeutlichen die Befunde jedoch, dass Erwachsene sensitiv gegenüber der Wort-Konfiguration von Idiomen sind, weshalb eine Mehrwort-Repräsentation wahrscheinlich ist.

Frage(2) zielte darauf ab, widersprüchliche Befunden in der Literatur bezüglich der Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomen zu untersuchen (Diskussion siehe Kapitel 3.3). Insbesondere standen hierbei methoden-relatierte Unterschiede im Fokus. Deshalb führten wir die Experimente mit verschiedenen Methoden, verschiedenen Modalitäten und damit verbundenen Präsentationsarten durch: auditiv präsentierter Wortstrom in Eye-Tracking- und EKP-Experimenten und visuell präsentierte Wort-für-Wort-Folge in einem EKP-Experiment. Wir fanden Hinweise für die Aktivierung der Einzelwort-Bedeutung des letzten, erwarteten Wortes während der Verarbeitung gesprochener Idiome, jedoch nicht während der Verarbeitung geschriebener Idiome. In dem EKP-Experiment mit geschriebener Sprache replizierten wir die Ergebnisse von Rommers, Dijkstra und Bastiaansen (2013), die für die Verarbeitung

geschriebener Sprache ebenfalls keine Unterschiede zwischen relatierten und unrelatierten Vervollständigungen fanden. Da in der vorliegenden Studie das gleiche linguistische Material für die beiden EKP-Experimente mit gesprochener und geschriebener Sprache verwendet wurde, lassen sich die Unterschiede zwischen den Experimenten lediglich auf die verwendete Modalität zurückführen. Daher ist es auch unwahrscheinlich, dass der verwendete Kontext (minimal, neutral vs. figurativ) oder die Verarbeitungsmechanismen (Bottom-Up vs. Top-Down) widersprüchliche Befunde zwischen der Studie von Rommers, Dijkstra und Bastiaansen (2013) und anderen Studien (Beck & Weber, 2016a; Holsinger, 2013; Smolka et al., 2007; Sprenger et al., 2006; van Ginkel & Dijkstra, 2019) erklärt.

Die Eye-Tracking- und EKP-Daten für gesprochene Idiome deuten darauf hin, dass Aktivierung der Einzelwort-Bedeutung zeitlich relativ früh, aber nur kurzzeitig auftritt. Im Eye-Tracking-Experiment zeigten sich Fixationsunterschiede zwischen relatierten und unrelatierten Wörtern auf dem Bildschirm zeitgleich mit der Antizipation des korrekten idiom-finalen Wortes. Fixationsunterschiede zwischen relatierten und unrelatierten Wörtern war aber nur für eine kurze Dauer vorhanden. Deshalb deuten die Eye-Tracking-Ergebnisse darauf hin, dass mit dem Erkennen des Idioms das letzte Wort und zeitgleich dessen semantische Merkmale kurzfristig aktiviert werden. In dem EKP-Experiment mit gesprochener Sprache fanden wir ebenfalls Amplitudenunterschiede zwischen relatierten und unrelatierten Vervollständigungen nur in einer frühen Komponente, die spezifisch für die Verarbeitung auditiv präsentierter Wörter ist (vgl. Kapitel 2.2), und in der N400-Komponente, nicht aber für die späte PNP-Komponente. Da das korrekte idiom-finale Wort nicht im tatsächlichen Input in den inkorrekten Bedingungen präsentiert wurde, kann davon ausgegangen werden, dass die Aktivierung der Einzelwort-Bedeutung ausschließlich auf der Erwartung des letzten Wortes aufbaut. Somit bieten die Ergebnisse in der auditiven Modalität Evidenz für eine frühe, aber kurzweilige semantische Aktivierung des erwarteten, idiomatisch korrekten Wortes. Diese Befunde sind vergleichbar mit denen von Sprenger und Kolleg:innen (2006), die semantische Effekte vor allem während der zeitlich frühen Darbietung von Primewörtern gefunden haben.

Die kurzzeitige Aktivierung der semantischen Information könnte die Messbarkeit dieser beeinflussen. Deshalb könnte es zum einen kritisch sein, wann die Einzelwort-Bedeutung aktiviert wird und zum anderen zu welchem Zeitpunkt diese Aktivierung gemessen wird. Falls die semantische Aktivierung zum Zeitpunkt der Messung bereits abgeflacht ist, könnte sie nicht mehr messbar sein (siehe Abbildung 17). Dies könnte widersprüchliche Befunde zwischen einzelnen Studien erklären.

Die kurzzeitige Aktivierung der Einzelwort-Bedeutung könnte die fehlenden semantischen Effekte während der Verarbeitung geschriebener Idiome in **Studie I** und der Studie von Rommers, Dijkstra und Bastiaansen (2013) erklären. Bedingt durch die Wort-für-Wort-Präsentationsart verlangsamt sich die Präsentationsdauer der jeweiligen Sätze im Vergleich zu

gesprochener Sprache (vgl. Abbildung 10). Daher verlängert sich auch der Abstand zwischen dem Erkennen des Idioms, welches in Erwartungsexperimenten die einzige Quelle für die Vorhersage des Einzelwortes und dessen Bedeutung wäre, und der Messung der Einzelwort-Bedeutung. Während der Wort-für-Wort-Präsentation geschriebener Idiome könnte die Messung der Einzelwort-Bedeutung zu einem Zeitpunkt geschehen, zu dem deren Aktivierung bereits wieder abgeflacht und daher nicht mehr messbar ist (siehe Abbildung 17).

Zusätzlich könnte die Abflachung der semantischen Aktivierung auch die Unterschiede zwischen den Studien von Rommers, Dijkstra und Bastiaansen (2013) und Sprenger und Kolleg:innen (2006) erklären. Beide Studien untersuchten, ob während des Lesens von Idiomen top-down die Bedeutung des idiom-finalen Wortes aktiviert wird. Rommers und Kollegen präsentierten Idiome Wort-für-Wort und fanden keine semantische Aktivierung. In diesem Fall ist der zeitliche Abstand zwischen Erkennen des Idioms und Messung der Einzelwort-Bedeutung relativ groß. Bei einer abflachenden semantischen Aktivierung, könnte diese daher nicht messbar sein. Sprenger und Kollegen präsentierten das gesamte Idiomfragment zur gleichen Zeit, wodurch das Idiom relativ schnell nach der Präsentation erkannt werden könnte, und fanden eine semantische Aktivierung. Unter dieser Bedingung konnten die Autor:innen die Aktivierung der Einzelwort-Bedeutung jedoch auch nur früh nach der Präsentation des Idiomfragmentes messen.

Weiterhin könnte die Abflachung der semantischen Aktivierung die Unterschiede zwischen den Ergebnissen aus den Erwartungsexperimenten mit geschriebenen Idiomen in **Studie I** und von Rommers, Dijkstra und Bastiaansen (2013) und aus Priming-Experimenten (Beck & Weber, 2016a; Holsinger, 2013; Smolka et al., 2007; van Ginkel & Dijkstra, 2019) erklären. Je nach Paradigma kann die Aktivierung der Einzelwort-Bedeutung bottom-up aufgrund der Präsentation oder top-down aufgrund der Erwartung erfolgen (siehe Abbildung 17; in dem Priming-Experiment von Smolka und Kolleg:innen (2007) wurden stark vorhersagbare Idiome verwendet, deshalb könnte in diesem Experiment die Aktivierung der Einzelwort-Bedeutung aufgrund der Erwartung und der Präsentation erfolgen). In Erwartungsparadigmen ist die top-down Erwartung des Idiom-Bestandteiles die einzige Quelle für eine Aktivierung dessen Bedeutung. In den Erwartungsexperimenten mit geschriebenen Idiomen in **Studie I** und von Rommers, Dijkstra und Bastiaansen (2013) wurden einzelne Wörter Wort-für-Wort präsentiert. Dadurch war der zeitliche Abstand zwischen der Aktivierung der Einzelwort-Bedeutung und deren Messung zu groß, um sie messen zu können. Priming-Experimente finden konsistent positive Befunde für eine Aktivierung der Einzelwort-Bedeutung. In diesen Experimenten wurde die Aktivierung der Einzelwort-Bedeutung zeitlich relativ kurz nach der Präsentation des kritischen Idiom-Bestandteiles gemessen. Zu diesem Zeitpunkt war die semantische Aktivierung vermutlich noch relativ stark. Direkte Vergleiche zwischen Priming- und Erwartungsexperimenten in zukünftigen Studien könnten diese Vermutung untersuchen.

Kurze Abstände zwischen Bottom-Up Aktivierung der Einzelwort-Bedeutung und Messung könnte auch erklären, dass positive Evidenz unabhängig des Kontextes gefunden werden. Die Einbettung von Idiomen in einen figurativen Kontext kann dazu führen, dass diese früher erkannt werden (Fanari et al., 2010). Beck und Weber (2016b), Holsinger (2013) und Rommers, Dijkstra und Bastiaansen (2013) führten Experimente durch, in denen Idiome in einem Kontext präsentiert wurden, der sehr stark die figurative Interpretation bahnte (entweder Satz- oder experimenteller Kontext). Wie bereits beschrieben ist es unwahrscheinlich, dass der Kontext in der Studie von Rommers, Dijkstra und Bastiaansen (2013) fehlende Effekte einer Einzelwortaktivierung ausschließlich erklärt. Diese Effekte stammen mit einer größeren Wahrscheinlichkeit von einer erschwerten Messbarkeit aufgrund des großen zeitlichen Abstandes zwischen Erkennen des Idioms und Messung der Einzel-Wort-Bedeutung. Nichtsdestotrotz könnte der Kontext das Erkennen des Idioms beschleunigen. In Experimenten, in denen der zeitliche Abstand zwischen Präsentation des kritischen Idiom-Bestandteiles und der Messung dessen Aktivierung klein war, fanden positive Effekte trotz bahndem Kontext (Beck & Weber, 2016b; Holsinger, 2013). Die Hypothese, dass die Messbarkeit der Einzelwortaktivierung im zeitlichen Verlauf rapide abnimmt, könnte diese Unterschiede zwischen den Studien erklären.

Eine Limitation der **Studie I** im Hinblick auf die beschriebene Hypothese zur automatischen Aktivierung der Einzelwort-Bedeutung ist, dass wir den Zeitpunkt, ab dem die jeweiligen Idiome als solche erkannt werden (Idiom Key, Cacciari & Tabossi, 1988, siehe Kapitel 1.1), nicht direkt gemessen haben. Der Idiom Key wird typischerweise in offline Cloze-Probability-Aufgaben gemessen. Der Punkt, an dem ein Satzfragment idiomatisch vervollständigt wird, kennzeichnet dessen Idiom Key. In vorhersagebaren bzw. langen idiomatischen Ausdrücken können diese schon vor dem Präsentationsende als solche erkannt werden (Fanari et al., 2010; Tabossi et al., 2005). In den vorliegenden Experimenten wurden lange Idiome mit einer hohen Cloze Probability des letzten Wortes verwendet. Dies erhöht die Wahrscheinlichkeit, dass die Idiome vor dem Präsentationsende erkannt wurden, was sich ebenfalls in den Fixationsdaten widerspiegelt. Allerdings ist die zeitliche Analyse der Fixationsdaten im Eye-Tracking-Experiment an das Präsentationsende des auditiven Signals ausgerichtet. Alternativ könnte man die zeitliche Analyse der Fixationsdaten an dem jeweiligen Idiom Key der Idiome ausrichten, um präziser für individuelle Idiome zu untersuchen, wie viel Zeit zwischen dem Zeitpunkt des Erkennens und der Messung der Einzelwort-Bedeutung vergeht.

Gesprochene Idiome können im Vergleich zu geschriebenen Idiomen zusätzlich prosodische Informationen enthalten, die Hinweise darauf geben können, dass nicht die wörtliche Interpretation der Phrase intendiert ist. Typischerweise unterscheidet sich der idiomatische Gebrauch eines Ausdrucks von dem wörtlichen Gebrauch desselben Ausdrucks hinsichtlich der phonetischen Länge. Dies äußert sich darin, dass der idiomatische Gebrauch schneller produziert wurde als der wörtliche (Sivanova-Chanturia & Lin, 2018). Diese phonetischen Unterschiede sorgen dafür, dass Muttersprachler:innen sehr genau die intendierte Bedeutung bestimmen

können (Van Lancker Siddis, 2003; Yang, Ahn & Van Lancker Siddis, 2010). Bisher wurde nicht untersucht, ob Idiome anhand dieser phonetischen Merkmale auch während der laufenden Sprachverarbeitung schneller erkannt werden können als geschriebene Varianten der gleichen Idiome. In der vorliegenden Studie kontrollierten wir phonetische Merkmale nicht und können deswegen keine konkreten Aussagen darüber treffen.

Insgesamt unterstützen die Ergebnisse von **Studie I** hybride Modelle, die sowohl die Repräsentation von MWE als Einheit als auch die Möglichkeit der Verarbeitung der Einzelwort-Bedeutungen postulieren (Arnon & Snider, 2010; Cacciari & Tabossi, 1988; Libben & Titone, 2008; Sprenger et al., 2006). Anhand der Ergebnisse von **Studie I** können keine Aussagen über die Flexibilität der Mehrwort-Repräsentation getroffen werden. Prinzipiell unterstützen die Daten das Superlemma-Modell von Sprenger und Kolleg:innen (2006). In diesem Modell sind Idiome sowohl als Einheit in sogenannten Superlemmas als auch als Einzel-Lemmas gespeichert. Die Studie von Sprenger und Kolleg:innen (2006) impliziert, dass die semantische Aktivierung der Einzel-Lemmas jedoch nur von kurzer Dauer ist. Da dieses Modell für die Produktion von Idiomen aufgestellt wurde, erhielt die zeitliche Komponente weniger Bedeutung. Die Ergebnisse von **Studie I** implizieren jedoch, dass der zeitliche Verlauf der Bedeutungsaktivierung für die Verarbeitung eine Rolle spielt. **Studie I** fokussierte insbesondere auf Unterschiede zwischen Hören und Wort-für-Wort-Lesen, welche den Zeitpunkt der Aktivierung der Einzelwort-Bedeutung relativ zum Erkennen des Idioms manipulieren könnten. Anknüpfend an das Superlemma-Modell (Sprenger et al., 2006), postulieren wir daher, dass Idiome als Mehrwort-Repräsentation abgespeichert und trotzdem die Einzelwort-Bedeutungen verfügbar sind. Allerdings sind letztere nur für eine kurze Zeit verfügbar, und deren Messbarkeit hängt von dem Zeitpunkt der Messung relativ zu deren Aktivierung ab. In den Abbildungen 17a und 17b ist der Zerfall der Aktivierung der Einzelwort-Bedeutung und dessen Messbarkeit zu unterschiedlichen Zeitpunkten in das Superlemma-Modell integriert. Je nachdem, ob die Aktivierung lediglich auf der Erwartung des Idiom-Bestandteiles aufbaut oder dieses direkt präsentiert wird, ist eine semantische Aktivierung messbar.

Weiterführend ist es wichtig zu untersuchen, inwiefern die Verschiebung des Zeitpunktes, an dem das Idiom als solches erkannt wird, einen Einfluss auf die Aktivierung der Einzelwort-Bedeutung hat. Dies könnte beispielsweise durch die Einbettung in einen figurativen Kontext oder bestimmte Idiom-Charakteristika realisiert werden. Demzufolge könnten die Ergebnisse aus **Studie I** auch das Konfigurationsmodell (Cacciari & Tabossi, 1988) und das multideterministische Modell (Libben & Titone, 2008) unterstützen. Beide Modelle gehen davon aus, dass Idiome sich im Zeitpunkt ab welchem diese erkannt werden unterscheiden und dies die Aktivierung der figurativen Bedeutung, die direkt mit der Konfiguration im mentalen Lexikon verbunden ist, beeinflusst.

Aus den vorliegenden Befunden von **Studie I** kann geschlussfolgert werden, (1) dass erwachsene Muttersprachler:innen Erwartung über Wörter innerhalb von MWE aufbauen. Dies unterstützt die Annahme von Mehrwort-Repräsentationen von MWE im mentalen Lexikon. (2) kann geschlussfolgert werden, dass erwachsene Muttersprachler:innen die Einzelwort-Bedeutungen relativ kurzzeitig aktivieren. Dies könnte deren Messbarkeit erschweren und unterschiedliche Ergebnisse zwischen Studien erklären. Anschließend an **Studie I** stellt sich die Frage, wie diese hybriden Mehrwort-Repräsentationen entstehen und sich im Entwicklungsverlauf verändern (**Studien II** und **III**).

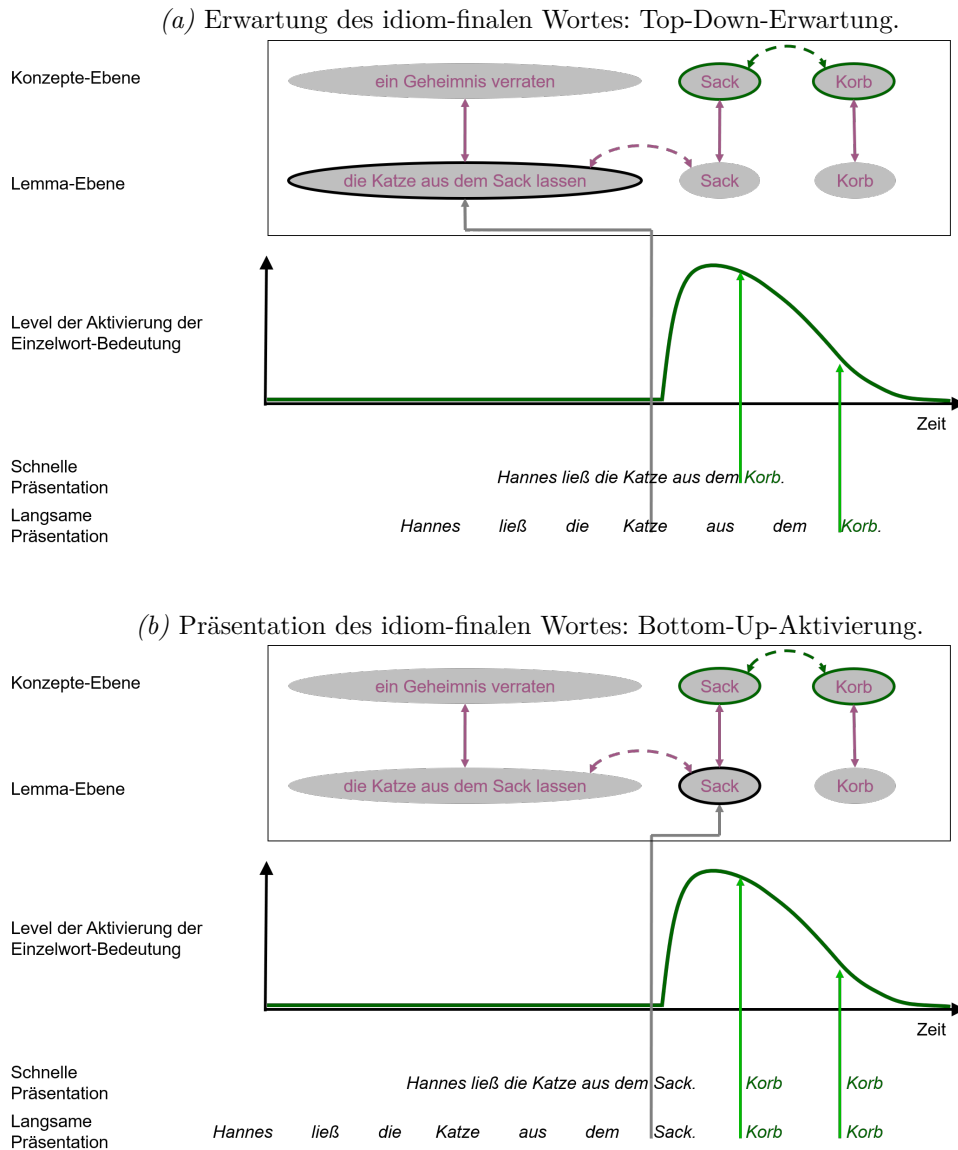


Abbildung 17: Schematische Darstellung der kurzzeitigen Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomem und deren Messbarkeit in Abhängigkeit von dem Verarbeitungsmechanismus ((a) und (b); additive Effekte von Erwartung und Präsentation sind nicht dargestellt), jeweils: **Oberer Teil:** mentale Repräsentation von Idiomem in Anlehnung an das Superlemma-Modell von Sprenger et al. (2006). Idiomem sind als Mehrwort-Repräsentationen gespeichert die Verbindungen zu den einzelnen Wörtern besitzen (durchgehende Pfeile = gemeinsame Elemente; gestrichelte Pfeile = Relationen). **Mittlerer Teil:** Zeitverlauf der Aktivierung der idiom-finalen Einzelwort-Bedeutung (dunkelgrüne Linie); Zeitpunkt der Aktivierung des idiom-finalen Wortes aufgrund (a) der Erwartung oder (b) der Präsentation (graue Pfeile); Zeitpunkt der Messung der Aktivierung der Einzelwort-Bedeutung (hellgrüne Pfeile). **Unterer Teil:** Schematische Darstellung unterschiedlich schneller Präsentationsarten (z.B. schnell: auditiver Wortstrom; langsam: Wort-für-Wort, visuelle Präsentation).

5 Verarbeitung und Repräsentation von Mehrworteinheiten im Erstspracherwerb

langsam Gestalt annehmen

Grundlegend für die Repräsentation von MWE ist, wie diese während des Spracherwerbs entstehen und sich verfestigen. Erwachsene Muttersprachler:innen scheinen MWE als Mehrwort-Repräsentation im mentalen Lexikon abzuspeichern. Gleichzeitig haben sie aber auch kurzzeitig Zugriff auf einzelne Bestandteile dieser Repräsentation (Kapitel 4.3). Laut gebrauchsbasierten Modellen (z.B. Bod, 2006, Diskussion siehe Kapitel 1) können sich die Repräsentationen sprachlicher Einheiten abhängig von dem Gebrauch und der Expositionshäufigkeit der jeweiligen Ausdrücke verändern.

Einige Untersuchungen legen nahe, dass zumindest das Wissen und die Bekanntheit formelhafter Ausdrücke mit der Expositionshäufigkeit steigt (Bonin et al., 2013; Sprenger et al., 2019; Tabossi et al., 2011). Das Idiomwissen oder -vokabular steigt mit dem Alter und dementsprechend auch mit der Erfahrung an (la Roi et al., 2019; Sprenger et al., 2019). Außerdem sind Idiome, die früh im Entwicklungsverlauf erlernt wurden, im Erwachsenenalter bekannter, als Idiome die später erlernt wurden (Bonin et al., 2013; Tabossi et al., 2011). Diese empirischen Befunde legen daher nahe, dass sich die Bekanntheit der jeweiligen MWE mit steigender Expositionshäufigkeit während des Spracherwerbs erhöht. Im Erwachsenenalter stehen sowohl die Häufigkeit als auch die Bekanntheit von MWE in Zusammenhang mit der Verarbeitungsgeschwindigkeit dieser Ausdrücke (z.B. Arnon & Cohen Priva, 2013; Carrol & Conklin, 2020, Diskussion siehe Kapitel 3.1). Daher stellt sich die Frage, ob mit steigender Expositionshäufigkeit und Bekanntheit im Entwicklungsverlauf auch die Verarbeitung und mentale Repräsentation von MWE verändert.

5.1 Erfahrungsbedingter Verarbeitungsvorteil für Mehrworteinheiten

Im Erwachsenenalter lassen sich Effekte von häufiger Exposition von MWE im bisherigen Sprachgebrauch auf die Verarbeitungsgeschwindigkeit finden (Arnon, McCauley & Christiansen, 2017; Bonin et al., 2013). Beispielsweise werden Idiome, die früh erworben wurden, schneller gelesen als solche, die spät erworben wurden (Bonin et al., 2013). Wie in Kapitel 3.1 beschrieben, verstärkt sich auch der Verarbeitungsvorteil von MWE gegenüber Nicht-MWE mit zunehmender Phrasenhäufigkeit (Arnon & Cohen Priva, 2014; Arnon & Snider, 2010; Kapatsinski & Radicke,

2009; Sosa & MacFarlane, 2002; Tremblay & Baayen, 2010). In einer Lernstudie konnten Conklin und Carrol (2020) außerdem zeigen, dass erwachsene Versuchspersonen bereits nach vier- bzw. fünfmaliger Exposition neuartige Binomiale (z.B. engl. *wires and pipes*, dt. *Kabel und Rohre*) bereits schneller verarbeiteten als umgekehrte, unbekannte Varianten (z.B. engl. *pipes and wires*, dt. *Rohre und Kabel*). Diese Befunde implizieren somit, dass die zunehmende Exposition häufig wiederkehrender Wort-Konfigurationen zu einer Festigung der Gedächtnisspuren dieser Wort-Konfigurationen führt. Dementsprechend verstärkt sich die Repräsentation von MWE im mentalen Lexikon.

In Bezug auf den Zweitspracherwerb konnten Siyanova-Chanturia, Conklin und van Heuven (2011) zeigen, dass der Verarbeitungsvorteil von Binomialen (engl. *bride and groom*, dt. *Bräutigam und Braut*) im Vergleich zu umgekehrten Varianten (engl. *groom and bride*, dt. *Bräutigam und Braut*) mit steigender Sprachkompetenz ansteigt. Die Autor:innen begründen dies damit, dass mit steigender Sprachkompetenz die Exposition mit formelhaften Ausdrücken ebenfalls steigt und diese daher schneller verarbeitet werden als umgekehrte Varianten. Da mit zunehmender Exposition sich auch die Gedächtnisspuren jeweiliger Ausdrücke stärken, können MWE von erfahreneren Zweitsprachler:innen schneller verarbeitet werden als von weniger erfahrenen Zweitsprachler:innen.

Auch im Hinblick auf die kindliche Sprachentwicklung konnte gezeigt werden, dass die Expositionshäufigkeit in Zusammenhang mit der Verarbeitungsgeschwindigkeit formelhafter Ausdrücke steht. Bannard und Matthews (2008) untersuchten, ob Kinder Ausdrücke, die häufig in der sprachlichen Umgebung von Kindern auftraten (engl. *sit in your chair*, dt. *setz dich auf deinen Stuhl*), schneller verarbeiten als seltene Ausdrücke (engl. *sit in your truck*, dt. *setz dich in deinen Lastwagen*). Dazu führten sie eine Satz-Wiederholungs-Studie mit zwei- bis dreijährigen Kindern durch. Die Kinder sollten die Ausdrücke wiederholen, die sie zuvor gehört hatten. Sie konnten häufige Ausdrücke schneller produzieren als seltene Ausdrücke. Eine ähnliche Sensitivität gegenüber der Häufigkeit von MWE konnte bei der Produktion von unregelmäßigen Pluralformen gezeigt werden (Arnon & Clark, 2011). Beispielsweise konnten Kinder unregelmäßige Pluralformen innerhalb eines Kontextes, in dem sie häufig auftreten (engl. *brush your teeth*, dt. *putz deine Zähne*), besser produzieren als in seltenen Konstellationen (engl. *brush my teeth*, dt. *putz meine Zähne*). Diese Befunde belegen, dass die Häufigkeit, mit der Kinder mit bestimmten Wort-Konfigurationen konfrontiert werden, einen positiven Zusammenhang mit der Verarbeitungsgeschwindigkeit aufweist. Dies könnte ebenfalls darauf hindeuten, dass mit steigender Exposition sich die Repräsentation der jeweiligen Ausdrücke im mentalen Lexikon verstärkt.

Im Kindesalter steht neben der Expositionshäufigkeit auch die Bekanntheit formelhafter Ausdrücke in Zusammenhang mit der Verarbeitungsgeschwindigkeit. Qualls et al. (2003) führten ein Experiment durch, in welchem Kinder der fünften Klasse verschiedene Sätze lasen und so

schnell wie möglich entscheiden sollten, ob es sich jeweils um ein Idiom handelte oder nicht. Kinder erkannten bekanntere Idiome schneller als weniger bekannte Idiome. Diese Effekte ähneln denen von Erwachsenen (vgl. Tabossi et al., 2009, siehe Kapitel 3.1) und deuten darauf hin, dass die jeweilige Bekanntheit mit den Ausdrücken ein Faktor bei der Verarbeitungsgeschwindigkeit von Idiomen ist. Da Expositionshäufigkeit und Bekanntheit eng miteinander verknüpft sind (vgl. Kapitel 1), kann man ihre Effekte auf die Verarbeitungsgeschwindigkeit nicht direkt voneinander trennen.

Schlussfolgernd kann man sagen, dass mit steigender Expositionshäufigkeit die Bekanntheit von MWE zunimmt und gleichsam die Geschwindigkeit abnimmt, mit der Kindern und Erwachsenen diese verarbeiten. Indirekt impliziert dies, dass, sobald Kinder individuelle MWE erworben haben, sie MWE als Mehrwort-Repräsentation abspeichern und diese sich mit steigender Häufigkeit und Bekanntheit verstärkt. Bisher gibt es keine direkte Evidenz dafür, dass Kinder anhand von MWE-Fragmenten weitere Bestandteile vorhersagen können. Diesen Aspekt betreffend untersuchen die **Studien II** und **III**, inwiefern Kinder unterschiedlicher Altersgruppen (zehn und dreizehn Jahre) finale Bestandteile innerhalb von bekannten Idiomen vorhersagen. Somit können Rückschlüsse auf die Art der Repräsentation von MWE im Spracherwerb gezogen werden.

5.2 Verarbeitung der Einzel-Bestandteile innerhalb von Mehrworteinheiten im Spracherwerb

Bisher liegen keine Studien vor, die die Rolle der Einzelwort-Bestandteile bei Kindern beziehungsweise im Entwicklungsverlauf direkt untersuchten. In Bezug auf Idiome ist anzumerken, dass Kinder dazu tendieren Idiome wörtlich zu interpretieren, bevor sie diese erlernt haben. Deshalb hat die Einzelwort-Bedeutung in diesem Falle eine große Relevanz. Einige Studien deuten darauf hin, dass insbesondere jüngere Kinder Idiome zunächst wörtlich interpretieren (Ackerman, 1982; Caillies & Le Sourn-Bissaoui, 2006; Levorato & Cacciari, 1995). Beispielsweise untersuchte Ackerman (1982), wie Kinder im Alter von sechs, acht und zehn Jahren und junge Erwachsene Idiome interpretieren, die entweder in neutralen, figurativen oder wörtlichen Kontexte eingebettet waren. Die Ergebnisse dieser Studie zeigen, dass mit steigendem Alter auch eine wörtliche Interpretation ab und eine figurative Interpretation zunahm. Ab einem Alter von ungefähr zehn Jahren interpretieren Kinder die meisten Idiome nicht mehr wörtlich. Außerdem konnten Kinder in diesem Alter Idiome auch in einem Kontext, der nicht die figurative Bedeutung unterstützte, figurativ interpretieren. Dies spricht dafür, dass Kinder bereits bekannte Ausdrücke nicht mehr wörtlich interpretieren.

Arnon und Christiansen (2017) argumentieren, dass, obwohl Kinder MWE als untersegmentierte Einheiten erwerben, sie trotzdem Einzelbestandteile dieser verfügbar haben. Weiterhin zeigen empirische Befunde für erwachsene Muttersprachler:innen, dass der Effekte der

Einzelwort-Häufigkeiten auf den Verarbeitungsvorteil von MWE mit steigender Phrasenhäufigkeit sinkt (Arnon & Cohen Priva, 2014). Dies impliziert, dass mit steigender Phrasenhäufigkeit die Mehrwort-Repräsentation von MWE gestärkt wird. Daher ist die Aktivierung der einzelnen Bestandteile innerhalb bekannter MWE im Hinblick auf deren Repräsentation im mentalen Lexikon bedeutsam. Die Fragestellung der **Studien II** und **III** bezieht sich darauf, ob Kinder in einem Alter von zehn und dreizehn Jahren Einzelwort-Bedeutungen innerhalb von Idiomen aktivieren. Anhand dieser Untersuchungen kann gemessen werden, ob sich die Repräsentation von MWE mit zunehmendem Alter verstärkt.

6 Studie II

keine Grenzen kennen

6.1 Forschungsfrage

Studie II zielte darauf ab die mentale Repräsentation bereits erworbener Idiome bei sprachlernenden Kindern im Altern von zehn Jahren zu untersuchen. Daher untersuchten wir, ob Kinder sensitiv gegenüber der konventionalisierten Wort-Konfiguration von Idiomen sind und ob sie einzelne Bestandteile dieser aktivieren. Diesbezüglich liegen generell noch wenig empirische Befunde vor. Insgesamt könnte diese Untersuchung mit zehnjährigen Kindern zusammen mit den Ergebnissen aus **Studie I** mit Erwachsenen Rückschlüsse auf Modelle zum Erwerb und zur Veränderung der Repräsentation von MWE im mentalen Lexikon (Kapitel 1.1) zulassen.

Im Kindesalter ist es noch sehr variabel, welche Idiome den Individuen bekannt sind (Sprenger et al., 2019). Deshalb haben wir versucht die Verwendung von Idiomen, die den Kindern bekannt waren, auf verschiedene Arten zu realisieren. Erstens verwendeten wir nur Idiome, die Erwachsenen sehr bekannt waren (siehe **Studie I**). Zweitens untersuchten wir Kinder einer Altersgruppe, die bereits relativ fortgeschritten im Idiom-Erwerb ist. Typischerweise kennen und verstehen Kinder im Alter von ca. zehn Jahren relativ viele Idiome (Ackerman, 1982; Bernicot, Laval & Chaminaud, 2007; Qualls et al., 2003; Vulchanova, Vulchanov & Stankova, 2011). Die Kinder sollten in den Experimenten jeweils bestimmen, welches die korrekte Idiom-Form war. Somit wollten wir kontrollieren, ob die individuellen Kinder die verwendeten Idiome kannten. Da wir untersuchen wollten, wie Kinder Idiome verarbeiten, die sie bereits kannten, schlossen wir nur die korrekt erkannten Idiome in die Analyse ein.

Das Design von **Studie II** lehnt sich an das aus **Studie I** mit erwachsenen Versuchspersonen an. Zu diesem Zweck führten wir das gleiche semantische Erwartungsparadigma mit dem gleichen linguistischen Material (siehe Anhang) in einem Visual-World-Eye-Tracking- und einem EKP-Experiment mit zehnjährigen Kindern durch. Dennoch beschränkten wir uns in **Studie II** nur auf die auditive Modalität, zum einen um die individuelle Lesefähigkeiten der Kinder nicht als konfundierende Variable einzuschließen, und zum anderen, weil die Aktivierung der Einzelwort-Bedeutung in **Studie I** nur für die auditive Modalität messbar war. Vorteilhaft an der Verwendung des semantischen Erwartungsparadigmas ist, dass sowohl untersucht werden kann, ob Kinder idiom-finale Wörter erwarten und ob sie die Bedeutung

dieser Wörter ko-aktivieren. Unter Verwendung des gleichen experimentellen Designs können Vergleiche zwischen den Altersgruppen in den **Studien I** und **II** angestrebt werden. Falls zehnjährige Kinder unterschiedliche Verarbeitungsmuster als Erwachsene zeigen, könnte dies auf eine Veränderung der mentalen Repräsentation mit zunehmendem Alter hinweisen.

Studie II zielte darauf ab, folgende Fragen zu beantworten (Abbildung 18):

- (1) Bauen Kinder eine Erwartung für das idiom-finale Wort auf? Kinder zeigen einen Verarbeitungsvorteil für häufige MWE (z.B. Bannard & Matthews, 2008). In **Studie II** soll konkret untersucht werden, ob Kinder anhand bekannter, aber unvollständiger Idiome (*Hannes ließ die Katze aus dem ...*) deren weitere Bestandteile (*Sack*) vorhersagen können. Innerhalb des Visual-World-Experimentes, erwarteten wir, dass sich ein solcher Effekt in mehr Fixationen auf der idiomatisch korrekten Vervollständigung als auf inkorrekten Vervollständigungen manifestiert. Wir erwarteten, in Anlehnung an die Daten der erwachsenen Versuchspersonen in diesem Paradigma (**Studie I** in Kapitel 4), dass der Start der Antizipation im zeitlichen Verlauf der Fixationen sichtbar ist. Innerhalb des EKP-Experimentes erwarteten wir, dass die Erwartung des idiom-finalen Wortes in Amplitudenunterschieden der N400- und PNP-Komponente zwischen korrekten und inkorrekten Vervollständigungen erkennbar ist. Falls Kinder diese Verarbeitungsunterschiede zeigen, würde dies die Annahme stützen, dass sie die Wort-Konfiguration von MWE repräsentieren.
- (2) Falls Kinder idiom-finale Wörter vorhersagen, aktivieren sie gleichzeitig auch deren Bedeutung? **Studie II** zielte darauf ab, zu untersuchen, ob Kinder die Einzelwort-Bedeutungen aktivieren, nachdem sie diese erkannt haben. Falls dies der Fall ist, erwarteten wir Fixationsunterschiede im Visual-World-Experiment und Amplitudenunterschiede im EKP-Experiment zwischen relatierten (*Korb*) und unrelatierten Vervollständigungen (*Arm*). Erwachsene Muttersprachler:innen scheinen die Einzelwort-Bedeutung innerhalb von Idiomen automatisch, aber nur kurzzeitig zu aktivieren, nachdem sie die Ausdrücke erkannt haben (**Studie I**). Demnach ist der zeitliche Verlauf der Aktivierung der Einzelwort-Bedeutung ausschlaggebend für die Art der Aktivierung.



Abbildung 18: Bauen zehnjährige Kinder eine Erwartung für das idiom-finale Wort auf und aktivieren sie währenddessen dessen Einzelwort-Bedeutung?

Zusätzlich führten wir in **Studie II** ein EKP-Kontrollexperiment mit durch, um zu untersuchen, ob Kinder im Alter von zehn Jahren in einem wörtlichen Kontext semantische Erwartungen über Einzelwörter aufbauen (siehe Kapitel 2.2). In diesem Experiment hörten Kinder wörtliche, neuartige Sätze (z.B. *Im Urlaub bauten sie eine Burg aus. . .*), die entweder mit einem erwarteten Wort (*Sand*), einem semantisch relatierten Wort (*Felsen*) oder einem unrelatierten Wort (*Kuchen*) vervollständigt wurden. Wir erwarteten, dass falls Kinder semantische Erwartungen aufbauen, dies in Amplitudenunterschieden zwischen relatierten und unrelatierten Vervollständigungen messbar wäre. Falls Kinder in **Studie II** unterschiedliche Verarbeitungsmuster für Idiome zeigen als Erwachsene in **Studie I**, könnten diese daher in Relation zu generellen Alterseffekten gestellt werden.

6.2 Manuskript: Delayed prediction of idiom constituent meaning points to weak holistic multi-word representation in children

Abstract

Children might store formulaic phrases (like the idiom *to let the cat out of the bag*) as multi-word representations. We investigated evidence for multi-word representations in ten-year-olds. Using eye-tracking and ERPs, we measured whether and when children predict idiom-final words (*bag*) and semantic associates of these words (*basket*). Both measures indicated that children predict idiom-final words which points to a multi-word representation of these expressions. Although, ten-year-olds are generally capable of forming rapid semantic predictions (as demonstrated in a control experiment), within idioms they activate semantic associates only in a late processing stage. Thus, delayed facilitation of semantic associates of predicted words was restricted to idiom constituents. We conclude that children rapidly activate multi-word units for idioms and decompose these units only after a short delay. This effect might relate to unitary representation of idioms, which will strengthen with an age-related increase of exposure to these expressions.

Introduction

Theories of language acquisition often discuss the size of “building blocks” with which children acquire and understand language (Arnon & Christiansen, 2017; Peters, 1983). Generative accounts of language learning assume words to be building blocks of language (e.g., Pinker, 1991). In contrast, usage-based accounts (e.g., Bannard & Matthews, 2008; Tomasello, 2000) posit that language acquisition and representation largely depend on statistical regularities beyond the level of single words. According to this view, children perceive and store frequently co-occurring words as large lexical units and do not necessarily analyze each single word when processing these multi-word representations. That is, formulaic multi-word expressions, such as collocations (*black coffee*), phrasal verbs (*dig into something*), and idioms (*kick the bucket*), could be acquired as “underanalyzed strings” (Arnon & Christiansen, 2017; Cruttenden, 1981; Wray, 2005). The present paper seeks evidence for multi-word units in children and asks how the cognitive architecture of those possible chunks looks like (i.e., whether or to what extent children analyze constituent meanings within these lexical units).

For adults, a well-established processing advantage of formulaic language indicates a special status of these fixed phrases in the mental lexicon (Gibbs, 1980; Swinney & Cutler, 1979). This advantage is reflected in faster judgements, in shorter reading times as well as in better recall of fixed formulaic phrases compared to novel expressions (e.g., Arnon & Snider, 2010; Carrol & Littlemore, 2020; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay & Baayen, 2010; Tremblay, Derwing, Libben, & Westbury, 2011; Underwood, Schmitt, & Galpin, 2004). Based on holistic processing accounts, it was argued that formulaic sequences are retrieved faster from

the mental lexicon than novel controls because there is no need for the parser to access single word meanings. This line of argumentation assumes that adults store formulaic language in form of relatively fixed multi-word representations.

Multi-word representations in adults appear to guide predictive processes (Cacciari & Corradini, 2015; Carrol & Conklin, 2020; Kessler, Weber, & Friedrich, accepted; Rommers, Dijkstra, & Bastiaansen, 2013; Underwood et al., 2004; Van Lancker Sidtis, Cameron, Bridges, & Sidtis, 2015; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2010). For example, Event-Related Potential (ERP) and eye-tracking experiments demonstrated that the linguistic system predicts idiom-final words (Kessler et al., accepted; Rommers et al., 2013). ERPs revealed indices of expectancy violation for sequences of written words and for strings of spoken words starting with typical idiom constituents respectively, but ending up with an unexpected word (e.g., *to let the cat out of the *arm*). That means, N400 amplitudes were enhanced compared to correct idiom completions (see below for further discussion of N400 effects in the ERP). In addition, eye-tracking showed that participants fixate a written version of the final idiom constituent well before this word appears in the auditory signal of a spoken version of this idiom (Kessler et al., accepted).

As prediction of idiom-final constituents has to be assisted by stored information about co-occurring words within idioms, predictive processing within idioms points to an involvement of some kind of respective memory representations such as multi-word representations. Stored information about co-occurrence might result from higher frequency of formulaic phrases or higher familiarity with these phrases compared to novel phrases (Arnon & Snider, 2010; Carrol & Conklin, 2020; Tabossi, Fanari, & Wolf, 2009). In general, some models assume that these effects reflect storage of multi-word expressions in form of multi-word representations in the mental lexicon (Arnon & Christiansen, 2017; Cacciari & Tabossi, 1988; Sprenger, Levelt, & Kempen, 2006; Tremblay & Baayen, 2010; Van Lancker Sidtis, 2009).

For children, some studies report a processing advantage of formulaic language similar to that obtained in adults (Bannard & Matthews, 2008; Qualls, Treaster, Blood, & Hammer, 2003; Reuterskiöld & Van Lancker Sidtis, 2013). For example, Bannard and Matthews (2008) report that two- and three-year-old children produce frequently occurring expressions (e.g., *sit in your chair*) faster than infrequent expressions (e.g., *sit in your truck*). Qualls and colleagues (2003) showed that fifth graders identified highly familiar idioms faster than less familiar idioms. Processing advantages of formulaic phrases in children might point to multi-word representations in the developing mental lexicon. Yet, we are not aware of any reports so far, in which evidence for such representations has been tested. Therefore, the present study directly investigated evidence for multi-word representations in children by measuring predictive processing within idioms.

If there are multi-word representations in children, the present study seeks to characterize their cognitive architecture. Are multi-word representations in children rather holistic or do they allow access to single word meanings? Idioms like *to let the cat out of the bag* might be perfectly suited to answer this research question. As idioms are formulaic phrases conveying a figurative meaning (to reveal a secret), which cannot be deduced from single constituent meanings (*to let, cat, out of, bag*), activation of idiom constituent meanings is not necessary for understanding the idiom. Therefore, hints to semantic activation of single constituents might be taken as indexing representation and processing at single word level.

To measure semantic activation within idioms, we relied on theories of semantic spread (Collins & Loftus, 1975). Words in the semantic memory that share semantic features are assumed to be closely connected (e.g., *candle – lamp*). The resulting network-like structure of lexico-semantic memory allows immediate spread of activation along these connections (Collins & Loftus, 1975; Kutas & Federmeier, 2000). Once a semantic representation is activated, semantically related words that are closely connected in the network get automatically co-activated. Thus, the activation of words that are semantically related to idiom constituents indexes the activation of the constituent meanings itself. Here, we will use indices of immediate spreading semantic activation exerted from idiom constituents as indicating processing of idioms at the single word level.

In adults, predictive processing within idioms points to relatively flexible multi-word representations, which allow short-living access to the semantic meaning of single constituents. We formerly showed predictions of semantic associates for idiom constituents that are not entailed in the input, but only predicted by participants (Kessler et al., accepted). When adult participants listened to idiom onsets up to the pre-final constituent, they not only fixated correct idiom completions, but also semantic associates to predicted idiom-final words well before the onset of the idiom-final word (e.g., fixations towards BASKET when listening to *let the cat out of the . . .*). In our former ERP experiment, we found reduced indices for violation of expectancy if the final word of an idiom was replaced by a semantically associated word (e.g., *to let the cat out of the basket*) compared to completely unrelated words. For both measures, the effect appeared only early during processing and declined rapidly. This evidence for semantic spread at the single word level converges with priming results pointing to automatic access to constituent meanings by adults (Beck & Weber, 2016; Rabanus, Smolka, Streb, & Rösler, 2008; Smolka, Rabanus, & Rösler, 2007; van Ginkel & Dijkstra, 2019).

The aim of the present study was to explore the mental representation of idioms by language learners, firstly by examining whether they predict final constituents, and secondly by measuring semantic spreading activation from this prediction. To this end, we investigated children’s analytic strategies during the processing of idioms. We used idioms like *to let the cat out of the bag* because they are perfectly suited in this respect. Prediction of idiom-final

words should indicate memory traces for these fixed multi-word expressions, while prediction of semantic associates should indicate some flexibility within respective mental representations. If children learn multi-word expressions as undersegmented chunks from the input without necessarily accessing their constituents (Arnon and Christiansen (2017), this should be reflected in initially rather fixed multi-word representations (i.e. no or less access to single word meaning compared to adults). We relied on our previously used experimental designs (Kessler et al., accepted, eye-tracking and ERP). In Experiment 1, we measured the time-course of prediction and constituent activation by means of an eye-tracking experiment. In Experiment 3, we employed a semantic expectancy paradigm with idioms and measured ERPs. Due to the lack of ERP experiments on semantic expectancy in children, we conducted a control experiment with literal, non-formulaic expressions in Experiment 2.

In order to measure whether children represent conventionalized forms of idioms but also co-activate constituent meanings, children should be already familiar with these formulaic phrases. To account for this, we chose the age group of nine- to ten-year-old children, that is age groups for which it was likely that children already knew most of the expressions. By the age of ten, idiom understanding and knowledge is quite developed (Ackerman, 1982; Bernicot, Laval, & Chaminaud, 2007; Qualls et al., 2003; Vulchanova, Vulchanov, & Stankova, 2011).

Experiment 1

In Experiment 1, we investigated whether children are able to predict idiom-final constituents including their semantic features before the offset of the phrase by means of a visual world eye-tracking paradigm during which participants hear spoken idioms and see written words presented on a screen (for an introduction into the visual world paradigm, see Huettig, Rommers, & Meyer, 2011). Eye fixations of adults, and even two-year-old children reveal that participants use language input to build up predictions about likely up-coming words (e.g., Altmann & Kamide, 1999; Kamide, 2008; Mani & Huettig, 2012, 2014). For example, upon hearing sentences like *the boy eats the big cake*, adults and children tend to look more towards eatable objects already shortly after the verb phrase *eat* (Altmann & Kamide, 1999; Mani & Huettig, 2014). Moreover, previous research using this paradigm shows that adults and children are sensitive towards semantic relations between acoustic linguistic stimuli (*biscuit*) and either pictures or printed words (*milk*) on the screen (Borovsky, Elman, & Fernald, 2012; Cortés-Monter, Angulo-Chavira, & Arias-Trejo, 2017; Huang & Snedeker, 2011).

The visual world paradigm has recently been used to study the time-course of idiom processing in adults (Holsinger, 2013; Kessler et al., accepted). These experiments showed that adult listeners consider literal word meanings within idioms. Kessler and colleagues (accepted) investigated whether adults predict idiom-final words alongside their semantic features. Participants heard highly predictable idioms where the final word was expected, but not presented auditorily (e.g., *Hannes let the cat out of the [bag]*) and simultaneously they

saw four written words (one in each quadrant of the screen). Written words were the correct idiom completions (*BAG*), distractors that were semantically related to the correct completion (*BASKET*), and two unrelated distractors (*STOMACH*, *ARM*). On the screen they presented four words in written forms that were the correct completions (*BAG*), distractors that were semantically related to the correct completion (*BASKET*), and two unrelated distractors (*STOMACH*, *ARM*). Adults showed anticipatory fixations towards the correct completion already 460 ms prior to the potential onset of the idiom-final word. The authors concluded that this early pre-activation of the idiom-final word indicates an early recognition of the memorized idiom configuration). Furthermore, adults showed early but briefly declining fixations towards semantic associates of correct completions. This indicated automatic constituent meaning activation simultaneously to the prediction of the idiom-final constituent.

In Experiment 1, we adopted the paradigm by Kessler and colleagues (accepted) to measure whether children show anticipatory fixations towards correct completions within familiar idioms. Because idiom knowledge largely differs between children (Sprenger, la Roi, & van Rij, 2019), we included idioms that were known by individual children. To this end, children had to choose which of the displayed words was the correct completion. Furthermore, we aimed to investigate whether this prediction leads to a co activation of semantic features. We hypothesize that semantic activation would be evident in more fixations towards semantically related distractors compared to unrelated distractors.

Methods

Participants Forty-one 4th graders participated in Experiment 2. Children were recruited from local schools. Prior to the experiment, parents gave informed consent. Data of all participating children could be included in the analysis ($N = 41$, mean age = 9;10 years, range = 9;3–10;4, 20 female and 21 male). Participants had no history of hearing disorders and normal or to normal corrected vision. All participating children were native, monolingual speakers of German. Children received a voucher for a local toy store as compensation for their participation. The Ethics Committee for Psychological Research at the University of Tübingen approved this study (reference number: 2016/1027/22).

Stimuli and Design We used the same design and stimuli as Kessler and colleagues (accepted, Experiment 1), and, hence, the description of the design is the same as in our previous manuscript. We selected 20 well-known German idioms (see Appendix). The idioms were embedded in sentences with a comparable structure (see Table 6): (i) a person carrying out the action of the sentence, (ii) a sentence body that originated from a German idiom, and (iii) the final target word of the idiom (which was not presented auditorily in Experiment 1). All idiomatic sentences were spoken in their complete form by a native speaker of German and digitally recorded. For Experiment 1, we removed the final target word from the recording. Participants heard each idiomatic sentence fragment once, while seeing four visual words on a computer

screen. The four words represented one of these four types: (1) Correct Completion: correct completion of the idiomatic phrase, (2) Related Distractor: semantic associate of the correct completion, and (3&4) Distractors Unrelated 1 and Unrelated 2: semantically unrelated to the correct completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Correct Completions and Related Distractors used with other sentence fragments in the experiment (avoiding phonological overlap). All words on the screen had the same grammatical gender fitting the preceding sentence context.

We ensured semantic relatedness between correct and related words by comparing pairwise semantic spaces using the R package LSAfun (Günther, Dudschig, & Kaup, 2015) and testing these similarity values with a Wilcoxon signed rank test. On average, semantic similarity between correct-related word pairs was significantly higher than between correct-unrelated1 ($Z = 189$, $p < .001$) and correct-unrelated2 ($Z = 185$, $p = .002$) word pairs. Semantic similarity between correct-unrelated1 and correct-unrelated2 word pairs did not differ ($Z = 75$, $p = .28$). Furthermore, semantic similarity between correct-related word pairs was significantly higher than between related-unrelated1 ($Z = 180$, $p = .004$) and related-unrelated2 ($Z = 182$, $p = .003$) word pairs.

Tabelle 6: German Example Sentence for Types 1-4 with English Equivalent

(i) Person	(ii) Sentence Body	(iii) Target Words			
		(1) Correct	(2) Related	(3) Unrelated 1	(4) Unrelated 2
Hannes	ließ die Katze aus dem	Sack	Korb	Bauch	Arm
<i>Hannes</i>	<i>let the cat out of the</i>	<i>bag</i>	<i>basket</i>	<i>stomach</i>	<i>arm</i>

Displayed words were presented in white font (Arial, font size 28) on a grey background. The position of the displayed words was counterbalanced across items and participants. The order of the trials was randomized.

Procedure Participants completed the experiment in a single session. For the experimental task, participants received both written and oral instructions. Prior to the experimental task, each participant received a 5-point grid for calibration and a practice block consisting of five trials.

An exemplary trial scheme is displayed in Figure 19. Each trial began with a 1500 ms inter-trial interval followed by a 500 ms presentation of a fixation cross. Then the visual

display of the set of four words appeared on the screen and remained until the end of the trial. The presentation of the audio stimuli started after a total of 2150 ms and was presented via headphones. After they heard the auditory stimuli, the task of the participants was to decide for each item which of the visually presented words was the best completion for the idiom by saying their choice out loud. The experimenter noted the participants' responses. Participants were instructed to press a button after their oral response in order to continue on to the next trial.

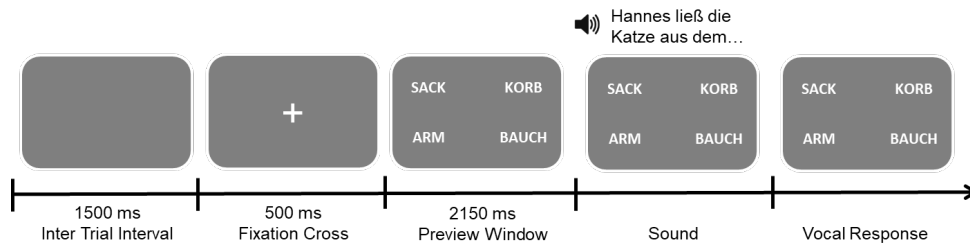


Abbildung 19: Example with times indicating the duration of the respective displays.

We recorded fixations using a portable Tobii eye-tracker with a sampling size of 60 Hz. In total, the eye-tracking experiment took around 20 minutes including instructions, calibration and the experimental task, which took around 10 minutes.

Analysis For the analysis, we divided the screen into four areas of interest. The analysis time window was aligned to the offset of each audio stimulus (offset = 0 ms). For the statistical analysis, we only included items responded to correctly, that is in which the participants completed the sentence aloud with the correct final word of the idiom (these were 79.76 % of all trials). Running t-tests comparing fixations towards correct completions and unrelated distractors at succeeding measurement points (every 16.67 ms) showed that participants' fixations were biased towards the correct idiomatic completion 112 ms prior to the offset of the audio stimuli ($p < .01$). This can be interpreted as the recognition point of the idiom. To compare the amount and time course of fixations towards related and unrelated distractors, we conducted a Growth Curve Analysis (GCA) with orthogonal polynomials (Mirman, Dixon, & Magnuson, 2008). As the starting point of the GCA time window, we chose the start of observable anticipation (112 ms prior to the offset) for a duration of 1200 ms.

Results

Figure 20 shows fixation patterns for words from all categories on the screen and the GCA model. Fixation proportions for Distractor Type were modeled with third-order orthogonal polynomials. Estimates for all terms are summarized in Table 7. There was an effect on the intercept of the model that indicates that overall participants fixated more on related than unrelated distractors. Distractor Type also had a significant effect on the quadratic term. This term reflects the inflection around the center. Fixations for unrelated distractors are more negative around the

center than for related distractors. Furthermore, the effect on the cubic term is marginally significant. This term relates to the inflections at the extremities of the curve. This effect would indicate that curves for fixations on both distractors converge. As a conclusion, participants fixate related distractors more than unrelated distractors and more specifically, this difference is strongest in the center of the analysis time window.

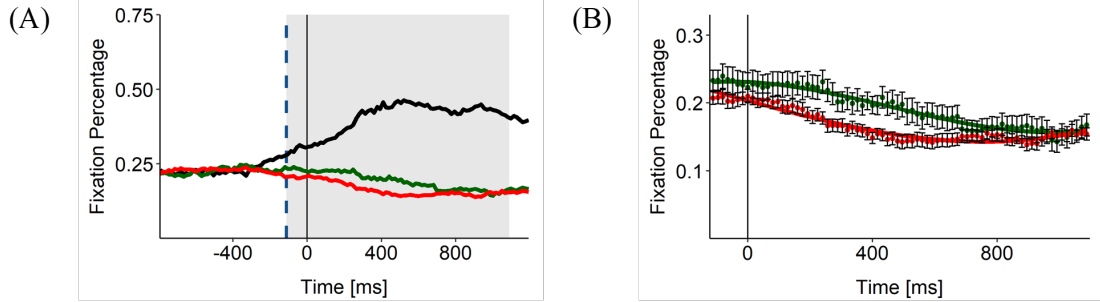


Abbildung 20: Panel (A) Fixation percentage for correct completions (black), related distractors (green) and mean of unrelated distractors (red); black vertical line = offset of spoken stimuli (0 ms); blue vertical, dashed line = start of the anticipation (-112 ms); grey background = time window for GCA. Panel (B) Fixation percentage for semantically related and unrelated distractors (points = mean; error bars = standard error) with fit of the growth curve model (line).

Tabelle 7: Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).

Term	Estimate	Standard Error	t	$p <$
Intercept	-0.01	0.00	-2.62	.01
Linear	0.02	0.03	0.84	.40
Quadratic	0.03	0.00	6.71	.00
Cubic	-0.01	0.00	-1.85	.06

Discussion

In Experiment 1, we aimed to investigate whether children predict idiom-final words and their semantic features in a visual world paradigm. We included only idiom that individual children recognized correctly. This resulted in an analysis of 79.76% of eye-tracking data. On average, children showed predictive fixations on correct idiomatic completions starting around 110 ms prior to the offset. Notably, it takes around 200ms to program a saccade (Saslow, 1967). Therefore, we assume that the anticipation of the correct completion occurred around 300 ms before the potential onset of the final word. This is a later start of anticipation for children

compared to adults for which anticipatory fixations towards the correct completions started around 460 ms (corrected for saccade programming: 660 ms) prior to the potential onset of the final idiom constituent (Kessler et al., accepted). Nevertheless, it seems that children also recognize and activate the conventionalized idiomatic configuration before the idiom-final word.

Children also tended to fixate more on related distractors than on unrelated distractors. Respective anticipatory fixations started approximately 300 ms after the onset of the fixation bias towards correct completions (not corrected for saccade programming: 110 ms). This indicates that children indeed access literal constituent meanings in idioms. But when comparing these results to the adults fixation pattern (Kessler et al., accepted) some differences become obvious. Relative to the start of the anticipation, increased fixations on the related distractor started later in children than in adults. While adults showed simultaneously increased fixations to correct completions and related distractors, children's fixations to related distractors started later than their fixations to correct completions. This might indicate that children do activate constituent meanings within idioms but not as automatically as adults do.

However, the simultaneous visual presentation of both the correct completion and the related distractor might explain the delayed semantic co-activation in children compared to adults. Auditorily presented idiom fragments did not contain the critical, idiom-final word. Nevertheless, it was presented as printed word on the screen. In this case the semantic spread might have occurred from the visual presentation instead of from the prediction of the correct idiomatic completion in children. In adults this explanation is unlikely, because they activate semantically related words in parallel with the correct completion (for further discussion, see also Kessler et al., accepted). However, children might show increased fixations towards related distractors later than adults, because this effect might be caused by bottom-up semantic spread from the printed, correct completion. To rule out, this explanation we conducted a semantic expectancy ERP experiment with idioms (Experiment 3). In this experiment, the correct completion was never presented in the input when measuring the processing of semantically related completions. Yet, because we did not recognize any ERP research on semantic expectancy in literal, non-formulaic language in children, we first conducted a control experiment addressing this question.

Experiment 2

In Experiment 2, we conducted a control experiment to investigate semantic expectancy for children with literal, non-formulaic language. Typically, semantic expectancy in literal, non-formulaic language is tested by means of amplitude differences of the N400 ERP component (Kutas & Federmeier, 2011). First, the N400 component is modulated by sentence-based expectations (e.g., Diaz & Swaab, 2007; Kutas & Hillyard, 1984). Sentence-final words that are highly expected show reduced N400 amplitudes compared to unexpected words. Second, the N400 component reflects pre-activation of semantic features (e.g., DeLong, Chan, & Kutas, 2018; Federmeier & Kutas, 1999; Ito, Corley, Pickering, Martin, & Nieuwland, 2016). In such

a paradigm using novel expressions (*They wanted to make the hotel look more like a tropical resort. So along the driveway, they planted rows of...*), sentences ended either with a highly expected word (*palms*), with a word that is not expected but belongs to the same semantic category as the expected exemplar (*pinés*) or with a word from a different semantic category (*tulips*). For these sentences, the N400 was modulated by semantic expectancy (Federmeier & Kutas, 1999). Not only did both unexpected exemplars elicit a more negative N400 amplitude than expected exemplars. In addition, exemplars from the same semantic category elicited a reduced N400 amplitude compared to exemplars from a different semantic category. Thus, due to the expectation of a word, respective semantically related words are easier to process than unrelated words. This N400 effect of semantic expectancy is not only observable for word-by-word reading but also for natural, connected speech (Federmeier, McLennan, De Ochoa, & Kutas, 2002).

While there is at least some evidence that children show reduced N400 for expected words compared to unexpected words within sentences (Meng, Tian, Jian, & Zhou, 2007), ERP evidence for semantic expectancy in children is lacking. So far, there is only offline evidence that children use sentential information to predict gradually matching words. Mahler and Chenery (2018) used an auditory word repetition paradigm in which six- to fourteen-year-old children had to repeat the last presented word in a sentence, naming latencies served as index of activation and recognition of the given word. They presented sentences (*The dog buried the...*) that either ended with a highly expected word (*bone*), a semantically congruous word that was not expected (*stick*) or an incongruous, not expected word (*air*). The results suggested a graded pattern of activation, with the highly expected word profiting the most from the sentential context and more activation for congruous compared to incongruous, unexpected words. While naming latencies reflect offline semantic prediction, no research so far considered online facilitation of semantically related words in children.

In order to test semantic expectancy within idioms in Experiment 3, we first conducted a control experiment to measure online semantic predictions in ten-year-old children in Experiment 2. To this end, we presented sentences with high expectancy of the final word (e.g., *On vacation, they built a castle out of...*). The word presented in final position was either the expected exemplar (*sand*), an exemplar that was semantically related to the expected exemplar (*rocks*) or an unrelated exemplar (*cake*). We hypothesize that if sentence predictions of the expected exemplar are strong enough for spreading activation this will be reflected in reduced N400 amplitudes for related compared to unrelated exemplars.

Methods

Participants Thirty-three 4th graders participated in Experiment 2. Children were recruited from local schools. As compensation, children received a voucher for either a local bookstore or a local toy store. Prior to the experiment, parents gave informed consent. Three children

were excluded from the analysis because of the following reasons: two due to artifactual EEG recordings and one due to left-handedness. All participants included in the analysis ($N = 30$, mean age = 9;8 years, range = 9;3 – 10;6, 15 male and 12 female) were right-handed, as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971), and had no history of neurological, psychiatric or hearing disorders. All children were native, monolingual speakers of German. The ethical committee of the German Psychological Association approved this study (reference number: 2017_0904_80).

Stimuli Sentences presented in the experiment consisted of a highly constraining sentence body and a target word in sentence-final position. Each sentence body was presented once in each of four conditions. Across the conditions, the characteristics of the respective target word differed: (1) Expected: the target word was the expected completion of the sentence, (2) Related: the target word was not expected, but semantically related to the respective Expected completion, (3&4) Unrelated 1 and Unrelated 2: the target words were semantically unrelated to the Expected completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Expected and Related Conditions (no phonological overlap, matching grammatical gender). In total, we presented 160 items, 40 in each of the four conditions. The same German-native as in Experiment 1 produced the stimuli.

In order to assure expectancy of the final words in the Expected completion, we collected Cloze Probability ratings, the probability with which participants complete sentences with a specific word. The cloze probability for the expected target word was on average 95.29%.

Relatedness between target words from different conditions was tested by comparing latent semantic spaces (LSA fun R package by Günther et al., 2015) with a Wilcoxon signed rank test. Semantic similarity was higher between expected vs. related target words than between the expected vs. both unrelated target words ($Z_{unrelated1} = 819$, $p < .001$; $Z_{unrelated2} = 814$, $p < .001$).

Procedure In the procedure, recordings and analysis we again relied on those applied in (Kessler et al., accepted) and description of these overlap partially with theirs.

Participants completed the experimental task in a single session. They sat in a comfortable chair facing a computer screen in a dimly lit room. During the experimental task, they were instructed to sit still and avoid eye-movements including blinking. Later, participants took part in a calibration task at the beginning and the end of the session. In this task, eye movements were systematically evoked for offline ocular correction. Before the experimental task, participants received both written and oral instructions. The participants received a practice block consisting of eight trials to ensure that they were familiar with the procedure and the task.

For each experimental trial, a sentence was presented auditorily via loudspeakers on both sides of the computer screen. During the presentation of the sentences, participants viewed a fixation cross at the center of the screen. After the auditory presentation, the task of the participants was to decide for each sentence whether it made sense or not by pressing buttons with the index fingers of the right or the left hand. The side for yes- and no-buttons was counterbalanced across participants. The response type was a delayed response; 1200 ms after onset of the target stimulus a question mark appeared at the center of the screen to signal the start of the response window for the participants. If they responded before the start of the response window, participants were given feedback (*too fast*). The interval between succeeding trials was 1500 ms.

The experiment consisted of 8 blocks of 20 trials, 160 trials in total, with five trials in each condition in each block. The order of trials was pseudorandomized in such a way that the same sentence body or target word never occurred in the same block. After each block, participants had the opportunity to take a self-timed break. The order of blocks was randomized using the Latin Square method. In total, the EEG-experiment took around 1.5 h including electrode application, instruction, calibration and the experimental task; the experimental task itself took around 15–20 minutes.

Electrophysiological Recordings and Analysis Electrophysiological brain potentials were recorded with 46 active electrodes (Ag/AgCl) mounted in an elastic cap (EasyCap GmbH, Herrsching, Germany) according to the 10-20 system (see Figure 21), online referenced to the nose. The ground electrode was positioned at the location of the AF3. In order to record eye movements, we attached two ocular electrodes below both eyes. The raw data were sampled at 500 Hz (bandpass filter 0.01–100 Hz, BrainAmpStandard, Brain Products, Gilching, Germany).

For the ERP analysis, the raw data were re-referenced offline to the average reference and filtered with a 0.3 Hz Low-Cut-Off filter. Using surrogate MultipleSource EyeCorrection (MSEC) by Berg and Scherg (1994), we removed horizontal and vertical eye movements as well as blinks from the continuous EEG signal. The EEG data were segmented into trials in epochs from 100 ms before and 1000 ms after the stimulus onset with a 100 ms pre-stimulus baseline subtraction. We excluded trials contaminated with artifacts and in which participants responded before the onset of the response time window (1200 ms after stimulus onset). Further, we only included individual items that participants responded to correctly in the Expected Condition in the analysis in all conditions (96.2%). For further analyses, we aggregated the conditions Unrelated1 and Unrelated2 into one condition Unrelated by averaging the mean voltages of the two conditions for each participant. Following this process, the final three conditions discussed in the analyses were: Expected, Related, and Unrelated.

We chose six Regions of Interest (ROIs), covering lateral and midline anterior and posterior sites (see Figure 21). Both lateral anterior ROIs included six electrode positions over both temporal cortices (left: F9, F7, FT9, FT7, FC5, T7; right: F10, F8, FT10, FT8, FC6, T8). The anterior midline ROI covered six fronto-central electrodes (F3, Fz, F4, FC1, FCz, FC2). Both lateral posterior ROIs included six temporo-parietal electrode positions (left: TP9, TP7, CP5, P7, PO9, O1; right: CP6, TP8, TP10, P8, PO10, O2). The posterior midline ROI covered six centro-parietal electrode positions (CP1, CP2, P3, Pz, P4, POz).

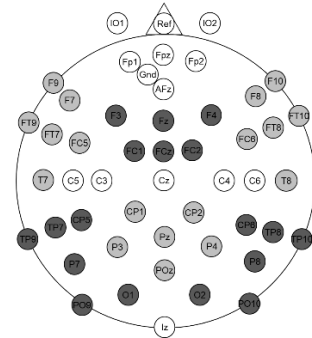


Abbildung 21: Electrode configuration used in the experiment. Anterior-Left, Anterior-Right, and Posterior-Central ROIs are highlighted in light gray. Anterior-Central, Posterior-Left, and Posterior-Right ROIs are highlighted in dark grey.

For statistical analysis, we conducted a 3 x 3 x 2 repeated-measures ANOVA with the within-participant factors Condition (Expected, Related, Unrelated), Hemisphere (Left, Central, Right) and Region (Anterior, Posterior). First, we conducted RM-ANOVAs for each 100 ms time window to identify relevant time windows. Based on the literature and this first step of analysis, we decided for three time windows that we aggregated for further analysis (see Table 8): 100–300 ms (PMN, N200, Nieuwland, 2019), 300–500 ms (N400, Kutas & Federmeier, 2011) and 600–1000 ms (PNP, Van Petten & Luka, 2012).

Results

Figure 22, Panel A depicts Grand-Average ERPs aggregated over Regions of Interest, Panel B shows difference topographies for relevant time windows. All reported p-values are Greenhouse-Geisser or Bonferroni (for post-hoc t-tests) corrected, respectively.

100–300 ms (pre-N400) For the pre-N400 time window, we found an interaction for Condition and Hemisphere ($F(4, 116) = 5.73, p = .001$). Post-hoc tests revealed main effects for Condition across all Hemispheres (Left: $F(2, 58) = 4.47, p = .02$; Central: $F(2, 58) = 7.22, p = .002$; Right: $F(2, 58) = 4.97, p = .01$). Across Left sites, there was a marginal significant difference between amplitudes for Expected and Related Conditions ($t_1(29) = -2.54, p = .051$), amplitude difference between Expected vs. Unrelated ($t_2(29) = -2.14, p = .122$) and Related vs. Unrelated ($t_3(29) = 1.10, p = .837$) did not differ. Across Central sites, we found significant amplitude differences between Expected vs. Related ($t_1(29) = 3.06, p = .014$) and Expected vs. Unrelated

Table 8: RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms	800–900 ms	900–1000 ms
C									
CxR			*	*		*	*	*	*
CxH	*	*	*	*	*	*			
CxRxH			*	*					*

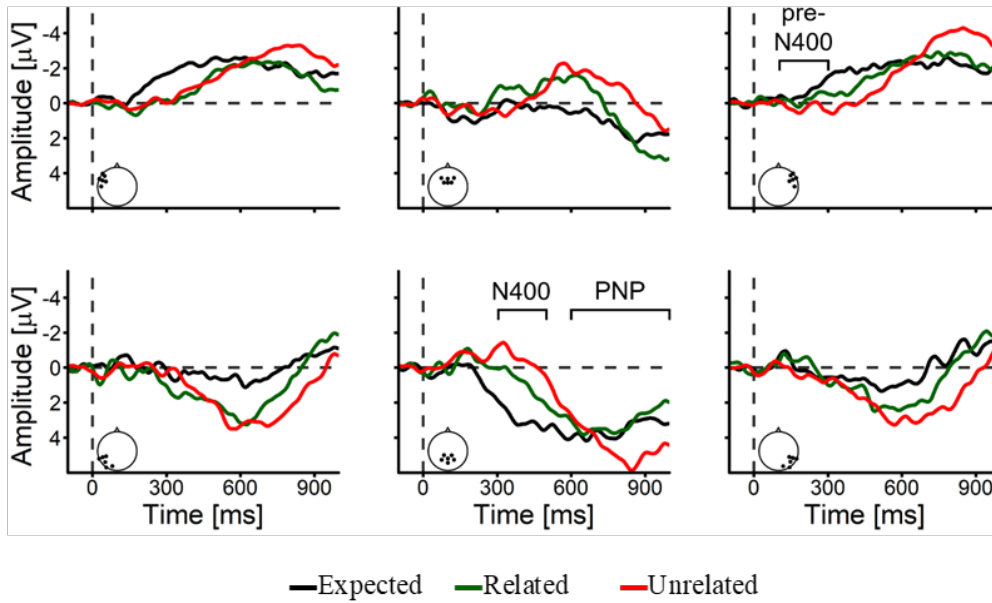
($t2(29) = 3.75, p = .002$), but not between Related vs. Unrelated ($t3(29) = 0.25, p = 1$). Across Right sites, there was no amplitude difference between Expected vs. Related ($t1(29) = 0.39, p = 1$), but between Expected vs. Unrelated ($t2(29) = -2.62, p = .042$) and Related vs. Unrelated ($t3(29) = -3.31, p = .008$). Thus, over central regions we find an effect expectancy and over right regions an effect that was more sensitive to semantic relations.

300–500 ms (N400) In the N400 time window, we found a significant three-way interaction for Condition, Region and Hemisphere ($F(4, 116) = 7.65, p < .0001$). Across Left electrodes, we found a main effect for Condition ($F(2, 58) = 11.59, p < .001$) with significant amplitude differences between Expected vs. Related ($t1(29) = -4.08, p = .001$) and Expected vs. Unrelated ($t2(29) = -3.25, p = .009$), but not between Related vs. Unrelated ($t3(29) = 1.70, p = .299$). Across Central sites, we found an interaction between Condition and Region ($F(2, 58) = 15.85, p < .0001$). Although there was a main effect for Condition ($F(2, 58) = 3.53, p = .04$) across Left electrodes, there were no significant amplitude differences between all three conditions (all $t(29) \leq |2.35|$, all $p \geq .077$). Across Posterior sites the main effect for Condition ($F(2, 58) = 28.61, p < .0001$) was due to significant amplitude differences between all three conditions: Expected vs. Related ($t1(29) = 4.09, p = .001$), Expected vs. Unrelated ($t2(29) = 7.68, p < .001$), Related vs. Unrelated ($t3(29) = 3.31, p = .008$). Across Right electrodes we did also find a main effect for Condition ($F(2, 58) = 8.10, p = .001$). Here, amplitudes for Expected and Unrelated Conditions differed ($t2(29) = -4.79, p < .001$), but amplitudes for Expected and Related ($t1(29) = -1.71, p = .296$), and Related and Unrelated ($t3(29) = -2.06, p = .147$) did not differ. Difference topographies in Figure 22 (Panel B) indicate that the N400 effect is most prominent over central posterior sites and graded for semantic expectancy.

600-1000 ms (PNP) In late time window, we found an interaction between Condition and Region ($F(2, 58) = 26.33, p < .0001$). There was a main effect for Condition for both Anterior

($F(2, 58) = 31.52, p < .0001$) and Posterior ($F(2, 58) = 22.20, p < .0001$) regions that showed the same effect: amplitude differences between Expected vs. Unrelated ($t2(29) \geq |5.64|, p < .001$) and Related vs. Unrelated ($t3(29) \geq |5.77|, p < .001$), but not between Expected and Related ($t1(29) \leq |1.09|, p \geq .858$). Thus, amplitudes for Unrelated differed from both other conditions being more negative over Anterior regions and more positive over Posterior regions.

Panel (A)



Panel (B)

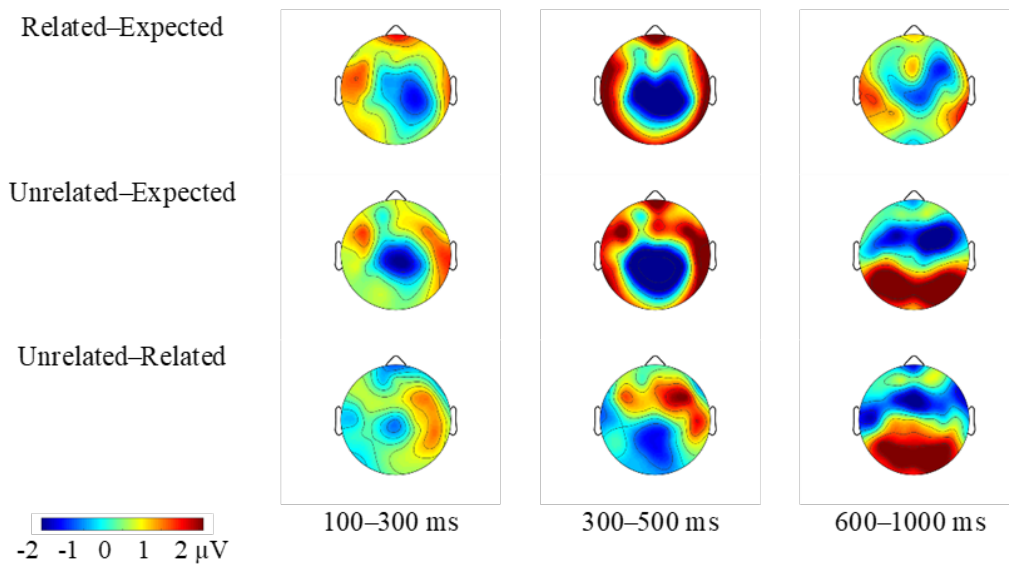


Abbildung 22: Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right; time windows for statistical analysis in boxes. (B) Difference topographies for the time windows 100–300 ms, 300–500 ms, and 600–1000 ms

Discussion

In Experiment 2, we investigated whether children form semantic predictions that are measurable in ERPs. Therefore, we conducted a typical ERP-experiment on semantic expectancy with sentences (e.g., *On vacation, they built a castle out of...*) that ended either with an expected completion (*sand*), an unexpected completion that was semantically related to the expected completion (*rocks*), or an unexpected completion (*cake*). We found a prominent N400 effect across central, parietal electrodes between 300 and 500 ms. This effect was graded for semantic relatedness: more negative amplitudes for unrelated words than for related words than for expected words. The present ERP data complement behavioral data by Mahler and Chenery (2018) and show that ten-year-old children do form rapid predictions of likely upcoming words and automatically activate semantic associates of these predictions.

In addition to the posteriorly distributed N400, we found an early effect between 100 and 300 ms indicating word form and meaning activation. This early ERP deflection showed mixed effects: Over central regions we found differences between expected and both related and unrelated conditions. Instead, over right regions, we found a difference between unrelated conditions and both expected and related conditions. There is a debate on whether the auditory N400 in adults is accompanied by an earlier negativity that is functionally different from the N400 (Nieuwland, 2019). Beyond that, our results might be interpreted in the light of the study by Friedrich and Kotz (2007) who found left-hemispheric differences in (phonological) form activation and right-hemispheric differences in meaning activation.

Further, we found a late positivity following the N400 that might be related to reinterpretation of unexpected exemplars with semantic contextual information (Brothers, Swaab, & Traxler, 2015; Kuperberg & Wlotko, 2020). Kuperberg and Wlotko (2020) reported a sensitivity for the plausibility of events. We find a late effect also for ten-year-old children with amplitude differences between unrelated and both expected and related completions, respectively, but no differences between expected and related exemplars. We cannot make a clear statement about the underlying mechanism, because we did not manipulate plausibility directly. However, plausibility might still have affected the amplitudes, because related completions are easier to reintegrate into the expected sentence concept than unrelated completions.

Concluding, observed graded N400 effects in our control experiment provide evidence that children form rapid semantic predictions in high cloze probability, non-formulaic sentences. The nature of a possible pre-N400 effect and post-N400 positivity need further investigation. Based on these results, we aimed to investigate whether children also form similar semantic predictions within idioms. Positive evidence would converge with results like described in Experiment 1.

Experiment 3

In Experiment 3, we applied the semantic expectancy ERP paradigm to test whether children predict idiom-final constituents and co-activate the meaning of these single words. Like in Experiment 2, we used an auditory paradigm to rule out reading-related differences between children. So far, the semantic expectancy ERP paradigm has been adopted in two studies to examine semantic activation within idioms for adult native speakers (Kessler et al., accepted; Rommers et al., 2013). Both studies presented idioms where the final word was highly predicted but either correct (*to let the cat out of the bag*), semantically related to the correct completion ([...] *basket*) or unrelated ([...] *arm*). Overall, adults show a high expectation of the correct idiom-final constituent. Based on the highly familiar, conventionalized idiom configuration adults pre-activate words within idioms.

Regarding ERP evidence for the activation of semantic associates in adults, Kessler and colleagues (accepted, Experiment 2) found an effect of semantic expectancy on the N400 for spoken idioms (for diverging results with written idioms, see Kessler et al., accepted; Rommers et al., 2013). Because children showed semantic expectancy effects on the N400 for non-formulaic phrases in Experiment 2, we expect that if children co-activate semantic features of idiom constituents we would observe graded N400 effects in Experiment 3.

The experiment by Kessler and colleagues (accepted, Experiment 2) with adults on spoken idioms found an early pre-N400 effect also reflecting semantic activation. This effect showed a parallel processing pattern: it was sensitive to semantic expectancy over anterior-left regions and sensitive to expectancy only over anterior-central and -right regions. Since this early effect was absent for the written modality in the experiment by Kessler and colleagues (accepted, Experiment 3) and the experiment by (Rommers et al., 2013), it seems to be specific for the processing of auditory stimuli (cf., Nieuwland, 2019). In the present Experiment 2, we show that a comparable early effect is also present in children for spoken, literal language processing. For the pre-N400 component, children also showed meaning activation for words within non-formulaic sentences. In contrast, results of Experiment 1 suggest that children do not form rapid semantic predictions in idioms: relative to the start of the anticipation of the correct idiomatic completion children show a late activation of constituent meanings. However, this late effect might also result from bottom-up semantic spread from the visually presented words. In Experiment 3 with ERPs, we were therefore interested in whether for idiomatic sentences children also activate single word meanings early on.

In the ERP-experiments by Rommers and colleagues (2013) and Kessler and colleagues (accepted, Experiment 2, auditory processing), a late effect succeeded the N400 effect that was not sensitive to relations between expected and semantically related completions. Both related and unrelated violations of the idiom yielded comparable amplitude differences com-

pared to correct idiomatic completions. For literal stimuli, the PNP has been observed when predictions are not met, irrespective of semantic relations, and has therefore been related to integration difficulties of semantic concepts for adults (for review see, Van Petten & Luka, 2012) and children (Experiment 2). In the context of idioms, it might thus reflect difficulties to integrate a manipulated form into the stored idiom configuration (Kessler et al., accepted, Experiment 2; Rommers et al., 2013). However, straight-forward predictions are hampered as Kessler and colleagues (accepted, Experiment 3) did not replicate this effect for written idioms.

For the present ERP-experiment on idiom processing in children, we used the paradigm by Kessler and colleagues (accepted), which is similar to Experiment 2. Stimuli were the same as in Experiments 1 and 2. As in Experiment 2, we presented complete versions of idioms (*Hannes let the cat out of the bag*), or manipulated version where the final highly predicted word was replaced by a semantically related word (*basket*) or unrelated words (*stomach, arm*). Again, children had to indicate whether they thought the respective expression was a correct idiom. We expect that amplitude differences between related and unrelated completions will give further insights into analyzing strategies in children.

Methods

Participants Thirty 4th graders participated in Experiment 3. Children were recruited from local schools. As compensation, children received a voucher for either a local bookstore or a local toy store. Prior to the experiment, parents gave informed consent. Six children were excluded from the analysis because of the following reasons: one due to technical problems and five due to too few trials (excluded when less than 10 trials out of 40 per condition). All participants included in the analysis ($N = 24$, mean age = 9;9 years, range = 9;3–10;4, 12 female and 12 male) were right-handed, as assessed by the Edinburgh Handedness Questionnaire (Oldfield, 1971), and had no history of neurological, psychiatric or hearing disorders. All children were native, monolingual speakers of German. The ethical committee of the German Psychological Association approved this study (reference number: RK 112015).

Stimuli We used the same design and materials as in Kessler and colleagues (accepted, Experiment 2). In the ERP experiment we used 40 idiomatic phrases using the same criteria of familiarity and predictability as in Experiment 1. Linguistic stimuli resulted from the combination of the sentence body with the four sentence final target words in four conditions with a combination logic following that of Experiment 1 (see Table 6). The conditions were the following: (1) Correct Condition: the target word was the correct completion of the idiomatic phrase, (2) Related Condition: the target word was semantically related to the correct completion, and (3&4) Conditions Unrelated 1 and Unrelated 2: the target word was semantically unrelated to the correct completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Correct and Related Conditions used with other sentence bodies in the experiment (no phonological or semantic overlap). Each sentence body was repeated four times, once in all four conditions. This

resulted in 160 different combinations of sentence bodies and target words. The same native speaker of German as in Experiment 1 spoke all linguistic stimuli. The linguistic stimuli that were repeated across conditions (sentence body and final words) were realized as the same recordings.

We had conducted rating studies to determine some characteristics of the materials essential for ERP research. In a cloze probability task, 17 adult participants read the 40 sentence bodies and filled in the word that they considered to be the most likely completion. The mean cloze probability of the correct idiom-final word was 93.82% ($SD = 9.69$).

Furthermore, we had controlled for the semantic relatedness between critical words by means of a second rating study. Fifteen adult participants received lists of word pairings of the target words and judged their relatedness on a scale from one to seven. The association strength between words presented as critical words in the Correct Condition (i.e., between the correct idiom completion) and words presented in the Related Condition (see Table 6) was significantly higher than the association strength of critical words presented in the Correct Condition and both Unrelated Conditions (Wilcoxon signed rank test: $Z_{unrelated1} = 120$, $p < .001$; $Z_{unrelated2} = 120$, $p < .001$). The association strength between critical words presented in the Correct Conditions and those presented in both Unrelated Conditions did not differ ($Z = 78$, $p = .32$).

Procedure The procedure was the same as in Experiment 2. In Experiment 3, the task of the participants was to listen to the sentences and decide for each item whether it was a correct idiomatic phrase or not. To indicate their response, they pressed a button with the index fingers of the right and the left hand. The side for yes- and no-buttons was counterbalanced across participants. Again, the response type was a delayed response with a response window starting 1200 ms after onset of the target stimulus.

Electrophysiological Recordings and Analysis Same as in Experiment 2. Again, we only included individual items that participants responded to correctly in the Correct Condition in the analysis in all conditions. We assumed that when participants recognized the idiom correctly in the Correct Condition (70.10%), they established an expectation of the final up-coming word. For further analyses, we aggregated the conditions Unrelated 1 and Unrelated 2 into one condition Unrelated by averaging the mean voltages of the two conditions for each participant. Following this process, the final three conditions discussed in the analyses were: Correct, Related, and Unrelated.

As in Experiment 2, we conducted a 3 x 3 x 2 repeated-measures ANOVA with the within-participant factors Condition (Correct, Related, Unrelated), Hemisphere (Left, Central, Right) and Region (Anterior, Posterior). First, we conducted RM-ANOVAs for each 100 ms time window to identify relevant time windows. Based on the literature, the results of Experiment 2

and this first step of analysis, we decided for three time windows that we aggregated for further analysis (see Table 9): 100–300 ms (PMN, N200, Nieuwland, 2019), 300–500 ms (N400, Kutas & Federmeier, 2011) and 600–1000 ms (PNP, Van Petten & Luka, 2012).

Table 9: RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions

	100–200 ms	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms	800–900 ms	900–1000 ms
C									
CxR						*	*	*	*
CxH	*	*	*	*	*	*			
CxRxH									

Results

Figure 23, Panel A depicts Grand-Average ERPs aggregated over Regions of Interest, Panel B shows difference topographies for relevant time windows. All reported p-values are Greenhouse-Geisser or Bonferroni (for post-hoc t-tests) corrected, respectively.

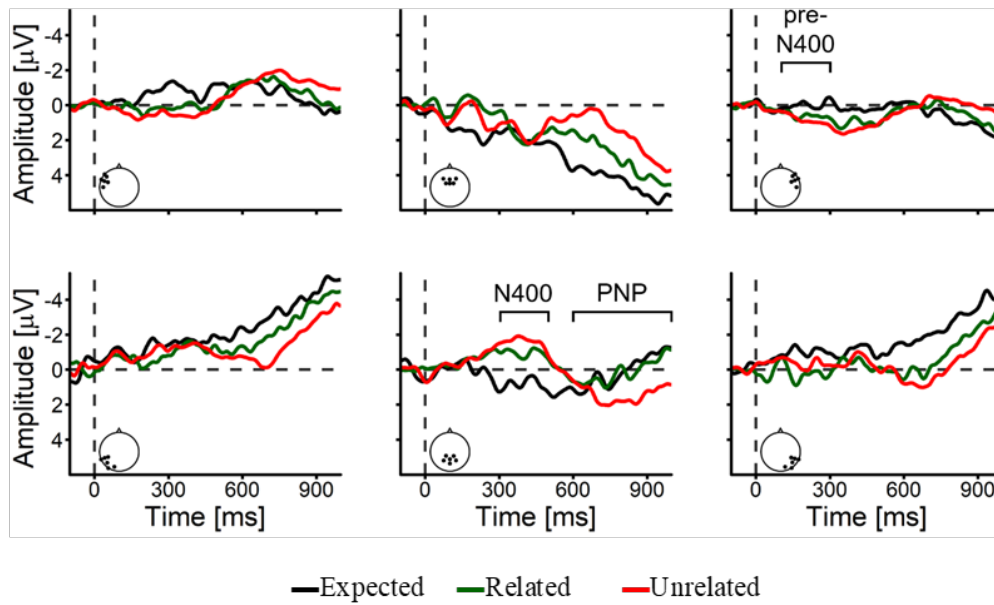
100–300 ms (pre-N400) In the pre-N400 time window we found a significant interaction of Condition and Hemisphere ($F(4, 92) = 5.20, p = .002$). Post-hoc tests revealed a main effect of Condition over Central ($F(2, 46) = 7.93, p = .002$) and Right sites ($F(2, 46) = 7.47, p = .002$). Over both sites this was due to amplitude differences between Correct vs. Related ($t1(23) \geq |3.48|, p \leq .006$) and Correct vs. Unrelated ($t2(23) \geq |3.33|, p \leq .009$), but not amplitude differences between Related vs. Unrelated ($t3(23) \leq |0.77|, p = 1$). Thus, related and unrelated violations yield in similar amplitudes indicating an effect of expectancy.

300–500 ms (N400) In the N400 time window, we found an interaction of Condition and Hemisphere ($F(4, 92) = 4.88, p < .001$). Across Central ($F(2, 46) = 6.86, p = .005$) and Right sites ($F(2, 46) = 4.32, p = .02$) post-hoc tests revealed amplitude differences between Correct vs. Unrelated ($t2(23) \geq |-2.87|, p \leq .026$), but not between Correct vs. Related ($t1(23) \leq |2.21|, p \geq .112$) and Related vs. Unrelated ($t3(23) \leq |0.97|, p = 1$). A similar trend was observable across Left sites but did not reach significance. Together, we found only amplitude differences

between Correct and Unrelated Conditions differed most prominently across Central and Right Hemispheres. However, when looking at difference topographies the N400 effect seems to be located over posterior central sites where also a difference between related and correct values is visible.

600–1000 ms (PNP) In the time window between 600–1000ms we found an interaction of Condition and Region ($F(4, 92) = 10.03, p < .001$). Post-hoc tests revealed main effects of Condition for both Anterior ($F(2, 46) = 9.06, p < .001$) and Posterior regions ($F(2, 46) = 10.20, p < .001$). For both Regions, amplitudes for Correct and Related Conditions did not differ ($t1(23) \leq |2.13|, p \geq .113$), amplitudes for Correct and Unrelated Conditions did differ significantly ($t2(23) \geq |4.16|, p = .001$). Amplitude differences between Related vs. Unrelated Conditions reached marginal significance level (Anterior: $t3(23) = 2.52, p = .057$, Posterior: $t3(23) = -2.50, p = .060$).

Panel (A)



Panel (B)

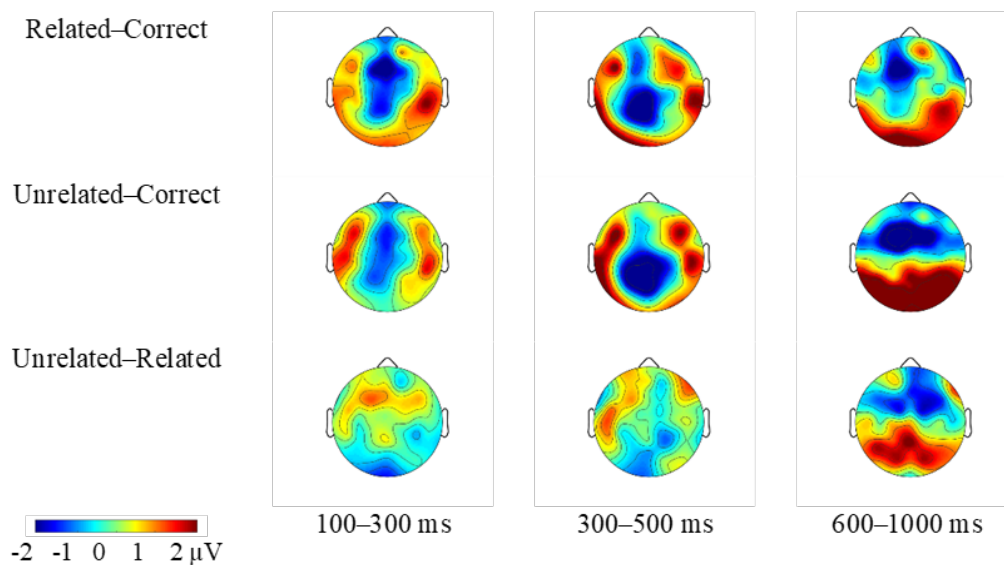


Abbildung 23: Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right; time windows for statistical analysis in boxes. (B) Difference topographies for the time windows 100–300 ms, 300–500 ms, and 600–1000 ms

Discussion

In Experiment 3, we used ERPs to test whether children aged nine to ten years predict idiom-final constituents due to the conventionalized idiomatic configuration and co-active semantic features of these constituents (as they did for word-by-word processing in non-formulaic sentence contexts in Experiment 2). Analysis was only conducted on idioms familiar to the individual children (in total 70.10 % of the trials).

In the N400 time window (300–500 ms, similar to Experiment 2), we only observed a difference between correct idiomatic and unrelated completions. Similar to ERP-experiments by Kessler and colleagues (accepted) and Rommers and colleagues (2013), we interpret this as an effect of expectancy. Children do predict idiom-final words in familiar idioms. Thus, ERP data converge with findings in the eye-tracking experiment and indicate multi-word representations.

In contrast to processing of non-formulaic phrases in Experiment 2, we did not obtain a difference between related and unrelated completions in the N400 component for idioms. This might indicate that children co-activate semantic associates during the prediction of idiom-final words. This effect was observable for central sites. However, typically the semantic N400 is located over posterior sites like also in idioms (Kessler et al., accepted; Rommers et al., 2013) and in literal, non-formulaic sentences in children (Experiment 2). In the present experiment with idioms, difference topographies also indicate that the N400 effect is located over posterior sites, even though the interaction of condition and region was not significant for central electrodes. Across central posterior electrodes, visual inspection of the ERP data suggest at least amplitude differences between correct and related completions that were not statistically significant for averaged central electrodes. However, since there was no statistical difference between related and both correct and unrelated conditions, we assume some kind of single word meaning activation.

Additionally, we find an early effect preceding the N400 that was not related to semantic expectancy. For auditory processing of idioms in adults (Kessler et al., accepted, Experiment 2) and of literal language in children (Experiment 2), a pre-N400 effect that is sensitive to semantic relatedness was observable. Both previous experiments show evidence for parallel processes of meaning and phonological form activation across different regions. In the present experiment, amplitudes significantly differed between correct and violations of the correct completion with no regional differences. Amplitudes between related and unrelated completions did not differ in this early effect. It seems that late semantic effects in the eye-tracking experiment might not be attributed to bottom-up semantic spread from words printed on the screen, but reflect that the meaning of the idiom constituent is not available for children early on.

Furthermore, we found a late effect following the N400 in a time window between 600 and 1000 ms that was affected by semantic relations between expected and unexpected completions. A post-N400 positivity was previously seen for literal language processing in adults (for review see, Van Petten & Luka, 2012) and children (Experiment 2), and for idioms in adults (spoken idioms: Kessler et al., accepted, Experiment 2; written idioms: Rommers et al., 2013). In contrast to the previous idiom studies, the effect here was affected by semantic relatedness. Amplitudes for correct completions were statistically indistinguishable from related completions. Amplitudes for unrelated completions differed from correct and with a tendency also from related completions. Assuming that the late effect reflects integration difficulties of both violations with the correct idiomatic form that is stored in the semantic memory (cf. Rommers et al., 2013), we conclude that the representation of idioms is not strong in ten-year-old children yet. However, especially because the underlying mechanisms of the PNP are not well established in children and the PNP for idioms is not consistent across experiments in adults (Kessler et al., accepted; Rommers et al., 2013), we can only make tentative interpretations at this point.

In general, the results converge with findings in the eye-tracking experiment (Experiment 1). In both experiments, children anticipate correct idiomatic completions. Therefore, we argue that children are sensitive to idiom configurations and store idioms as multi-word representations. In both experiments, children also show only effects of late semantic activation which might indicate weakness in these multi-word representations.

General Discussion

In the present study, we investigated online processing of idiomatic expressions in ten-year-old children. Global processing advantages of fixed formulaic phrases compared to novel expressions (faster judgements, shorter reading times, better recall) point to a special status of formulaic phrases in the mental lexicon of adults and children (e.g., Carrol & Conklin, 2019; Qualls et al., 2003; Tabossi et al., 2009; Tremblay & Baayen, 2010). However, at least in adults, this special status of respective multi-word representations does not suppress analysis of individual constituents: Several studies with adults showed that they activate semantic associates of individual components within idioms (Beck & Weber, 2016; Holsinger, 2013; Kessler et al., accepted; Smolka et al., 2007; Sprenger et al., 2006). This might be different in children, who are assumed to acquire formulaic phrases as “underanalyzed” units (Arnon & Christiansen, 2017). To test this assumption, we conducted eye-tracking and ERP experiments with familiar, spoken idioms in ten-year-old children.

Based on our former research with adults (Kessler et al., accepted), we measured semantic expectancy of final constituents of familiar idioms (i.e., we compared the amount of fixations and ERP amplitudes elicited by words semantically related to the expected final constituent of an idiom with those elicited by unrelated words). In both experiments, children performed an idiom recognition task. In the analysis, we only included those items that

children correctly recognized as idioms to assure familiarity with respective phrases. In the eye-tracking experiment, children recognized roughly 80 % of the presented idioms correctly and in the ERP experiment, they recognized roughly 70 % correctly. Percentages are numerically higher than in the experiment by Qualls and colleagues (2003), a study in which fifth graders correctly identified 62 % of the presented idioms. This confirms that we selected idioms that were relatively familiar to children although we relied on familiarity ratings from adult participants.

In addition to the behavioral data, both eye-tracking and ERP results confirm that, children recognized idioms correctly and predicted idiom-final constituents based on only partial input of the idioms (indicated by increased fixations towards correct, idiomatic completions in the eye-tracking and by amplitude difference between correct, idiomatic and incorrect completions in the ERPs). Based on these results, we conclude that children build up strong expectations for idiom-final constituents, showing that they are familiar with the idiom configuration. These effects show that children store some information about the co-occurrence of single constituents in idioms, that is, they point to multi-word representations of idioms in children's mental lexicon.

Compared to a previous study with adults (Kessler et al., accepted), the present eye-tracking data suggest that children start to anticipate the correct completion of the idiomatic expression later than adults do. On the one hand, these differences might result from general age-related differences in language abilities (vocabulary size, production skills: Borovsky et al., 2012; Huettig & Pickering, 2019; Mani & Huettig, 2012, 2014). On the other hand, they might relate to age-dependent differences in the frequency of exposure: Adults encountered these idiomatic expressions more often than children and, therefore, adults might anticipate idiomatic completions earlier than children do. This could include that with increasing age and experience with the idioms, the multi-word representation becomes stronger.

Responses to semantic associates of idiom constituents (in the related condition respectively) reveal that multi-word representations for idioms allow some access to single constituents in children. In Experiment 1, children fixated related distractors more often than they fixated unrelated distractors. In Experiment 3, we found ERP amplitude differences between related and unrelated completions of the idiom (at least during a late time window between 600 and 1000 ms). Thus, both experiments were indicative of some access that children have to individual idiom constituents.

However, we find possible temporal differences in activation of constituent meanings between children in the present experiment and our former study with adults (see Kessler et al., accepted). Children started to fixate semantic associates considerably later than they started to fixate the correct completions. Adults did not show such a delay: They considered the correct completion and its semantic associate simultaneously (Kessler et al., accepted). In the ERPs in Experiment 3, children showed amplitude differences between related and

unrelated completions only in a late component, but not in a pre-N400 effect. This contrasts to our former data obtained with adults, who showed amplitude differences between related and unrelated completions in a pre-N400 component, but not in a late component. Thus, processing mechanisms involved in the activation of constituent meanings seem to differ between native adults and children. Notably, children appear to process semantic associates of idiom constituents with a delay (compared to correct completions).

In adults, early effects of semantic expectancy have been attributed to an automatic activation of constituent meanings (Kessler et al., accepted). Experiment 2 shows that children are generally able to exploit this mechanism: When listening to literal sentences with highly predictable sentence endings, children show early ERP effects for semantic associates of highly expected sentence endings (compared to unrelated sentence endings). Therefore, we conclude that the present differences between adults and children in idiom processing are not caused by general age-related processing differences, but are specifically related to idiom processing. It appears that children do not automatically activate constituent meanings within idioms. They rather seem to build up strong expectations for familiar idioms without automatically analyzing the semantic features of the single constituents making up these multi-word units.

We find that children do show an effect of semantic activation later in idiom processing which might point to weak multi-word representations. In adults, missing semantic activation in a late ERP effect in adults has been interpreted as reflecting the violation of the expected idiom configuration that is represented in the mental lexicon (Kessler et al., accepted; Rommers et al., 2013). Accordingly, we might conclude that the mental representation of idiomatic expressions in the developing mental lexicon is still weak. This leads to an increased acceptability of related violations of the expected idiom configuration. It appears that children store information about the familiar idiom configurations and build up expectations about final words. However, this multi-word representation of the idiom might not be strong enough yet.

We assume that frequency of exposure might be driving the strength of mental representations of familiar idioms. This is in line with some accounts on the acquisition of multi-word units assuming that frequent exposure of these expressions leads to an entrenchment in the mental lexicon (Arnon & Cohen Priva, 2014; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay & Baayen, 2010). Adults and children process expressions that occur frequently in corpus data faster than less frequent phrases (e.g., Arnon & Cohen Priva, 2014; Arnon & Snider, 2010; Bannard & Matthews, 2008). Furthermore, the age at which idioms or other formulaic phrases are typically acquired relates to the effort adults need for processing them (Arnon, McCauley, & Christiansen, 2017; Bonin, Meot, & Bugajska, 2013): For example, Bonin and colleagues (2013) showed that adults read early acquired idioms faster than later acquired idioms. This might relate to more exposure to early-acquired phrases. Therefore, we assume that the differences of effects between children in the present study and adults in Kessler and

colleagues (accepted) can also be explained by age-related differences in the frequency both groups of participants have heard and read idiom across their life so far.

To our knowledge, this is the first study examining online analyzing strategies that children apply to idioms. In continuing this new line of research, future studies on acquisition of multi-word expressions should focus on relevant characteristic of idioms like individual familiarity. Although native adults rated the idioms that we presented as being highly familiar, this does not necessary imply that these idioms are highly familiar to children as well. Even though children correctly recognized many idioms that we presented, we did not test the strength of their familiarity with individual idioms that might play a role for processing (e.g., Carrol & Conklin, 2019; Qualls et al., 2003).

Taken together, we suggest that if children are familiar with an idiomatic expression, they expect the final constituent. Despite limited experience with the expressions, it appears that children learn idioms as fixed units and store them as multi-word representations (cf. Reuterskiöld & Van Lancker Sidtis, 2013). When processing these fixed units, children do not have constituent meanings automatically available. Multi-word representation might become stronger with accumulating experiences (cf. Arnon & Clark, 2011; Siyanova-Chanturia, Conklin, & Schmitt, 2011). Compared to children, strong, but not entirely fixed multi-word representations might allow adults to rapidly predict correct idiom completions as well as their semantic associates. Corroborating with Arnon and Christiansen (2017) we propose that children store multi-word representations and have constituent information partially available.

Acknowledgement

We would like to thank Julia Blumenthal, Babette Jakobi, Sören Koch, Julia-Larissa Maier, Dr. Anne Rau, Johanna Schläger, Holger Schneider, Charlotte Veil and Miriam Wild. We also warmly thank all children that participated in the experiments as well as their parents. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 75650358 – SFB 833. The founding source had no involvement in the study.

References

- Ackerman, B. (1982). On Comprehending Idioms: Do children get the picture? *Journal of Experimental Child Psychology*, 33(3), 439-454.
- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1-L2 differences. *Topics in Cognitive Science*, 9(3), 621-636.

- Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth – Children’s word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7(2), 107-129.
- Arnon, I., & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, 9(3), 377-400.
- Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 265-280.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62(1), 67-82.
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children’s repetition of four-word combinations. *Psychological Science*, 19(3), 241-248.
- Beck, S. D., & Weber, A. (2016). Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, 7, 1305.
- Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, 39(12), 2115-2132.
- Bonin, P., Meot, A., & Bugajska, A. (2013). Norms and comprehension times for 305 French idiomatic expressions. *Behavior Research Methods*, 45(4), 1259-1271.
- Borovsky, A., Elman, J. L., & Fernald, A. (2012). Knowing a lot for one’s age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112(4), 417-436.
- Brothers, T., Swaab, T. Y., & Traxler, M. J. (2015). Effects of prediction and contextual support on lexical processing: prediction takes precedence. *Cognition*, 136, 135-149.
- Cacciari, C., & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idioms processing: a reading-time study. *Journal of Cognitive Psychology*, 27(7), 797-811.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27(6), 668-683.
- Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, 63(1), 95-122.
- Carrol, G., & Littlemore, J. (2020). Resolving figurative expressions during reading: The role of familiarity, transparency, and context. *Discourse Processes*, 57(7), 1-18.
- Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82(6), 407-428.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29(1), 72-89.

- Cortés-Monter, D. R., Angulo-Chavira, A. Q., & Arias-Trejo, N. (2017). Differences between skilled and less-skilled young readers in the retrieval of semantic, phonological, and shape information. *Journal of Research in Reading, 40*, S170-S189.
- Cruttenden, A. (1981). Item-learning and system-learning. *Journal of Psycholinguistic Research, 10*(1), 79-88.
- DeLong, K. A., Chan, W. H., & Kutas, M. (2018). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology, e13312*.
- Diaz, M. T., & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research, 1146*, 85-100.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41*(4), 469-495.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E., & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39*(2), 133-146.
- Friedrich, C. K., & Kotz, S. A. (2007). ERP evidence of form and meaning coding during online speech recognition. *Journal of Cognitive Neuroscience, 19*(4), 594-604.
- Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition, 8*(2), 149-156.
- Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods, 47*(4), 930-944.
- Holsinger, E. (2013). Representing idioms: syntactic and contextual effects on idiom processing. *Language and Speech, 56*(3), 373-394.
- Huang, Y. T., & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language, 38*(3), 644-661.
- Huettig, F., & Pickering, M. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences, 23*(6), 464-475.
- Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica, 137*(2), 151-171.
- Ito, A., Corley, M., Pickering, M. J., Martin, A. E., & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language, 86*, 157-171.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass, 20*(4), 647-670.
- Kessler, R., Weber, A., & Friedrich, C. K. (accepted). Activation of literal word meanings in idioms: Evidence from Eye-tracking and ERP experiments. *Language and Speech*.
- Kuperberg, G., & Wlotko, E. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience, 32*(1), 12-35.
- Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences, 4*(12), 463-470.

- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, *62*, 621-647.
- Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, *307*(5947), 161-163.
- Mahler, N. A., & Chenery, H. J. (2018). A developmental perspective on processing semantic context: Preliminary evidence from sentential auditory word repetition in school-aged children. *Journal of Psycholinguistic Research*, *48*(1), 81-105.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake-but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843-847.
- Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264-279.
- Meng, X., Tian, X., Jian, J., & Zhou, X. (2007). Orthographic and phonological processing in Chinese dyslexic children: An ERP study on sentence reading. *Brain Research*, *1179*, 119-130.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475-494.
- Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? A critical review. *Neuroscience and Biobehavioral Reviews*, *96*, 367-400.
- Nippold, M. A., & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech, Language, and Hearing Research*, *36*(4), 728-737.
- Oldfield, R. C. (1971). The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*, *9*(1), 97-113.
- Ortony, A., Turner, T. J., & Larson-Shapiro, N. (1985). Cultural and instructional influences on figurative language comprehension by inner city children. *Research in the Teaching of English*, *19*(1), 25-36.
- Peters, A. M. (1983). *The Units of Language Acquisition* (Vol. 1). Cambridge: Cambridge University Press.
- Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530-535.
- Qualls, C. D., Treaster, B., Blood, G. W., & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: a reaction time study. *Journal of Communication Disorders*, *36*(4), 245-261.
- Rabanus, S., Smolka, E., Streb, J., & Rösler, F. (2008). Die mentale Verarbeitung von Verben in idiomatischen Konstruktionen. *Zeitschrift für Germanistische Linguistik*, *36*(1), 27-47.
- Reuterskiöld, C., & Van Lancker Sidtis, D. (2013). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy*, *29*(2), 219-231.

- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25(5), 762-776.
- Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, 57(8), 1024-1029.
- Sivanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27(2), 251-272.
- Sivanova-Chanturia, A., Conklin, K., & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(3), 776-784.
- Smolka, E., Rabanus, S., & Rösler, F. (2007). Processing verbs in German idioms: Evidence against the Configuration Hypothesis. *Metaphor and Symbol*, 22(3), 213-231.
- Sprenger, S., la Roi, A., & van Rij, J. (2019). The Development of Idiom Knowledge Across the Lifespan. *Frontiers in Communication*, 4, 29.
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54(2), 161-184.
- Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18(5), 523-534.
- Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, 37(4), 529-540.
- Tomasello, M. (2000). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, 11(1/2), 61-82.
- Tremblay, A., & Baayen, R. H. (2010). Holistic Processing of Regular Four-word Sequences: A Behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151-173). London: Continuum.
- Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61(2), 569-613.
- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 155-172). Amsterdam: John Benjamins.
- van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, 23(1), 131-147.
- Van Lancker Sidtis, D. (2009). Formulaic and novel language in a 'dual process' model of language competence. *Formulaic Language*, 2, 445-470.
- Van Lancker Sidtis, D., Cameron, K., Bridges, K., & Sidtis, J. J. (2015). The formulaic schema in the minds of two generations of native speakers. *Ampersand*, 2, 39-48.

- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190.
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682-1700.
- Vulchanova, M. D., Vulchanov, V., & Stankova, M. (2011). Idiom comprehension in the first language: a developmental study. *Vigo International Journal of Applied Linguistics*, *8*, 207-234.
- Wray, A. (2005). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

6.3 Schlussfolgerungen

Studie II untersuchte, wie sprachlernende Kinder Idiome verarbeiten, um daraus Rückschlüsse auf den Erwerb dieser Repräsentation ziehen zu können. Zu diesem Zweck untersuchten wir mittels eines Eye-Tracking- und eines EKP-Experimentes, (1) ob Kinder das idiom-finale Wort erwarten, nachdem sie das Anfangsfragment gehört haben, und, (2) ob Kinder währenddessen die Einzelwort-Bedeutung aktivieren. In **Studie I** konnten wir zeigen, dass erwachsene Muttersprachler:innen sowohl idiom-finale Wörter vorhersagen als auch deren Bedeutung kurzzeitig ko-aktivieren. Verarbeitungsunterschiede zwischen den Altersgruppen könnten auf eine Veränderung der Repräsentation mit zunehmendem Alter hinweisen.

Zunächst zeigten die Ergebnisse von **Studie II**, dass zehnjährige Kinder relativ viele Idiome korrekt erkennen können (Eye-Tracking-Experiment: 79.76 % korrekt, EEG-Experiment: 70.10 % korrekt). Bezüglich der Frage (1) fanden wir Evidenz, dass Kinder für diese Idiome eine starke Erwartung über die korrekten, idiom-finalen Wörter aufbauen. Dieser Effekt war sowohl in Fixationsunterschieden als auch in EKP-Amplitudenunterschieden zwischen korrekten und inkorrekten Vervollständigungen evident. Dadurch bestätigen diese Daten, dass die Ausdrücke den Kindern bekannt gewesen sein müssen. Falls die Kinder die korrekte Antwort nur erraten hätten, würden wir nicht erwarten, Evidenz für die Erwartung in den Eye-Tracking- und EKP-Daten zu finden. Ferner zeigten die Eye-Tracking-Daten, dass Kinder die idiomatisch korrekten Vervollständigungen vor dem Präsentationsende des Idiomfragmentes fixieren, das heißt, bevor dieses Wort im Input realisiert worden wäre. Daher kann man annehmen, dass den Kindern die Wort-Konfiguration der Idiome bekannt war und diese als Mehrwort-Repräsentation verfügbar ist.

Eine Limitation der **Studie II** ist jedoch, dass, obwohl wir auf verschiedene Weise versucht haben, die Bekanntheit der Idiome für die Kinder sicherzustellen, keine direkten Daten für die Stärke der individuellen Bekanntheit vorliegen. Die Bekanntheit von MWE steht aber in einem positiven Zusammenhang mit der Verarbeitungsgeschwindigkeit (z.B. Carrol & Conklin, 2020; Qualls et al., 2003; Tabossi et al., 2009). Im Erwachsenenalter spielen Faktoren wie deren Zerlegbarkeit (wie stark einzelne Wörter innerhalb des Idioms zu dessen Phrasen-Bedeutung relatiert sind) bei der Verarbeitungsgeschwindigkeit eine größere Rolle für weniger bekannte Idiome als für bekannte Idiome (Carrol & Conklin, 2020; Carrol & Littlemore, 2020). Daher ist es von besonderer Wichtigkeit für Experimente mit Kindern spezifischere Maße für den Grad der individuellen Bekanntheit zu verwenden. Obwohl man detailliertere Daten zu der individuellen Bekanntheit für Kinder allgemein erfragen kann, erweist es sich als schwierig, innerhalb eines Experimentes, wie in **Studie II** durchgeführt, diese korrekt zu messen. In dieser Studie konnten wir messen, ob den individuellen Kindern die jeweiligen Idiome bekannt waren oder nicht, aber nicht, wie sehr sie ihnen bekannt waren.

In Bezug auf die Aktivierung der Einzelwort-Bedeutung innerhalb eines Idioms (Frage (2)) zeigten die Ergebnisse von Studie II zunächst sowohl zeitlich relativ späte Fixations- als auch Amplitudenunterschiede zwischen relatierten und unrelatierten Vervollständigungen eines Idiomfragments. Es gibt keine Evidenz dafür, dass Kinder diese gleichzeitig mit der Vorhersage des finalen Wortes aktivierten, obwohl sie im Rahmen wörtlicher, nicht-formelhafter Ausdrücke dazu in der Lage sind (Kontroll-Experiment). Im Eye-Tracking-Experiment waren Fixationsunterschiede zwischen relatierten und unrelatierten Vervollständigungen nicht direkt mit dem Start der antizipatorischen Fixationen auf die korrekte Vervollständigung evident. Diese Unterschiede treten erst zeitlich später während der Verarbeitung auf. Ebenso zeigt sich keine frühe semantische Aktivierung in den EKP. Amplituden zwischen relatierten und unrelatierten Vervollständigungen unterschieden sich nicht in einer frühen Komponente, obwohl dies der Fall für wörtliche, nicht-formelhafte Sprache ist. Für die N400-Komponente fanden wir keine Amplitudenunterschiede zwischen relatierten und unrelatierten Vervollständigungen. Jedoch unterschieden sich relatierte und idiomatisch korrekte Vervollständigungen auch nicht signifikant. Dies weist bereits tendenziell auf eine Aktivierung der Einzelwort-Bedeutung hin. In der PNP hingegen unterschieden sich Amplituden zwischen relatierten und unrelatierten Vervollständigungen. Daher zeigten Kinder sowohl in Eye-Tracking- als auch EKP-Daten insgesamt einen späten Effekt der Aktivierung der Einzelwort-Bedeutung.

Im Vergleich zu Erwachsenen (**Studie I**), die sehr früh, aber nur kurzzeitig während der Verarbeitung die Einzelwort-Bedeutung aktivieren, trat dieser Effekt bei zehnjährigen Kindern später auf (ausführliche Diskussion folgt in Kapitel 8). In dem Kontroll-Experiment mit wörtlichen, neuartigen Sätzen konnten wir zeigen, dass Kinder generell in der Lage sind schnelle semantische Repräsentationen zu aktivieren. In diesen Sätzen zeigten Kinder signifikante Amplitudenunterschiede zwischen relatierten und unrelatierten Vervollständigungen bereits in der prä-N400- und der N400-Komponente. Daher gehen wir davon aus, dass der späte semantische Effekt bei Idiomem für Kinder keine automatische semantische Aktivierung reflektiert. Alternativ sprechen die Ergebnisse dafür, dass die Kinder bisher noch keine starke Repräsentation der Idiome im mentalen Lexikon aufgebaut haben. Diese Schwäche der Repräsentation spiegelt sich insofern in der späten Aktivierung der Einzelwort-Bedeutung wider, als dass Kinder manipulierte, aber relatierte Varianten des Idioms leichter verarbeiten als unrelatierte Varianten.

Insgesamt lassen sich die Ergebnisse aus **Studie II** in Modelle einordnen, die einen besonderen Status von MWE im Spracherwerb postulieren (Arnon & Christiansen, 2017; Grimm et al., 2017; Reuterskiöld & Van Lancker Sidtis, 2013; Sprenger et al., 2019). Die Ergebnisse, dass Kinder (1) Erwartungen für idiom-finale Wörter aufbauen, deutet darauf hin, dass sie Idiome als Mehrwort-Repräsentation im mentalen Lexikon abspeichern. Außerdem unterstützen die vorliegenden Ergebnisse die Annahme, dass Kinder (2) zusätzlich auch die Einzelwort-Bedeutungen verfügbar haben (vgl. Arnon & Christiansen, 2017; Grimm et al., 2017; Lin, 2018). Dennoch scheint die Einzelwort-Bedeutung nicht wie bei Erwachsenen automatisch nach dem Erkennen

des Idioms aktiviert zu werden. Daher liegt die Vermutung nahe, dass Kinder im Alter von zehn Jahren zwar sensitiv gegenüber Mehrwort-Konfigurationen sind, deren Repräsentation im mentalen Lexikon jedoch noch schwach ist. Daran anschließend stellt sich die Frage, ob sich die Repräsentation mit zunehmendem Alter verstärkt (**Studie III**)

7 Studie III

aus einem anderem Holz geschnitzt

7.1 Forschungsfrage

Studie III untersuchte die Verarbeitungsmechanismen von Idiomen im Erst- und Zweitspracherwerb. In dieser Studie untersuchten wir sprachlernende Kinder im Alter von dreizehn Jahren. Im Vergleich mit den Studien mit Erwachsenen und zehnjährigen Kindern (**Studien I** und **II**) können somit potentielle Unterschiede der mentalen Repräsentation von MWE im Entwicklungsverlauf festgestellt werden. Zusätzlich können anhand dieser Studie Verarbeitungsmechanismen zwischen Sprachlernenden einer ersten- und einer zweiten Sprache kontrastiert werden. Beide Lerner:innengruppen könnten sich potentiell in diesen Mechanismen unterscheiden, da sie unterschiedliche Erfahrungen mit Sprache im Allgemeinen gemacht haben.

Erwachsene Lernende einer Zweitsprache unterscheiden sich von sprachlernenden Kindern in der Regel insofern, dass sie bereits Vorerfahrung mit ihrer Erstsprache haben. Dies beinhaltet beispielsweise, zu erkennen, wo sich Wortgrenzen befinden, oder auch explizites Wissen um die Verwendung von Idiomen und anderen MWE. Aufgrund dieses metalinguistischen Wissens kann die Einzelwort-Bedeutung für diese Lerner:innengruppe eine andere Rolle spielen als für Lernende, die über wenig Vorwissen verfügen. Daher besteht die Möglichkeit, dass Zweitsprachler:innen MWE der zweiten Sprache von Beginn an in ihre einzelnen Bestandteile zerlegen und diese als Kombinationen einzelner Wörter erlernen. Arnon und Christiansen (2017) argumentieren, dass aufgrund von unterschiedlichen Lernmechanismen der beiden Lerner:innengruppen unterschiedliche mentale Repräsentation von MWE entstehen. Kinder erlernen MWE als untersegmentierte Einheiten und zerlegen diese Einheiten in die einzelnen Wortbestandteile. Im Gegensatz dazu erlernen Zweitsprachler:innen MWE, indem sie die einzelnen Bestandteile zu einer Einheit zusammenfügen.

Studie III untersuchte die Vorhersage und Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomen bei Erst- und Zeitsprachler:innen. Sie sollte die potentiell unterschiedlichen Verarbeitungsmechanismen in Bezug auf die Vorhersage des idiom-finalen Wortes und der Aktivierung dessen Bedeutung zwischen Erst- und Zweitspracherwerb kontrastieren. **Studie III** verwendete das gleiche Visual-World-Eye-Tracking-Experiment wie **Studie I** und **Studie II**. Somit kann der zeitlichen Verlauf der Antizipation des idiom-finalen Wortes während der

Präsentation des Idiomfragmentes und der Aktivierung der Einzelwort-Bedeutung relativ zu letzterem untersucht werden.

Studie III befasste sich daher mit folgenden Fragestellungen (Abbildung 24):

- (1) Bauen Lernende einer ersten und zweiten Sprache Erwartungen über idiom-finale Wörter auf? Beide Lerner:innengruppen zeigen einen erfahrungsabhängigen Verarbeitungsvorteil für MWE im Vergleich zu Nicht-MWE (z.B. Bannard & Matthews, 2008; Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin & van Heuven, 2011), welcher die Annahme einer Mehrwort-Repräsentation unterstützt (siehe Kapitel 1.1). In einem Visual-World-Experiment sollte **Studie III** untersuchen, ob beide Lerner:innengruppen das letzte Wort (*Sack*) innerhalb bekannter Idiome (*Hannes ließ die Katze aus dem ...*) vorhersagen. Wir erwarteten, dass dies in mehr Fixationen auf der korrekten, idiomatischen Vervollständigung erkennbar ist. Falls die Kohärenz der Repräsentation erfahrungsbasiert ist, sollten sich Erst- und Zweitsprachler:innen nicht qualitativ in der Vorhersage unterscheiden. Da wir die Kompetenz beider Lerngruppen nicht direkt vergleichen können, könnten demnach quantitative Unterschiede in der Stärke der Vorhersage auftreten.
- (2) Falls Lernende einer ersten und zweiten Sprache das idiom-finale Wort vorhersagen, aktivieren sie ebenfalls dessen Bedeutung? Positive Evidenz für diese Fragestellung würde sich in Fixationsunterschieden zwischen relativen (*Korb*) und unrelativen Wörtern (*Arm*) manifestieren. Aufbauend auf der Annahme von Arnon und Christiansen (2017) könnte man qualitative Unterschiede zwischen beiden Lerner:innengruppen diesbezüglich erwarten.



Abbildung 24: Bauen dreizehnjährige Kinder (L1) und erwachsene Zweitsprachler:innen (L2) eine Erwartung für das idiom-finale Wort auf und aktivieren sie währenddessen dessen Einzelwort-Bedeutung?

7.2 Manuskript: L1 and L2 learners keep their eyes on the prize: Eye-tracking evidence during idiom recognition

Introduction

Idioms are both formulaic sequences as well as figurative expressions, identities which both present their own implications and possible challenges for language processing. Although idiomatic meaning appears to challenge compositional theories of language processing, for adult native speakers of a language, the use and comprehension of idioms is a breeze. In fact, research shows that not only are idioms like *to let the cat out of the bag* and *to have butterflies in one's stomach* understood figuratively by native speakers online (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988), but there may even be an advantage in speed of comprehension when compared with similar novel expressions (e.g., Gibbs, 1980; McGlone, Glucksberg, & Cacciari, 1994; Swinney & Cutler, 1979). This advantage might stem from predictive properties based on their formulaic nature, in which idiomatic strings are recognized early during comprehension. However, the same speed advantages may also stem from a direct mapping of the form to the idiomatic or figurative meaning, bypassing compositional processes and associated word-meaning activation processes. The figurative meaning of the idiom (to reveal a secret) is thus associated with a fixed, word-like form rather than building the meaning of an idiom *like to let the cat out of the bag* online from its individual constituents (i.e., CAT and BAG). In spite of this lack of apparent contribution of the literal words to the idiomatic meaning, it seems that, in many instances, native speakers still activate individual literal constituents (e.g., Beck & Weber, 2016a; Cacciari & Tabossi, 1988; Titone & Connine, 1994) and even literal phrasal meaning (e.g., Beck & Weber, 2019; Holsinger & Kaiser, 2013) without hindering the speed of access to the idiomatic meaning.

However, these processing advantages in idioms are not to be underestimated, particularly when comparing first (L1) and second (L2) language learners. For a language learner to understand an idiom, they must first recognize the phrase as an idiom, have been exposed to the figurative meaning of the idiom, and must ignore the possible literal interpretation achieved through traditional meaning computation. Additionally, they must do so quickly if predictive processes are to offer the same processing advantages as are evidenced in native speakers. However, language learners have less experience with idiomatic expressions, which might negatively influence idiom recognition and speed of processing. Furthermore, the role of literal meaning might differ both between adult native speakers as well as between L1 and L2 learners of a language. Language learners may activate literal meaning differently than experienced users of a language based on language experience. Both learner groups may also differ because of different underlying mechanisms of acquisition (Arnon & Christiansen, 2017). For instance, children may acquire formulaic sequences as units, and literal word meanings might be less prominent in processing of those sequences, while adult L2 learners are more

likely to combine words within formulaic sequences and rely more on literal word meaning.

The current study addresses both the issue of prediction and literal activation. Specifically, the current study asks to what extent adult L1 predictive processes associated with formulaic sequences also apply to language learners, and what processes occur following idiom recognition, particularly concerning literal meaning activation. In order to address these questions, we will first review these processes in native speakers and then attend to the current research in language learners, specifically proficient non-native (L2) speakers and child (L1) speakers.

Predictive Processing in Idioms

The processing advantages of idioms compared to similar novel phrases are well-documented (e.g., Conklin & Schmitt, 2008; Gibbs, 1980; McGlone et al., 1994; Siyanova-Chanturia, Conklin, & Schmitt, 2011; Swinney & Cutler, 1979). Some theories of idiomatic processing account for this processing advantage by suggesting that this holistic access to figurative meaning is either partially, in the case of hybrid theories (e.g., Cacciari & Tabossi, 1988; Sprenger et al., 2006), or even fully responsible for fast idiomatic processing (e.g., Swinney & Cutler, 1979). Alternatively, processing advantages are often also attributed to familiarity with and the conventionalized form of idioms (e.g., Tabossi, Fanari, & Wolf, 2009). Providing evidence for the latter instance, reading studies using eye-tracking techniques have consistently shown processing benefits for idioms as well as other formulaic sequences. Underwood, Schmitt, and Galpin (2004) used a reading paradigm that measured fixations on individual words within formulaic sequences, including idioms embedded in unambiguous contexts. They found that participants fixated less while reading all types of formulaic sequences compared to novel phrases. They also found fewer fixations on terminal words in the phrases, which indicates prediction of these words based on the previous elements of the sequence. This result was confirmed by Siyanova-Chanturia and colleagues (2011), who found that idioms, used in both literal and figurative contexts, incited fewer and shorter fixations than similar novel phrases (e.g., *at the end of the day vs. at the end of the war*). However, they did not find any differences between idiomatic and literal readings of the idioms providing even further evidence that the processing benefit associated with idioms may indeed relate to the conventionalized form of the phrase.

Additionally, idioms pose an interesting case because of their distinct idiomatic and phrasal meanings, and these meaning differences are a common testing ground for predictive processes or anticipatory online processing that occurs during listening. There is evidence of qualitative differences in idiomatic processing compared to other types of formulaic language. Carrol and Conklin (2020) compared processing in idioms, collocations and binomials and found that specific idiomatic properties such as phrasal frequency, familiarity, and decomposability play an important role in the speed of processing and that familiarity with the phrase might explain shorter reading times associated with idioms, unlike other types of formulaic language. Additionally, ERP studies conducted by Vespignani, Canal, Molinaro, Fonda, and Cacciari

(2009) and by Canal, Pesciarelli, Vespignani, Molinaro, and Cacciari (2017) provide further evidence for such differences by comparing literal and figurative processing. The former study found that prediction within idioms is qualitatively different from predictive processing in novel, literal language based on the timing and distribution of N400 and P300 effects. The latter study added to this finding with evidence of effects in biased idiomatic sentences as early as the first word of the idiom, suggesting that prediction and possible reintegration processes occur immediately at recognition of the idiom, which may be sooner than in other types of formulaic language. Thus, the reasons for idiomatic advantages are still in need of further investigation.

While these well-established general speed and processing advantages in idioms indirectly indicate predictive mechanisms for final words as result of their familiar and conventionalized form, only few studies have looked directly at prediction within idioms. Cloze-probability tests show that native speakers can produce predictions for constituent words within idioms. In fact, most native speakers predict the final word *BAG*, when encountering the phrase *let the cat out of the . . .* (e.g., Beck & Weber, 2016b; Van Lancker Sidtis et al. 2015), and some idioms allow for prediction very early in the phrase. An online study by Kessler, Weber, and Friedrich (accepted for publication) confirms these processes and found evidence for predictive processing both with eye-tracking and ERPs. In their eye-tracking study, native speakers listened to a non-biasing sentence ending with a predictable idiomatic phrase missing its final constituent. Participants saw the correct final constituent of the idiom, a literal distractor, and two unrelated distractors. Fixation rates on the correct completion were significantly higher than to all other options early during listening. In their ERP data, in line with the results of Rommers, Dijkstra, and Bastiaansen (2013), Kessler and colleagues found that when the final constituents of these predictable idioms were replaced with incorrect and unexpected words (e.g., *let the cat out of the BASKET/STOMACH*), N400 effects occurred, yielding higher amplitudes for incorrect completions compared to the correct completion. These results are in line with previous work showing the sensitivity of the N400 to predictive properties, such as lexical violations like unexpected words (Federmeier & Kutas, 1999). Thus, regardless of the root of idiomatic processing advantages, adult native speakers consistently show evidence of prediction during idiomatic processing.

Predictive Idiomatic Processing in Language Learners Most comparable research on predictive processing of idioms in language learners has been conducted with proficient non-native speakers. However, the results of these studies appear to be mixed. In several studies, the same advantages for formulaic sequences or idiomatic phrases compared to novel phrases, such as faster reading times and fewer fixations during the phrase and on terminal words, have also been present when compared directly to native speakers (e.g. Conklin & Schmitt, 2008; Siyanova-Chanturia, Conklin & Van Heuven, 2011; Underwood et al., 2004). Conklin and Schmitt (2008) tested both figurative and literal uses of idioms in a self-paced reading study on L1 and L2 readers and found a speed advantage for formulaic phrases compared to novel phrases in both reader groups. In direct contrast to these findings, Siyanova-Chanturia,

Conklin and Schmitt (2011) found differences between L1 and L2 readers in eye-movements depending on the phrase-type; unlike in their L1 reader group, idiomatic readings did not produce any speed or fixation differences from novel phrases. Furthermore, idiomatic readings produced a processing cost for L2 readers not present in L1 readers. However, the results of the study conducted by Siyanova-Chanturia, Conklin and Van Heuven (2011) suggest that the inconsistencies found between studies may be due to experience with the language. Namely, in their study on frequency in binomial processing (e.g., *bride and groom* vs. *groom and bride*), phrasal frequency interacted with proficiency in a manner suggesting that only known, memorized forms can be subject to phrasal frequency effects and this experience is necessary for processing effects to be visible in L2 readers.

Beyond the limited L2 research on the possible advantages of idiomatic and formulaic phrases, there is little research directly on L2 prediction in idiomatic processing. It's clear, however that L2 learners are also able to predict during processing, though these processes may be more limited than in L1 speakers (see Kaan, 2014 for an overview). For example, there is growing evidence that L2 speakers use contextual information to form predictions of upcoming words differently from L1 speakers in literal language (e.g., Dussias, Valdés Kroff, Guzzardo Tamargo, & Gerfen, 2013; Ito, Martin, & Nieuwland, 2017). For instance, Ito and colleagues (2017) conducted an ERP study to test prediction of form and meaning in highly predictable sentence-endings (e.g., *The student is going to borrow a . . . book.*) on Spanish-English bilinguals. While the authors found reduced N400 amplitudes for words semantically related to predicted words, this reduction was not dependent on predictability. However, as also suggested by the reading and eye-tracking studies discussed previously, proficiency may modulate prediction abilities rather than distinguishing them from native processing. More studies looking at prediction in idioms are needed in order to compare L1 and L2 predictive processes in idiom comprehension. To our knowledge, there are not yet comparable eye-tracking or ERP studies on prediction within idioms in the L2.

For L1 language learners, there is some evidence that children show a processing advantage for idioms over novel expressions. In a study by Qualls, Treaster, Blood, and Hammer (2003), 10-year-old children read correct idioms with high, moderate, or low familiarity. These were compared to control phrases that were manipulated idiom forms in which either the initial or the final word was replaced with an unrelated word (e.g., *put/stomp their heads together*; *go around in circles/trouble*). For each sentence, children had to judge whether it was an idiom or not. Children showed shorter response latencies for idiomatic than for control expressions but only when the final word was substituted. Furthermore, children showed shorter response times for highly familiar idioms reflecting the L1 adult data showing that higher familiarity leads to faster processing (Carrol & Conklin, 2020). Thus, similar to second language learners, children might show a stronger processing advantage if they are familiar with the canonical form of the idiom. Taken together, idioms also seem to have a pri-

vileged status for children, but direct evidence for predictive processing within idioms is missing.

L1 language learners are also able to perform predictions about upcoming words in literal language (e.g., Mani & Huettig, 2012, 2014). Mani & Huettig (2014) conducted a visual world experiment in which children listened to sentences like *the boy eats the big cake* while at the same time seeing pictures of eatable and non-eatable objects on the screen. Children showed more fixations towards eatable objects shortly after hearing the verb *eat* (Mani & Huettig, 2014). In another experiment, Mahler & Chenery (2018) looked at naming latency. Children were presented auditorily with high and low cloze sentences (*The dog buried the bone/stick*) and had to repeat the final word. They were faster at repeating words in high (*bone*) compared to low cloze words (*stick*) suggesting a benefit for predictable items. Still, more evidence is needed to explain predictive processing mechanisms in children, but overall research supports general predictive abilities.

Literal Activation in Idiomatic Constituents

When idioms are highly predictable, and an idiomatic meaning is activated, ongoing literal computation and therefore literal word activation seems unnecessary. However, evidence has repeatedly shown that online literal activation occurs in adult native speakers. In addition to consistent semantic priming effects for literal constituents of idioms (e.g. *kick the bucket primes PAIL*, Beck & Weber, 2016; Cacciari & Tabossi, 1988), eye-tracking also provides further evidence of such activation. For example, Holsinger (2013) tested literal word activation in a visual world paradigm. While listening to idioms (*kick the bucket*), participants saw four words on the screen: one related to the figurative meaning (DEATH), one literally related to an idiom constituent (FOOT) and two unrelated distractors (TRIANGLE, ANIMAL). Despite increased fixations towards the figuratively related word, participants also showed more fixations towards literally related words compared to unrelated distractors. Thus, eye-tracking also shows evidence of literal constituent meaning activation.

With regard to literal word activation during processing of highly predictable idioms, evidence is mixed. In a production study, Sprenger, Levelt, and Kempen (2006) found that speakers preparing to produce an idiom-final word (e.g., English translation: *Jan walked against the lamp*) produce words semantically related (*candle*) to the respective idiom-final word faster than unrelated words. In contrast, there is some evidence that literal constituents are not always activated during idiom processing. Rommers, Dijkstra, and Bastiaansen (2013) conducted a semantic expectancy paradigm using ERP in which the final word of Dutch idioms in biasing contexts was highly predictable, but the idiom was presented either with its correct completion or replaced by either a word semantically related to the correct completion or an unrelated word (e.g., English translation: *After many transactions the careless scammer eventually walked against the LAMP/CANDLE/FISH yesterday.*). Neither the N400 nor the P600 effect showed sensitivity towards semantic relatedness, which suggested that literal activation of the final constituent did

not occur. In contrast, a recent study of native German speakers using a similar paradigm by Kessler and colleagues (accepted for publication) challenges this finding when testing auditorily instead of visually presented idioms (*Hannes let the cat out of the BAG/BASKET/ARM*). The authors found evidence of semantic expectancy within idioms as indexed by reduced N400 amplitudes for related compared to unrelated words in neutral contexts during listening. In a second experiment, Kessler and colleagues (accepted for publication) adapted the paradigm used by Holsinger (2013) and tested semantic expectancy in a visual world paradigm. Participants listened to idioms missing the final, highly predictable constituent word while viewing the four possible words adapted from earlier studies on the screen. Increased fixations towards the correct completion around 460 ms prior to the offset of the acoustic stimulus indicated prediction of the correct idiom form. Simultaneously, there were more fixations to related than to unrelated distractors. Therefore, in spite of somewhat mixed evidence, literal word meaning is often co-activated during predictive processing after recognition of the idiom.

Literal Constituent Activation in Language Learners For language learners, it is possible that literal meaning has a different status than figurative meaning than for adult speakers of a language. And because of possibly different acquisition mechanisms involved (Arnon & Christiansen, 2017), L1 and L2 language learners might also differ in their reliance on individual constituents and thus, their literal word meaning activation. Although research with children has focused on the understanding and interpretation of figurative expressions (e.g., Bernicot et al., 2007), the topic of literal meaning activation has not yet been explored. However, this topic has been widely addressed in L2 learners.

For non-native speakers, a possible difference between L1 and L2 in literal activation may be its status as highly salient in comparison to figurative language as a whole (e.g., Cieślicka, 2006; Giora, 1997). This line of research suggests that not only does literal activation occur even when a figurative meaning is likely (i.e., in a biasing context), but activation of literal word and even phrasal meaning is obligatory for idioms as a general processing strategy (i.e., for *kick the bucket*, both the word BUCKET and the literal phrasal meaning of striking a pail with one's foot are activated automatically). In L2 cross-modal priming studies, both Cieślicka (2006) and later Beck and Weber (2016a) found that the literal meaning of idiom-final words was activated (e.g., *kick the bucket* activated PAIL), and these priming effects appeared to be stronger than the priming of the figurative meaning (e.g., DIE). However, Cieślicka (2006) did not include native speakers as a control group in her study. In contrast, Beck and Weber (2016a), and an additional study also using both speaker groups by van Ginkel and Dijkstra (2019) confirmed that there were not differences between L1 and L2 activation of literal meaning. Beck and Weber (2019; submitted) also looked into the special status of literal meaning in L1 and L2 speakers on a phrasal level in two self-paced reading studies. Readings of idiomatic sentences were biased literally or figuratively, and sentence endings either followed this expectation or went against them. While L1 readers (Beck & Weber, submitted) showed evidence that context

influenced literal and figurative readings, L2 readers (Beck & Weber, 2019) showed evidence only on figurative readings. However, there was no evidence that literal meaning is obligatorily computed in L2 speakers compared to L1 speakers, so phrasal differences in literal processing also remain unsubstantiated. Overall, while L2 speakers do activate literal constituent meaning during processing, it's unclear whether literal constituent words play more importance in L2 compared to L1 processing.

Research Questions

We conducted two visual world eye-tracking experiments using L1 and L2 language learners to address the questions:

- (1) Do L1 and L2 learners predict words within an idiom during listening?
- (2) Are literal words activated during this process, and if so, when?

By looking at both L1 and L2 language learners of German, we aim to fill in largely unexplored gaps on idiom processing in the progression of language development and look at processing differences that might result from different acquisition mechanisms.

In the current study, we use the visual world paradigm to answer our research questions. The visual paradigm is suitable to answer both research questions because it is sensitive towards predictive mechanisms (Altmann & Kamide, 1999, 2007; Kamide, Altmann, & Haywood, 2003) as well as semantic competition (Huettig & Altmann, 2005). Furthermore, this paradigm has been previously used to study lexical access in idioms (Holsinger, 2013; Kessler et al., accepted for publication). Literal word activation can be measured by looking at co-activation of semantically related words. Here, we used the experimental design of Kessler and colleagues (accepted for publication): participants listened to neutral sentences containing idioms without their final word (e.g., German: *Hannes ließ die Katze aus dem . . .* /English translation: Hannes let the cat out of the . . .). We chose idioms in which the final word was highly predictable. At the same time participants viewed words displayed on the screen (German/English translation) that were (a) correct completions of the idiom (*SACK*/*BAG*), (b) distractors semantically related to the correct completion (*KORB*/*BASKET*), and (c) two unrelated distractors (*ARM*/*ARM*, *BAUCH*/*STOMACH*).

In answering question (1), we hypothesized that prediction will be seen in many and early fixations towards the correct completion that is not presented auditorily. If evidence of an idiomatic processing advantage for both L1 and L2 listeners extends to prediction, then we should find looks to the correct completion before or around the offset of the final auditorily presented word. Concerning question (2), increased fixations towards related compared to unrelated distractors will index effects of semantic competition and thus, literal word activation. We expect the results of previous priming experiments to extend to the current experiment, and L2 par-

ticipants are expected to look more towards related compared to unrelated distractors. If there is a literal priority for L2 listeners, the amount of these fixations may be greater than in the L1 listeners groups examined in Kessler and colleagues (accepted for publication). As there is no evidence to the contrary for L1 language learners, we expect that they, too, will activate the literal meaning of idioms. Though, if only L2 learners show activation of the literal meaning, this would imply that L1 and L2 learners store and access idioms differently.

Experiment 1: L1 Language Learners

Methods

Participants Twenty-six 7th graders (17 female; mean age = 12;8 years) that we recruited from local schools participated in the experiment. L1 learners from this age group were chosen based on the great probability of exposure to the idioms included in the study while still early enough in language development to consider them learners. Parents gave informed consent prior to the experiment. All participants were native, monolingual speakers of German and had no history of hearing disorders and normal or to normal corrected vision. As compensation, the children received a voucher for a local toy store. The design of the study was approved by the local ethical committee (reference number: 2016/1027/22).

Materials Experimental items from Kessler and colleagues (accepted for publication) were used in the current study. These items consisted of twenty well-known German idioms (see Appendix) embedded in sentences using the following structure (see Table 10): (i) a person carrying out the action of the sentence, (ii) a sentence body that originated from a German idiom and (iii) the final target word of the idiom (which was not presented auditorily). Sentences were recorded using a native speaker of German, and the final target word was cut from the recording. Participants heard each idiomatic sentence once, during which four visual words were presented on a computer screen. The words presented visually were categorized as: (1) Correct Completion: correct completion of the idiomatic phrase, (2) Related Distractor: semantically related word of the correct completion, (3&4) Distractors Unrelated 1 and Unrelated 2: semantically unrelated to the correct completion. Unrelated 1 and Unrelated 2 words were matched word pairs from Correct Completions and Related Distractors used from other experimental sentences in the experiment (avoiding phonological overlap). All presented words matched the grammatical gender expected from the preceding sentence context.

Tabelle 10: German example sentence for types 1-4 with english equivalent.

(i) Person	(ii) Sentence Body	(iii) Target words			
		(1) Correct	(2) Related	(3) Unrelated 1	(4) Unrelated 2
Hannes	ließ die Katze aus dem	Sack	Korb	Bauch	Arm
<i>Hannes</i>	<i>let the cat out of the</i>	<i>bag</i>	<i>basket</i>	<i>stomach</i>	<i>arm</i>

Semantic relatedness between correct and related words was confirmed by comparing semantic spaces using the R package LSAfun (Günther, Dudschig, & Kaup, 2015). The semantic similarity between correct and related words was significantly higher than between correct words and both unrelated words (Wilcoxon signed rank test: unrelated1 $Z = 189$, $p < .001$; unrelated2 $Z = 185$, $p = .002$). There was no difference in semantic similarity between correct words and both unrelated words ($Z = 75$, $p = .28$). Target words were visually presented in white font (Arial, font size 28) on a grey background (see Figure 25), and the position of displayed words was counterbalanced across items and participants. The inter-trial interval was 1500 ms, and the presentation of a fixation cross for 500 ms marked the beginning of a new trial. Next, the set of four words were displayed on the screen, and the presentation of the audio stimuli began 2150 ms later. The order of the trials was randomized.

🔊 “Hannes ließ die Katze aus dem...”

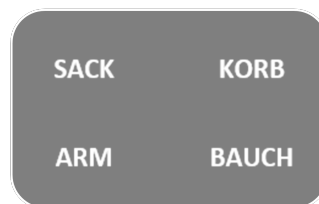


Abbildung 25: Example for visual display and auditory stimuli used in the experiment.

Procedure The experiment was conducted either at local schools or in the lab of the University in a single session. Participants were given both written and oral instructions for the experimental task. They were instructed to sit in a relaxed position and fixate on the screen during the task. The auditory stimuli were presented via closed headphones. Participants were asked to listen to each sentence and decide which of the visually presented words was the best completion for the idiom by saying their choice out loud. Participant responses were noted by the experimenter. After their oral response, participants were instructed to press a button in order to continue to

the next trial. Prior to the experimental task, a 5-point eye calibration was carried out for each participant followed by a practice block consisting of five trials. Fixations were recorded using a portable Tobii eye-tracker with a sampling size of 60 Hz. The eye-tracking experiment took around 20 minutes in total including instructions, eye-calibration and the experimental task; which lasted about 10 minutes on its own.

Analysis In order to analyze fixations on the four visually presented words, we divided the screen into four areas of interest (AOIs). The time window of the analysis was aligned to the offset of each audio stimulus (offset = 0 ms). Only items with correct responses were included in the analysis (i.e. trials in which the participants chose the correct completion of the idiom), we assume that correct answers indicate knowledge of the idiom forms. Values of the two unrelated distractors were averaged. We used the fixation data to conduct two analyses to determine 1) the point at which participants reliably looked to the correct response, and 2) the time-course of fixations to each response type. In our first analysis, we conducted running t-tests ($p < .01$) comparing fixations towards correct completions and unrelated distractors at succeeding measurement points (every 16.67 ms resulting from 60 Hz sampling rate of the eye-tracker). These tests were used to determine whether anticipation occurred by identifying a point of divergence between fixations on correct completions compared to unrelated distractors. In the second analysis, we conducted a Growth Curve Analysis (GCA) with orthogonal polynomials (Mirman, Dixon, & Magnuson, 2008) to compare the amount and time-course of fixations towards Distractor Types (related and unrelated). As the starting point of the GCA time window, we chose the start of the anticipation for a duration of 1200 ms.

Results

In 91.15% of the trials, participants responded correctly. In our first analysis, a running t-test showed that fixations on correct completions and unrelated distractors diverged around 128 ms before the stimulus offset and therefore indicate anticipation. The overall fixation pattern is displayed in Figure 26, Panel (A). The vertical, dashed line visualizes the start of the anticipation. In our second analysis, we modeled fixation patterns for related and unrelated distractors with third-order orthogonal polynomials. The modeled data are also depicted in Figure 26, Panel (B). Estimated parameter terms of Distractor Type are summarized in Table 11.

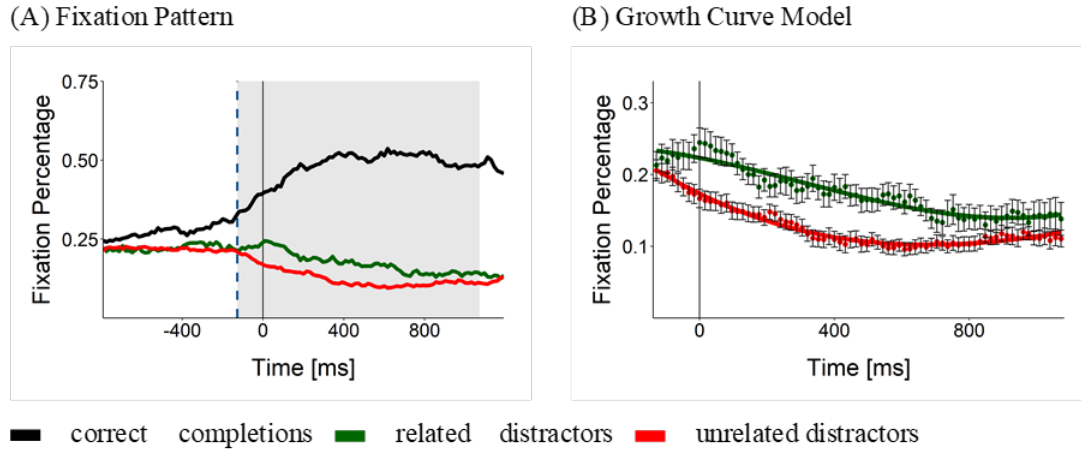


Abbildung 26: (A) Fixation patterns for L1 Children: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line).

We find a significant effect on the intercept that indicates more fixations on related than unrelated distractors. The quadratic term indicates that the curves for both distractors differ in their central inflection. Thus, the peak for unrelated distractors is more negative than for related distractors. The effect on the cubic term suggests that towards the extremities of the model inflections for both distractors differ, resulting in convergence of both curves. Together, these curves indicate that while participants generally fixate more on related distractors overall, this difference is largest over central parts of the model.

Tabelle 11: Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).

Term	Estimate	Standard Error	t	p
Intercept	-0.02	0.00	-4.78	.00
Linear	0.02	0.03	0.80	.42
Quadratic	0.03	0.00	6.35	.00
Cubic	-0.01	0.00	-2.93	.00

Discussion

In Experiment 1, we tested idiom processing in first language learners (7th graders). They showed relatively good performance on idiom recognition as displayed by correct responses in more than 90 % of the trials. This recognition of the idiom-final word was also observable in predictive fixations of the correct completion. Thus, children not only show a processing benefit for the recognition of highly familiar idioms (Qualls et al., 2003), but also fast online prediction of idiom-final constituents. This is in line with other studies showing the abilities of children to predict upcoming words in sentences (e.g., Mani & Huettig, 2012, 2014). However, unlike previous studies, the current study examines highly predictable idioms rather than using semantic cues for prediction (e.g., EAT increases looks to CAKE), and indicates an additional type of prediction in L1 children. We interpret these results as positive evidence for our first research question concerning prediction during listening, as seen via early anticipation of the idiom-final word.

Regarding lexical activation, children indeed show more fixations on related than unrelated distractors, and we interpret this as positive evidence for literal activation in L1 child listeners. Lexical activation seems to be strongest shortly after anticipation of the correct idiom-final word.

Experiment 2: L2 Language Learners

Methods

Participants We recruited 33 non-native speakers of German (English L1, 12 female; mean age = 25.7 years) from the University of Tübingen for the experiment. Participants were living in Germany at the time of the experiment and were asked only to participate if they had at least a B2 level of proficiency (“Common European framework of reference for languages,” 2001). Of the 33 participants, only 26 were included in the analysis because five participants fell below the criteria for appropriate language proficiency based on task performance (correct responses < 40 %), LexTale score (scores < 50/100), and the language background questionnaire (years of study < 2, average self-rated proficiency < 3 on a scale of 1-7). The 26 participants included in the analysis had an average of 5.05 ($SD = 2.34$) years of German instruction, an average LexTale score of 68.62 out of 100 ($SD = 9.39$), and an average self-reported proficiency (averaged across reading, writing, speaking and listening) of 5.01 ($SD = 0.94$) on a scale of 1-7. Participants had no history of hearing disorders and normal or to normal corrected vision.

Materials The same materials from Experiment 1 were used in Experiment 2.

Procedure The procedure was the same as in Experiment 1 with the addition of a LexTale vocabulary test (Lemhöfer & Broersma, 2012) prior to the eye-tracking task and a language background questionnaire at the end of the experimental session. These additional tasks were used as measures of language proficiency. The experiment lasted about 30 minutes.

Analysis The analyses were conducted following the same procedure as Experiment 1. Again, only items with correct responses were included in the analysis, and fixations were used to determine 1) the point at which participants reliably looked to the correct response via t-tests, and 2) the time-course of fixations to each response type via Growth Curve Analysis.

Results

On average, participants responded in 64.81 % of the trials with the correct completion of the idiom. The average responses to the related competitor were 17.88 % In our first analysis, a running t-test revealed a reliable difference in fixations towards correct completions and unrelated distractors starting around 48 ms prior to the stimulus offset. The overall fixation pattern is displayed in Figure 27, Panel (A). The vertical, dashed line visualizes the start of the anticipation. In our second analysis, differences between fixation patterns for related and unrelated distractors were again modeled using a Growth Curve Analysis (GCA) (Mirman et al., 2008) with third-order orthogonal polynomials because by visual inspection of the time-course, two bends were observable in the curve. The modeled data are also depicted in Figure 27, Panel (B). Estimated parameter terms of Distractor Type are summarized in Table 12.

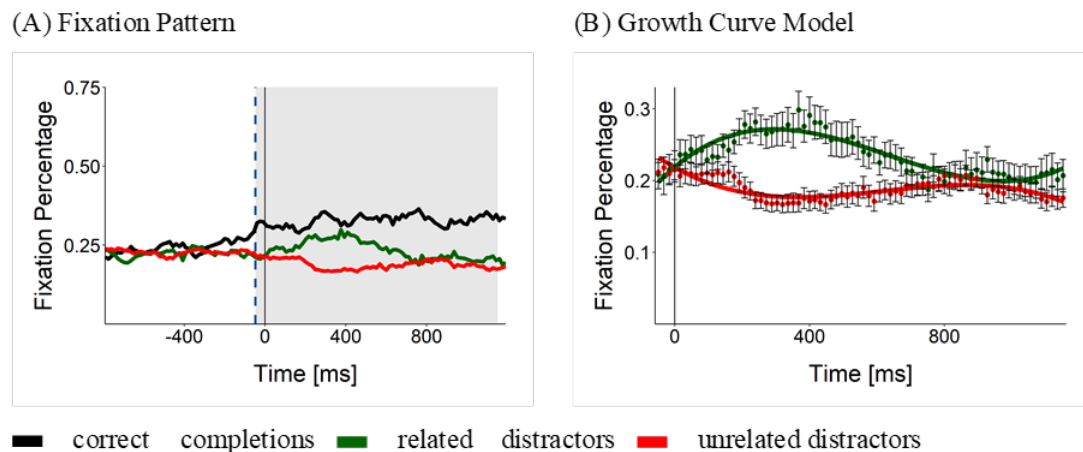


Abbildung 27: (A) Fixation patterns for L2 Adults: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line).

There was a significant effect of Distractor Type on all terms but the linear term (see Table 12). The main effect of Distractor Type indicates overall more fixations towards related than unrelated distractors. The quadratic term reflects that the curves for both distractors differ in their inflection around the center. Together with the depiction of the model in Figure 27, this indicates that the curve for related distractors peaks higher than for unrelated

distractors. Significant effects on the cubic term reflect differences at the extremities of the curve.

Tabelle 12: Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).

Term	Estimate	Standard Error	t	p
Intercept	-0.02	0.00	-4.16	0.00
Linear	0.04	0.03	1.06	0.29
Quadratic	0.05	0.01	8.39	0.00
Cubic	-0.07	0.01	-11.37	0.00

Discussion

In Experiment 2, we tested idiom processing in second language learners of German (English L1). Their performance, averaging about 65%, reflected moderate figurative language proficiency, which was lower than the proficiency of the L1 learner group in Experiment 1. However, when idioms were recognized, there was also evidence of prediction. Although looks to the correct completions only occurred 48 ms before the offset of the phrase, the programming of a saccade following a critical word takes approximately 200 ms (Saslow, 1967), a delay which suggests that recognition occurred prior to the anticipation point of 48 ms identified here. These results are in line with studies suggesting that there may be processing advantages associated with idioms, even for non-native learners (e.g., Conklin & Schmitt, 2008; Underwood et al., 2004). We interpret these results as positive evidence for our first research question, namely that L2 learners exhibit predictive processing of idioms during listening.

Concerning literal activation, more fixations on related distractors compared to unrelated distractors suggest that, like L1 learners, L2 learners activated the literal meaning of constituent words. This pattern of results is consistent with literature suggesting strong activation of literal constituents in L2 idiom processing (e.g., Beck & Weber, 2016a; Cieřlicka, 2006), especially considering that this literal activation continues to increase even after correct recognition of the final constituent. While this does not provide evidence of an L2 literal priority directly (e.g., Cieřlicka, 2006), it is further confirmation of the importance of literal word meaning even in figurative phrases. It should also be noted that activation occurs even without the auditory cue of the final word, suggesting that this activation is also predictive in nature and strengthens our conclusions on prediction in L2 listening. Overall, these results indicate that L2 learners also display patterns of predictive processing and semantic activation early on in idiom processing.

General Discussion

Overall, the results from Experiments 1 and 2 show evidence that both groups of L1 and L2 language learners are able to predict idiom-final constituents during listening in addition to activating the literal meaning of these constituents, yet there were differences related to language proficiency when comparing both groups to native adult speakers. Regarding the task itself (i.e. identifying the correct completion of each idiom), we obtained graded effects in accuracy and timing of eye movements. In a previous study, adult native speakers determined idioms' final word at ceiling level with 99 percent correct responses (Kessler et al., accepted for publication). In the present study using the same materials and procedures, L1 language learners determined 91 percent of the idiom completions correctly, while L2 language learners only determined 65 percent correctly. Similarly, overall looks to the correct response were higher for the L1 language learners than for the L2 language learners (see Figures 26 and 27 Panel A). Furthermore, adult native speakers directed their gazes earlier to the correct completion (ca. 460 ms before its onset) than L1 language learners (ca. 130 ms) and L2 language learners (ca. 50 ms), respectively. Graded accuracy in fixating and recognizing the correct idiom completions might simply reflect that adult native speakers had the most experience with the language, and L1 children had more experience than our L2 learners did. Differences in proficiency and therefore experience and familiarity with respective idioms, may impact stored frequency information (e.g., Lee, Hsin-Yi Lu, & Garnsey, 2013; Qualls et al., 2003; Siyanova-Chanturia, Conklin & Van Heuven, 2011; Tabossi et al., 2009) resulting in differences in the efficiency of predictions based on the same phrases.

Like native adults (Kessler et al., accepted for publication), language learners appeared to activate literal constituents of idioms for predictive processing. All three groups showed more fixations towards related distractors than compared to unrelated distractors. While predicting the correct completion *bag* of the idiom onset *to let the cat out of the ...*, participants' fixations were biased towards *BASKET* (being a semantic associate of *bag*). They appeared to predict single word constituents, which made spreading activation to semantic associates of these single constituents possible. Also similar to the timing obtained for fixations to the correct completions, the timing of fixations towards related distractors (relative to the anticipation of the correct completion) was graded between the three groups. Again, this outcome might relate to differences in proficiency. It might suggest less rapid spreading activation in language learners compared to proficient native adults (e.g., Gernsbacher & Faust, 1991).

Looks towards related distractors also revealed qualitative differences in processing between L1 and L2 speakers of German. While looks to semantically related competitors linearly decreased in time in L1 learners (see Figure 26, Panel B) and in L1 adults (Kessler et al., accepted for publication), they initially increased for L2 learners (see Figure 27, Panel B). In terms of qualitative differences, it has been suggested that literal meaning play a more important role in

L2 idiom processing compared to L1 idiom processing (e.g., Cieslicka, 2006; Giora, 1997), though evidence has been mixed before this study (e.g., Beck & Weber, 2016a; van Ginkel and Dijkstra, 2019). Responsible for the current result may indeed be an underlying learning mechanism that is more additive in nature for L2 learners compared to L1 learners (Arnon & Christiansen, 2017), which might account for a stronger reliance on constituent meanings. According to this line of argumentation, increased looks to semantically related competitors in L2 might reflect obligatory activation of literal meaning rather than a more automatic or shallow activation that may be displayed in both L1 groups (e.g., Peterson, Burgess, Dell, & Eberhard, 2001). Following this line of argumentation, single word meanings play a different role in online processing.

Alternatively, however, this outcome might also still be a result of graded language proficiency. Participants of the least proficient group (L2 learners) might either be less confident in their choice of the idiom-final word and/or weaker in inhibiting automatic activation of semantic associates. Furthermore, increased looks to the related distractors may be a result of second-guessing their given responses as these choices sometimes reflect a likely response (e.g., a cat may be let out of a *BAG/SACK* or a *BASKET/KORB*). In the case of the latter possibility, the visual presence of both the correct final constituent as well as its semantic associate may cause a greater challenge for non-native speakers, who are less skilled at inhibiting related but inappropriate information (e.g., Gernsbacher & Faust, 1991). As the final idiomatic constituent as well as its related distractor were available on the screen during the listening process, the current study is limited in its ability to clearly tease apart those alternative conclusions.

In conclusion, our data shows evidence that L1 and L2 language learners are able to predict during listening and activate the meaning of literal constituents. Whereas prediction indicates that both sets of learners interact with the idiom as a unit, literal constituent activation indicates that the parts of the unit remain available. Differences in timing between these groups and adult native speakers (Kessler et al., accepted for publication) suggest that proficiency plays a role in both the timing of anticipation during listening as well as the pattern of literal activation. Additionally, in line with growing evidence that L1 and L2 idiomatic processing are similar (e.g., Beck & Weber, 2016a, 2019; Conklin & Schmitt, 2008; Tabossi et al., 2009; van Ginkel & Dijkstra, 2019), we propose that the differences found between our L1 and L2 learners here may be an interaction between proficiency and L2-specific challenges rather than fundamental differences in storage and processing. Future studies must include more varied proficiencies in learner groups in order to better understand how prediction and activation evolves and use additional methods such as ERP (e.g., Kessler et al., accepted for publication) in which literal activation can be more clearly separated from other possibilities such as spreading activation by excluding target words from the experimental trials.

Acknowledgements

We would like to thank Stacie Boswell, Claudia K. Friedrich, Birte Herter, Kerstin Jendrysik, Thanh Lan Truong, Andrea Weber, Miriam Wild. We also warmly thank all participants. This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 75650358 – SFB 833.

References

- Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.
- Altmann, G. T. M., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, *57*(4), 502-518.
- Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, *9*(3), 621-636.
- Beck, S. D., & Weber, A. (2016a). Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, *7*, 1350.
- Beck, S. D., & Weber, A. (2016b). English-German Database of Idiom Norms (DIN). [Data file]. <https://uni-tuebingen.de/fakultaeten/philosophische-fakultaet/fachbereiche/neuphilologie/englisches-seminar/sections/english-linguistics/lehrstuhl-prof-dr-andrea-weber/staff/sara-d-beck/din/>
- Beck, S. D., & Weber, A. (submitted). Context and Literality in Idiom Processing: Evidence from Self-paced Reading.
- Beck, S. D., & Weber, A. (2019). Context matters, figuratively, for L2 readers: Evidence from self-paced reading. In *Proceedings of The 11th International*, *1*, e057.
- Bernicot, J., Laval, V., & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, *39*(12), 2115-2132.
- Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, *27*, 668–683.
- Canal, P., Pesciarelli, F., Vespignani, F., Molinaro, N., & Cacciari, C. (2017). Basic composition and enriched integration in idiom processing: An EEG study. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(6), 928–943.
- Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, *63*(1), 95-122.
- Cieślicka, A. B. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22*(2), 115–144.

- Common European framework of reference for languages: Learning, teaching, assessment.* (2001). Cambridge, U.K.: Press Syndicate of the University of Cambridge.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89.
- Dussias, P. E., Valdés Kroff, J. R., Guzzardo Tamargo, R. E., & Gerfen, C. (2013). When gender and looking go hand in hand: Grammatical gender processing in L2 Spanish. *Studies in Second Language Acquisition*, *35*(2), 353–387.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, *41*(4), 469–495.
- Gernsbacher, M. A., & Faust, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(2), 245–262.
- Gibbs, R. W. Jr. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition*, *8*(2), 149–156.
- Giora, R. (1997). Understanding figurative and literal language: The graded salience hypothesis. *Cognitive Linguistics*, *8*(3), 183–206.
- Günther, F., Dudschig, C., & Kaup, B. (2015). LSAfun-An R package for computations based on Latent Semantic Analysis. *Behavior Research Methods*, *47*(4), 930-944.
- Holsinger, E. (2013). Representing idioms: syntactic and contextual effects on idiom processing. *Language and Speech*, *56*(3), 373-394.
- Holsinger, E., & Kaiser, E. (2013). Processing (non)compositional expressions: Mistakes and recovery. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(3), 866–878.
- Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23-B32.
- Ito, A., Martin, A. E., & Nieuwland, M. S. (2017). On predicting form and meaning in a second language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(4), 635–652.
- Kaan, E. (2014). Predictive sentence processing in L2 and L1. *Linguistic Approaches to Bilingualism*, *4*(2), 257–282.
- Kamide, Y., Altmann, G. T. M., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, *49*(1), 133-156.
- Kessler, R., Weber, A., & Friedrich, C. K. (accepted for publication). Activation of literal word meanings in idioms: Evidence from Eye-tracking and ERP experiments. *Language and Speech*.
- Lee, E.-K., Hsin-Yi Lu, D., & Garnsey, S. M. (2013). L1 word order and sensitivity to verb bias in L2 processing. *Bilingualism: Language and Cognition*, *16*(4), 761–775.
- Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343.

- Mahler, N. A., & Chenery, H. J. (2018). A developmental perspective on processing semantic context: Preliminary evidence from sentential auditory word repetition in school-aged children. *Journal of Psycholinguistic Research*, *48*(1), 81-105.
- Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake-but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(4), 843-847.
- Mani, N., & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, *126*, 264-279.
- McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, *17*(2), 167–190.
- Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*(4), 475-494.
- Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(5), 1223–1237.
- Qualls, C. D., Treaster, B., Blood, G. W., & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: a reaction time study. *Journal of Communication Disorders*, *36*(4), 245-261.
- Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25*(5), 762–776.
- Saslow, M. G. (1967). Effects of components of displacement-step stimuli upon latency for saccadic eye movement. *Journal of the Optical Society of America*, *57*(8), 1024–1029.
- Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011): Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, *29*(2), 72–89.
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011): Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*(3), 776–784.
- Sprenger, S., Levelt, W., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, *54*(2), 161-184.
- Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*, 523–534.
- Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, *37*(4), 529–540.
- Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literalness. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(5), 1126–1138.

- Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences* (pp. 155–172). Amsterdam: John Benjamins.
- Van Ginkel, W., & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, *3*, 1–17.
- Van Lancker Sidtis, D., Cameron, K., Bridges, K., & Sidtis, J. J. (2015). The formulaic schema in the minds of two generations of native speakers. *Ampersand*, *2*, 39–48.
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, *22*(8), 1682–1700.

7.3 Schlussfolgerungen

Studie III untersuchte Verarbeitungsmechanismen von Idiomen bei Lernenden der Erst- und der Zweitsprache. In Visual-World-Eye-Tracking-Experimenten untersuchten wir, ob dreizehnjährige Kinder und Zweitsprachler:innen im jungen Erwachsenenalter (1) idiom-finale Wörter korrekt vorhersagen können und ob sie (2) deren Bedeutungen ko-aktivieren. In dieser Studie konnten wir die beiden Lerner:innengruppen nicht direkt vergleichen, daher werden lediglich deskriptive Vergleiche dargestellt. In Kapitel 8 werden zusätzlich Vergleiche der Verarbeitungsmechanismen von dreizehnjährigen Kinder mit denen von zehnjährigen Kindern und erwachsenen Muttersprachler:innen angestrebt.

Bezüglich der Frage (1) zeigten beide Lerner:innengruppen antizipatorische Fixationen auf das korrekte, idiom-finale Wort. Dies unterstützt die Annahme, dass Sprachlernende Informationen zu Konfigurationen von Wörtern innerhalb wiederkehrender, fester Ausdrücke abspeichern. Qualitativ gibt es keine Hinweise darauf, dass sich beide Lerner:innengruppen in der Art der Vorhersage unterscheiden. Allerdings fixierten dreizehnjährige Kinder die korrekte Vervollständigung des Idioms früher und auch stärker als Zweitsprachler:innen (junge Erwachsene). Wir vermuten, dass der Erfahrungsunterschied mit den Ausdrücken im Erst- und Zweitspracherwerb diesen Effekt erklären könnte, obwohl die Zweitsprachler:innen eine sehr hohe Sprachkompetenz in ihrer Zweitsprache aufwiesen. Obwohl wir die Spracherfahrung nicht direkt zwischen beiden Lerner:innengruppen vergleichen können, könnte man annehmen, dass Kinder bisher insgesamt mehr Exposition mit ihrer Erstsprache hatten als Zweitsprachler:innen mit ihrer Zweitsprache. Daher könnten Kinder auch mehr Erfahrung mit den in **Studie III** verwendeten Ausdrücken haben. Diese Annahme kann durch die behavioralen Daten unterstützt werden. Zweitsprachler:innen wählten weniger häufig die korrekte Vervollständigung der Idiome aus als Erstsprachler:innen (65 % vs. 91 %). In zukünftigen Studien könnte innerhalb der Lerner:innengruppen untersucht werden, ob Erfahrungsunterschiede tatsächlich zu früheren und stärkeren Vorhersagen des idiom-finalen Wortes führen. Des Weiteren liegen keine Bewertungen zu der generellen Bekanntheit der verwendeten Ausdrücke bei beiden Lerner:innengruppen vor. Dies könnte in weiteren Studien ebenfalls berücksichtigt werden.

In Bezug auf Frage (2) konnte in **Studie III** gezeigt werden, dass beide Lerner:innengruppen häufiger relatierte als unrelatierte Vervollständigungen des Idiomfragmentes fixierten. Dies indiziert, dass sie Einzelwort-Bedeutungen innerhalb von Idiomen aktivieren. Im Vergleich zu erwachsenen Muttersprachler:innen (siehe **Studie I**), trat diese Aktivierung der Einzelwort-Bedeutung allerdings nicht zeitgleich mit der Antizipation der korrekten Vervollständigung, sondern erst später auf. Bei erwachsenen Muttersprachler:innen deuten die Daten darauf hin, dass sie die Einzelwort-Bedeutung automatisch ko-aktivieren. Dies scheint nicht der Fall für Lernenden der ersten und zweiten Sprache zu sein.

Jedoch scheint es qualitative Unterschiede zwischen den beiden Lerner:innengruppen für die Aktivierung der Einzelwort-Bedeutung zu geben, die sich in der Richtung der Unterschiede zwischen den Fixationen auf relatierte und unrelatierte Wörter zeigt. Dreizehnjährige Kinder in **Studie III**, wie auch Erwachsene (**Studie I**) und zehnjährige Kinder (**Studie II**), zeigten eine weniger starke Reduktion der Fixationen auf semantisch relatierte Wörter im Vergleich zu unrelatierten Wörtern. Im Gegensatz dazu, zeigten Zweitsprachler:innen einen Anstieg an Fixationen auf semantisch relatierte Wörter relativ zu unrelatierten Wörter. Diese Ergebnisse könnten im Rahmen der Annahme von Arnon und Christiansen (2017) interpretiert werden. Demzufolge erwerben Kinder MWE als untersegmentierte Einheiten und zerlegen diese erst später in einzelne Bestandteile. Zweitsprachler:innen hingegen lernen MWE als Kombinationen aus einzelnen Wörtern, weshalb die Repräsentation von MWE stärker auf den einzelnen Bestandteilen aufbaut (Abbildung 28).

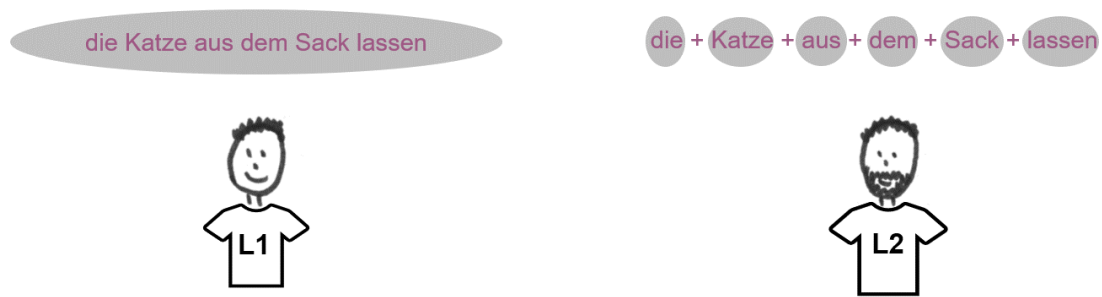


Abbildung 28: Schematische Darstellung: Dreizehnjährige Kinder repräsentieren MWE als eine Mehrwort-Repräsentation; Zweitsprachler:innen lernen MWE als Kombination aus den einzelnen Wörtern.

Alternativ kann das unterschiedliche Aktivierungsmuster zwischen den Lerner:innengruppen auch auf andere Faktoren wie sprachkompetenz-abhängige Unsicherheit mit den Ausdrücken oder geringe Fähigkeit der Inhibition dieser Aktivierung zurückzuführen sein. Um diese Alternativ-Erklärungen zu untersuchen, könnten weitere Studien die Aktivierung der Einzelwort-Bedeutung in Idiomen für Sprachlernende mit unterschiedlicher Sprachkompetenz messen und auch Inhibitionsfähigkeiten direkt dazu in Bezug setzen.

Insgesamt lässt sich schlussfolgern, dass (1) sowohl Kinder im Erstspracherwerb als auch Erwachsene im Zeitspracherwerb Erwartungen über die Idiom-Konfiguration aufbauen, die sich in antizipatorischen Fixationen auf das korrekte idiom-finale Wort widerspiegelt. (2) aktivieren beide Lerner:innengruppen die Einzelwort-Bedeutung des erwarteten Wortes. Es ist möglich, dass unterschiedliche Lernprozesse der Aktivierung der Einzelwort-Bedeutung zugrunde zu liegen. Generell benötigt es aber noch weitere Forschung, um unterschiedliche Lern- und Verarbeitungsmechanismen von MWE im Erst- und Zweitspracherwerb vergleichen zu können.

8 Interpretation und Zusammenfassung

alles unter einen Hut bringen

Ziel der vorliegenden Dissertation war es, die mentale Repräsentation von Idiomen (z.B. *die Katze aus dem Sack lassen*) als Exemplar für Mehrworteinheiten (MWE) bei erwachsenen Muttersprachler:innen (**Studie I**) und sprachlernenden Kindern im Altern von zehn (**Studie II**) und dreizehn Jahren (**Studie III**) zu untersuchen. MWE sind konventionalisierte Ausdrücke, die aus mehreren Wörtern bestehen. Aufgrund der Konventionalisierung ihrer Wort-Konfiguration und Phrasen-Bedeutung besteht die Frage, ob MWE im mentalen Lexikon als eine Einheit repräsentiert werden und wie diese erworben werden.

Die mentale Repräsentation von Idiomen kann indirekt anhand von Verarbeitungsmechanismen während der Sprachverarbeitung untersucht werden (Kapitel 3). Erwachsene Muttersprachler:innen und Kinder verarbeiten bekannte und häufig auftretende MWE schneller als neuartige Ausdrücke (z.B. Arnon & Snider, 2010; Bannard & Matthews, 2008; Carrol & Conklin, 2020; Conklin & Schmitt, 2008; Tremblay & Baayen, 2010; Tremblay et al., 2011). Außerdem können erwachsene Muttersprachler:innen Wörter innerhalb einer konventionalisierten Wort-Konfiguration vorhersagen (z.B. Carrol & Conklin, 2020; Rommers, Dijkstra & Bastiaansen, 2013; Underwood et al., 2004). Diese Befunde deuten darauf hin, dass MWE einen besonderen Status im mentalen Lexikon aufweisen (vgl. Sprenger et al., 2006; Swinney & Cutler, 1979; Van Lancker Sidtis, 2012). Bisher fehlten empirische Untersuchungen, ob Kinder ebenfalls einzelne Wörter innerhalb von Idiomen vorhersagen können. Außerdem könnten Unterschiede in der Vorhersage idiom-finaler Wörter zwischen unterschiedlichen Altersgruppen auf eine Veränderung der mentalen Repräsentation hinweisen.

Um weiterhin die Natur der mentalen Repräsentation zu ergründen, spielt die Aktivierung der Einzelwort-Bedeutungen innerhalb von Idiomen eine Rolle. Da die Einzelwort-Bedeutungen nicht notwendigerweise für das Verstehen der Phrasen-Bedeutung verarbeitet werden müssen, würde eine Aktivierung solcher Hinweise darauf geben, dass einzelne Bestandteile innerhalb von MWE ebenfalls verarbeitet werden. Während es bisher für erwachsene Muttersprachler:innen Evidenz für eine semantische Aktivierung der Einzelwörter gibt (Diskussion siehe Kapitel 4.3), lagen noch keine Befunde für Kinder bezüglich dieser Fragestellung vor. Falls MWE als eine Einheit erworben werden (vgl. Arnon & Christiansen, 2017; Sprenger et al., 2019; Wray, 2005) könnten die einzelnen Wörter innerhalb von MWE für Kinder weniger verfügbar sein als für Erwachsene.

Anhand der vorliegenden Dissertation sollte geschlussfolgert werden, in welcher Größe sprachliche Einheiten im mentalen Lexikon abgespeichert sind. Des Weiteren sollte erforscht werden, inwiefern sich die mentale Repräsentation von Idiomen mit zunehmendem Alter verändert. Mit zunehmendem Alter steigt die Expositionshäufigkeit und die Bekanntheit von Idiomen (Bonin et al., 2013; Sprenger et al., 2019; Tabossi et al., 2011). Da die Stärke der Repräsentation von Idiomen in starken Zusammenhang mit derer Auftretenshäufigkeit und der Bekanntheit steht (z.B. Arnon & Snider, 2010; Bonin et al., 2013; Carrol & Littlemore, 2020), erhält dieser Punkt besondere Relevanz. Unterschiede zwischen den Altersgruppen der Muttersprachler:innen in den **Studien I, II** und **III** begründen sich deshalb vermutlich auf erfahrungsbedingte Unterschiede mit diesen Ausdrücken.

Die vorliegende Dissertation im Spezifischen untersuchte, ob zehn- und dreizehnjährige Kinder und junge Erwachsene idiom-finale Wörter vorhersagen (z.B. *Sack* in dem Idiom *jemand ließ die Katze aus dem Sack*) und auch deren semantische Merkmale ko-aktivieren (die Bedeutung des Wortes *Sack*). Zu diesem Zweck wurden semantische Erwartungsparadigmen in Eye-Tracking- und EKP-Experimenten (vgl. Kapitel 2) durchgeführt. Anhand von deskriptiven Vergleichen der Daten aus den **Studien I, II** und **III** ließe sich potentiell eine Veränderung der mentalen Repräsentation von Idiomen während des Spracherwerbs erkennen. Dies wäre in unterschiedlichen Verarbeitungsmechanismen zwischen den drei untersuchten Altersgruppen detektierbar.

Zunächst unterschieden sich die drei Altersgruppen hinsichtlich ihres Wissens um spezifische Idiome (Tabelle 13). In den **Studien I, II** und **III** erkannten Versuchspersonen mit zunehmendem Alter den idiomatisch Ausdruck häufiger korrekt (Erwachsene: 93-99%, dreizehnjährige Kinder: 91%, zehnjährigen Kinder: 70-80% korrekte Antworten). Diese Befunde gehen einher mit den Befunden von Sprenger und Kolleg:innen (2019), die zeigen, dass mit zunehmendem Alter mehr idiomatische Ausdrücke bekannt sind.

Tabelle 13: Durchschnittliche Prozentzahl der idiomatisch korrekten Antworten für die drei Altersgruppen in den einzelnen Experimenten.

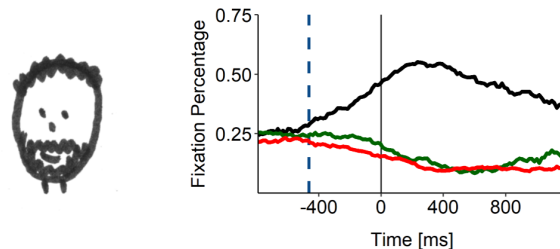
	Studie II Zehnjährige Kinder	Studie III Dreizehnjährige Kinder	Studie I Erwachsene Muttersprachler:innen
Visual-World-Experiment	79.76 % (9.56)	91.15 % (11.00)	98.87 % (1.85)
EKP-Experiment (auditiv)	70.10 % (10.22)		94.31 % (5.12)
EKP-Experiment (visuell)			93.80 % (4.70)

In Bezug auf die Vorhersage des idiom-finalen Wortes bauten alle Altersgruppen eine starke Erwartung dieser Wörter auf. Dies war sowohl in Eye-Tracking- als auch EKP-Daten erkennbar. Diese Ergebnisse bestätigen Befunde über Verarbeitungsvorteile von MWE im Vergleich zu Nicht-MWE bei Kindern und Erwachsenen (z.B. Bannard & Matthews, 2008; Carrol & Littlemore, 2020; Conklin & Schmitt, 2008; Tabossi et al., 2009). Zusammen bieten diese Befunde Hinweise auf eine Mehrwort-Repräsentation von Idiomem im mentalen Lexikon (z.B. Arnon & Cohen Priva, 2014; Reuterskiöld & Van Lancker Sidtis, 2013; Sprenger et al., 2006; Swinney & Cutler, 1979). Dementsprechend sind einzelne Wörter nicht die einzige Verarbeitungseinheit im mentalen Lexikon.

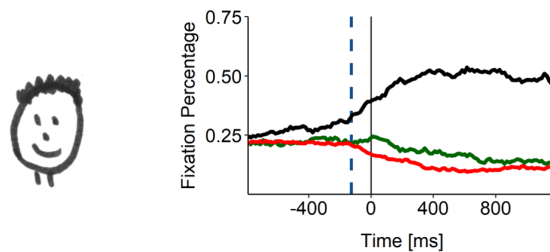
Dennoch geben die Eye-Tracking-Daten Anlass zu der Annahme, dass mit zunehmendem Alter das idiom-finale Wort zeitlich früher während der Verarbeitung vorhergesagt werden kann. In Abbildung 29 sind die Eye-Tracking-Daten für die drei Altersgruppen dargestellt. Der Beginn der Vorhersage ist durch eine vertikale, blau gestrichelte Linie gekennzeichnet. Numerisch verschiebt sich der Beginn der Vorhersage mit zunehmendem Alter zu einem früheren Zeitpunkt (Erwachsene Muttersprachler:innen: 464 ms, dreizehnjährige Kinder: 128 ms, zehnjährige Kinder: 112 ms vor Ende des Idiomfragmentes). Dies könnte ein genereller Effekt sein, der unabhängig von der Erfahrung mit MWE im Speziellen ist. Typischerweise stehen antizipatorische Fixationen im Visual-World-Paradigma in einem positiven Zusammenhang mit allgemeinen linguistischen Fähigkeiten wie der Größe des Vokabulars oder Lese- und Schreibfähigkeiten (z.B. Borovsky et al., 2012; Huettig, 2015; Huettig & Pickering, 2019; Mani & Huettig, 2012, 2014), die sich mit zunehmendem Alter ebenfalls weiterentwickeln können. Weiterhin könnte dieser Effekt aber auch durch Erfahrungsunterschiede mit MWE erklärt werden. Mit altersbedingt zunehmender Expositionshäufigkeit der jeweiligen Ausdrücke, können Versuchspersonen diese schneller erkennen. Weil erwachsene Muttersprachler:innen aufgrund ihrer größeren Erfahrung die Ausdrücke schneller erkennen als Kinder, aktivieren sie auch die Idiom-Bestandteile schneller. Dies spricht dafür, dass sich die mentale Repräsentation ebenfalls mit zunehmendem Alter und Expositionshäufigkeit verstärkt. Da wir in den **Studien I, II** und **III** nicht für allgemeine Sprachfähigkeiten

kontrolliert haben, lässt sich an dieser Stelle jedoch keine eindeutige Schlussfolgerung ziehen. Zukünftige Studien sollten demnach allgemeine linguistische Fähigkeiten, die in Zusammenhang mit Vorhersagefähigkeiten stehen, beachten.

(a) Studie I: Erwachsene Muttersprachler:innen (Beginn der Vorhersage: -464 ms).



(b) Studie III: Dreizehnjährige Kinder (Beginn der Vorhersage: -128 ms).



(c) Studie II: Zehnjährige Kinder (Beginn der Vorhersage: -112 ms).

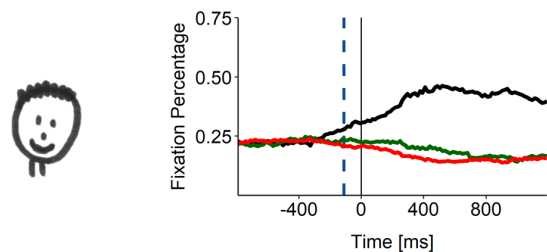


Abbildung 29: Vergleich der Fixationen für die drei Altersgruppen im Eye-Tracking-Experiment. Schwarze Linie = idiomatisch korrektes Wort; grüne Linie = semantisch relatives Wort; rote Linie = unrelatives Wort; vertikale, blau gestrichelte Linie = Beginn der Vorhersage.

Die Versuchspersonen aller Altersgruppen aktivierten die Einzelwort-Bedeutung innerhalb von Idioms. Dennoch unterscheiden sich die Altersgruppen hinsichtlich des Zeitpunktes der Bedeutungs-Aktivierung (Abbildung 29 für Vergleiche der Eye-Tracking-Daten und Tabelle 14 für Vergleiche der EKP-Daten). Sowohl Eye-Tracking- als auch EKP-Daten deuten darauf hin, dass erwachsene Muttersprachler:innen in **Studie I** das Einzelwort zeitgleich mit dessen Bedeutung aktivierten. Diese Aktivierung tritt aber nur kurzzeitig auf. Vermutlich sind diese Effekte für erwachsene Muttersprachler:innen deshalb nur für gesprochene Idiome messbar (Diskussion siehe Kapitel 4.3). Für die Verarbeitung geschriebener Idiome, die Wort für Wort präsentiert wurden, ist der zeitliche Abstand zwischen dem Erkennen des Idioms und der

Messung der semantischen Aktivierung zu groß, um diese kurzzeitige Aktivierung messen zu können. Es kann angenommen werden, dass erwachsene Muttersprachler:innen die Einzelwort-Bedeutung automatisch aktivieren. Im Gegensatz dazu zeigten Kinder in den **Studien II** und **III** keine Hinweise auf eine automatische, sondern eine zeitlich verzögerte Aktivierung der Einzelwort-Bedeutung innerhalb der Eye-Tracking- und EKP-Experimente. Die Eye-Tracking-Daten über die drei unterschiedlichen Altersgruppen hinweg legen nahe, dass mit zunehmendem Alter sich auch der zeitliche Abstand zwischen der Vorhersage des idiom-finalen Wortes und der Aktivierung dessen Bedeutung verkürzt. Ein genereller Alterseffekt in Bezug auf diese Tendenz ist unwahrscheinlich. Die EKP-Ergebnisse des Kontrollexperimentes in **Studie II** deuten darauf hin, dass Kinder im wörtlichen, nicht formelhaften Kontext fähig sind, schnelle semantische Erwartungen aufzubauen. Die frühe Einzelwortaktivierung bei Erwachsenen könnte daraus resultieren, dass mit steigendem Alter auch die Aktivierung der Einzelwort-Bedeutung automatischer erfolgt. Im Vergleich zu Kindern haben erwachsene Versuchspersonen in der Regel mehr Erfahrung mit den einzelnen Wörtern innerhalb der Ausdrücke, insbesondere auch in anderen Kontexten, gesammelt (vgl. Arnon & Clark, 2011), weshalb die Aktivierung dieser auch innerhalb von festen MWE im Erwachsenenalter automatisch erfolgt. Die späte Aktivierung der Einzelwort-Bedeutung bei Kindern könnte daher rühren, dass sie noch keine starke Mehrwort-Repräsentation aufgebaut haben. Der Argumentation folgend, wäre die zeitliche Verschiebung der Einzelwortaktivierung darauf zurückzuführen, dass sich die mentale Repräsentation von Idiomen mit zunehmendem Alter verstärkt.

Tabelle 14: Vergleich der EKP-Effekte für die Verarbeitung relatierter und unrelatierter Vervollständigungen von Idiomfragmenten in den **Studien I** und **II**.

				prä-N400		N400		PNP	
		Altersgruppe	Modalität	Zeitfenster	Effekt	Zeitfenster	Effekt	Zeitfenster	Effekt
Studie I	Experiment 2	Erwachsene Muttersprachler:innen	auditiv	100-200 ms	unrelatiert > relatiert unrelatiert = relatiert	300-500 ms	unrelatiert > relatiert	700-1000 ms	unrelatiert = relatiert
Studie I	Experiment 3	Erwachsene Muttersprachler:innen	visuell			300-400 ms	unrelatiert = relatiert		
Studie II	Experiment 3	Zehnjährige Kinder	auditiv	100-300 ms	unrelatiert = relatiert	300-500 ms	unrelatiert = relatiert	600-1000 ms	unrelatiert > relatiert

Unter der Annahme, dass mit zunehmendem Alter die Erfahrung mit den jeweiligen Ausdrücken steigt (Sprenger et al., 2019) und dementsprechend ebenfalls der Verarbeitungsvorteil für MWE (Arnon & Cohen Priva, 2014; Bannard & Matthews, 2008; Siyanova-Chanturia, Conklin & van Heuven, 2011), könnten die vorliegenden Unterschiede zwischen den Altersgruppen ebenfalls auf Erfahrungs- bzw. Expositionsunterschieden basieren. Mit altersbedingter Zunahme der Expositionshäufigkeit verstärkt sich demnach auch die Repräsentation der Idiome

im mentalen Lexikon (Abbildung 30). In den **Studien I, II** und **III** wäre die Stärkung dieser Repräsentation in einer zeitlich früheren Vorhersage des idiom-finalen Wortes und sinkender Toleranz gegenüber semantisch relatierten, modifizierten Varianten reflektiert. Dies unterstützt Modelle, die die Häufigkeit der Exposition als treibenden Faktor im Idiom-Erwerb sehen (Arnon & Christiansen, 2017; Grimm et al., 2017; Nippold & Rudzinski, 1993).

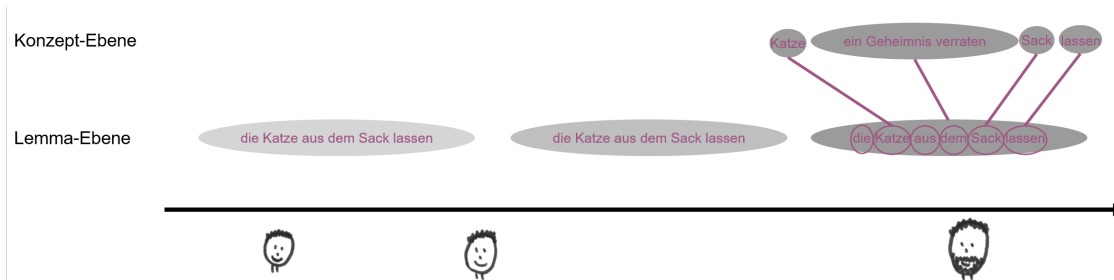


Abbildung 30: Schematische Darstellung des Modells, wie sich die mentale Repräsentation von MWE mit steigendem Alter verstärkt. Dunklere Farbe der Mehrwort-Repräsentation stellt eine stärkere Repräsentation dar. Anhand der vorliegenden Daten können keine Rückschlüsse auf die Verknüpfung von Phrasen-Bedeutung und Wort-Konfiguration gezogen werden.

Gleichzeitig könnte mit zunehmendem Alter auch eine Automatisierung der Aktivierung der Einzelwort-Bedeutung erfolgen. Dies könnte erklären, weshalb im Erwachsenenalter auch einzelne Bestandteile von MWE verfügbar sind (**Studie I**). Darauf aufbauend unterstützen die vorliegenden Ergebnisse ebenfalls hybride Modelle (Sprenger et al., 2006), die annehmen, das MWE sowohl als eine Mehrwort-Repräsentation als auch als einzelne Wörter im mentalen Lexikon abgespeichert sind.

Die hier gemachten Annahmen bezüglich der Altersunterschiede basieren lediglich auf deskriptiven Vergleichen zwischen den unterschiedlichen Altersgruppen. In dieser Hinsicht fehlen direkte statistische Vergleiche zwischen den Altersgruppen. Evidenz über einen positiven Zusammenhang zwischen Alter und Verarbeitungsvorteil bzw. Vorhersagefähigkeiten würde aufzeigen, dass mit steigender, altersbedingter Erfahrung mit MWE deren Repräsentationen im mentalen Lexikon gestärkt werden. Um Effekte allgemeiner Vorhersagefähigkeiten auszuschließen, sollte man beispielsweise individuelle Lesefähigkeiten und die Größe des Vokabulars kontrollieren. Ähnliche direkte Vergleiche zwischen den Altersgruppen könnte für die Aktivierung der Einzelwort-Bedeutungen durchgeführt werden.

Interessant wäre weiterhin, ob sich eine Stärkung der mentalen Repräsentation von MWE auch auf Ebene individueller Ausdrücke und Personen finden ließe. Bisher wurde der Verarbeitungsvorteil von MWE gegenüber Nicht-MWE in Abhängigkeit der generellen Auftretenshäufigkeit in Korpusdaten oder in Abhängigkeit von der generellen Bekanntheit mit den Ausdrücken untersucht (z.B. Arnon & Clark, 2011; Arnon & Cohen Priva, 2014; Arnon &

Snider, 2010; Bannard & Matthews, 2008; Bonin et al., 2013; Carrol & Conklin, 2020; Qualls et al., 2003). Insgesamt finden diese genannten Studien positive Zusammenhänge zwischen der Stärke der Repräsentation und der Auftretenshäufigkeit, der Bekanntheit bzw. dem Alter. In den **Studien I, II** und **III** verwendeten wir Idiome, die von erwachsenen Muttersprachler:innen generell als bekannt bewertet wurden. Für Kinder untersuchten wir lediglich, wie sie Idiome verarbeiteten, die sie wiedererkannt hatten. Dies ist jedoch nur ein dichotomes Maß für die Bekanntheit. Da das Wissen um MWE wie Idiome jedoch stark zwischen Personen variieren kann (Sprengrer et al., 2019), wäre wichtig zu untersuchen, inwiefern der Grad der individuellen Erfahrungen mit diesen Ausdrücken deren mentale Repräsentation verändern könnte.

Anschließend an die vorliegenden **Studien I, II** und **III** stellt sich die Frage, was zugrundeliegende Lernmechanismen für den Erwerb von MWE und Idiome sind. Unter der Annahme, dass individuelle, idiomatische Ausdrücke relativ selten im Sprachgebrauch auftreten (Moon, 1988), stellt sich die Frage, weshalb Kinder, die wahrscheinlich noch wenig Erfahrung mit diesen Ausdrücken haben, Idiome trotzdem als Mehrwort-Repräsentation im mentalen Lexikon abspeichern. Weiterhin scheint es unterschiedliche Lernmechanismen und daraus resultierende mentale Repräsentationen im Vergleich vom Erst- und Zeitspracherwerb zu geben (vgl. Arnon & Christiansen, 2017). Falls dies zutrifft, könnten die Ergebnisse von **Studie III** dafür sprechen, dass Kinder im Erstspracherwerb Idiome als eine Einheit erlernen. Daran anschließend wäre interessant, weshalb Kinder während des initialen Erwerbs Idiome nicht in einzelne Bestandteile zerlegen, sondern diese als Einheit lernen, obwohl sie nicht sehr häufig auftreten.

Die Lernstudie von Reuterskiöld und Van Lancker Sidtis (2013) legt nahe, dass besondere Merkmale von MWE zu einem Erwerb als holistische Einheit der Ausdrücke führen können. In dieser Studie konnten Kinder bereits nach einmaliger Exposition idiomatische Ausdrücke besser erinnern als nicht-idiomatische Ausdrücke. Die Autor:innen nehmen an, dass Idiome besondere Merkmale aufweisen, weshalb sie als eine Einheit gelernt werden.

Beck und Weber (eingereicht) führten ebenfalls eine Lernstudie durch und konnten zeigen, dass die Einzelwörterhäufigkeit mit dem Lernerfolg in Zusammenhang steht. In dieser Studie lernten erwachsene Muttersprachler:innen „neue“ Idiome besser, wenn diese Wörter beinhalten, die selten in Korpusdaten auftreten, als wenn diese häufige Wörter enthalten. Die Autor:innen argumentieren, dass Phrasen, die seltene Wörter beinhalten, salienter sind als Phrasen, die häufige Wörter beinhalten, und deshalb leichter zu merken sind.

Da Kinder im alltäglichen Spracherwerb hauptsächlich gesprochenen Input erhalten, könnten zusätzlich prosodische Merkmale eine Rolle für die Größe der erworbenen Einheiten spielen. Anhand ihrer prosodischen Merkmale (Sivanova-Chanturia & Lin, 2018) könnten MWE als zusammenhängende Einheit erkannt und erlernt werden (Überblick siehe Lin, 2018).

Um die Frage nach den Erwerbsmechanismen weiter zu untersuchen, könnten Lernstudien vergleichbar zu denen von Reuterskiöld und Van Lancker Sidtis (2013) und Beck und Weber (eingereicht) unter Berücksichtigung der besonderen Merkmale von MWE und Idiomen (Beck & Weber, eingereicht; Lin, 2018; Reuterskiöld & Van Lancker Sidtis, 2013) durchgeführt werden. Es besteht die Möglichkeit, dass MWE aufgrund von besonderen Merkmalen als eine Einheit erkannt und erlernt werden, und deshalb nach wenig Exposition bereits bekannte Ausdrücke sind.

In Bezug auf das mentale Lexikon kann man schlussfolgern, dass einzelne Wörter nicht die alleinige grundlegende Einheit für die Speicherung sprachlichen Wissens sind. Mehrworteinheiten weisen im Erwachsenen- und Kindesalter einen besonderen Stellenwert auf. Deshalb kann man ebenfalls davon ausgehen, dass diese sprachlichen Einheiten als Mehrwort-Repräsentationen im Gedächtnis abgespeichert sind. Dennoch sprechen empirische Befunde dafür, dass diese keine strikt holistischen Repräsentationen sind, sondern Einzelwort-Bedeutungen verfügbar sind. Altersunterschiede in der Verarbeitung von Idiomen deuten darauf hin, dass sich die Repräsentation mit zunehmender Erfahrung verstärkt.

Anhang

Land in Sicht

In den **Studien I, II** und **III** verwendetes Stimulusmaterial. Relevante Bewertungen von erwachsenen Muttersprachler:innen sind angegeben.

Item	Visual- Word- Eye- Tracking- Experiment	EKP- Experiment	Idiomfragment	Idiom- inales Word	Relatives Word	Bekanntheit (N=20), Skala von 1 (nicht bekannt) bis 7 (sehr bekannt)		Relation des Idiom-inales Wortes zu dem des Idioms (N=25), Skala von 1 (nicht relativiert) bis 7 (stark relativiert)		Class Probability (N = 17)	Wörtlichkeit (N=20), Skala von 1 (keine wörtliche Interpretation möglich) bis 7 (wörtliche Interpretation möglich)	
						Mittelwert	Standardfehler	Mittelwert	Standardfehler		Prozent	Mittelwert
1	x	x	Julia rutschte das Herz in die	Hose	Jacke	6,05	0,31	3	0,34	100%	1,3	0,11
2	x	x	Lena setzte alle Hebel in	Bewegung	Sprünge	6,35	0,21	4,92	0,33	88%	6,2	0,30
3	x	x	Marie stand Gabriel Rede und	Antwort	Frage	5,65	0,36	5,72	0,32	100%	5,45	0,47
4	x	x	Hannah schlug sich die Zeit um die	Ohren	Augen	5,35	0,44	2,17	0,37	100%	1,55	0,18
5	x	x	Sarah band sich einen Klotz aus	Bein	Knie	5,25	0,32	3,4	0,34	100%	6,55	0,18
6	x	x	Paula malte den Teufel an die	Wand	Tür	6,2	0,26	2,44	0,34	100%	5,9	0,38
7	x	x	Sofia brachte die Aufgaben unter Dach und	Fach	Schrank	5,6	0,37	3,08	0,36	100%	4,65	0,45
8	x	x	Annika war Balsam für die	Seele	Gefühle	5,95	0,29	4,96	0,35	94%	1,65	0,25
9	x	x	Amelie hatte einen Frosch im	Hals	Rücken	6	0,29	5,88	0,29	100%	2,8	0,43
10	x	x	Emma packte den Stier bei den	Hörnern	Zähnen	4,35	0,41	3,58	0,37	88%	5,6	0,46
11	x	x	Laura streute Salz in die	Wunde	Narbe	6,45	0,18	5,36	0,33	94%	6,6	0,18
12	x	x	Johanna hatte Tomaten auf den	Augen	Ohren	5,1	0,35	4,92	0,34	88%	6,25	0,30
13	x	x	Isabell hatte Schmetterlinge im	Bauch	Arm	6,5	0,24	4,96	0,34	100%	1,7	0,21
14	x	x	Jasmin lebte wie die Made im	Speck	Käse	4,4	0,41	3,12	0,36	94%	5,15	0,52
15	x	x	Lisa packte das Übel an der	Wurzel	Blüte	4	0,36	4,12	0,34	65%	3,7	0,52
16	x	x	Melina ließ die Kirche im	Dorf	Feld	5,15	0,32	2,36	0,29	94%	5,3	0,51
17	x	x	Eva war das schwächste Glied der	Kette	Linie	5,75	0,28	3,8	0,36	76%	5,05	0,48
18	x	x	Helena fiel ein Stein vom	Herzen	Magen	6,8	0,12	4,96	0,31	100%	1,3	0,13
19	x	x	Tabes hatte mehr Glück als	Verstand	Geist	6,15	0,28	5,76	0,31	94%	5,45	0,48
20	x	x	Nora hatte ihr Herz am rechten	Fleck	Platz	6,35	0,20	2,92	0,35	100%	4,4	0,56
21	x	x	Lukas war am Ende seines	Latein	Spanisch	5,25	0,37	3,28	0,43	71%	4,45	0,49
22	x	x	Leon nahm eine Mütze voll	Schlaf	Traum	4,15	0,35	6,16	0,32	65%	1,25	0,10
23	x	x	Max schlug zwei Fliegen mit einer	Klappe	Kiste	6,35	0,17	3,4	0,37	88%	6,85	0,11
24	x	x	Nico erblickte das Licht der	Welt	Venus	5,75	0,32	5,68	0,32	88%	6,1	0,30
25	x	x	Simon hatte bei Nina einen Stein im	Brett	Nagel	4,2	0,47	2	0,22	94%	3,15	0,41
26	x	x	Robin legte für Natalie die Hand ins	Feuer	Holz	6,45	0,15	3	0,32	100%	5,65	0,46
27	x	x	Linus verlor den Boden unter den	Füßen	Händen	6,35	0,21	3,84	0,38	100%	4,7	0,50
28	x	x	Laars fand in jeder Suppe ein	Haar	Kinn	4,85	0,40	2,96	0,34	100%	6,35	0,32
29	x	x	Jannis fiel die Decke auf den	Kopf	Bart	6,45	0,15	3,6	0,37	100%	5,1	0,52
30	x	x	Daniel brachte das Fass zum	Überlaufen	Auströcknen	6,55	0,14	4,88	0,41	100%	6,4	0,28
31	x	x	Florian brachte Marius an den Rand der	Verzweiflung	Angst	6,4	0,22	6,12	0,23	82%	2,3	0,42
32	x	x	Julian hielt den Kopf über	Wasser	Regen	5,1	0,27	3,42	0,37	100%	6,6	0,15
33	x	x	Erik fiel mit der Tür ins	Haus	Zelt	5,95	0,32	2,92	0,32	100%	6,05	0,34
34	x	x	Moritz begab sich in die Höhle des	Löwen	Hasen	5,45	0,37	3,44	0,4	100%	5,95	0,36
35	x	x	Fabian stellte Emils Geduld auf die	Probe	Übung	6,1	0,23	5,16	0,33	100%	3,65	0,55
36	x	x	Timo brachte den Stein ins	Rollen	Kugelh	5,2	0,32	4,4	0,33	100%	6,35	0,27
37	x	x	Hannes ließ die Katze aus dem	Sack	Korb	5,75	0,31	2,76	0,31	94%	6	0,41
38	x	x	Antons Entscheidung stand auf Messers	Schneide	Klinge	5,1	0,43	3,8	0,39	100%	4	0,55
39	x	x	Dennis packte die Gelegenheit beim	Schopfe	Scheitel	4,9	0,35	2,92	0,34	100%	1,3	0,13
40	x	x	Emil war das fünfte Rad am	Wagen	Zug	6,5	0,15	2,92	0,41	94%	4,25	0,59

Abbildungsverzeichnis

1	Schematische Darstellung der Modelle zur Repräsentation von MWE.	3
2	Schematische Darstellung für die Bottom-Up Aktivierung des Wortes „Katze“ nach der Präsentation des Wortes „Hund“.	9
3	Schematische Darstellung für die Top-Down Aktivierung des Wortes „Felsen“ nach der Erwartung des Wortes „Sand“.	10
4	Schematische Darstellung des Eye-Tracking im Visual-Word-Paradigma anhand eines Beispiels aus Huettig und Altmann (2005). Links: Person hört über Kopfhörer einen Satz, während ihre Fixationen auf dem Bildschirm über eine Kamera aufgezeichnet werden. Mitte: Geschriebene Wörter auf dem Bildschirm (schwarz = kritisches Wort (<i>Piano</i>); grün = semantisch relatierter Kompetitor (<i>Trompete</i>); rot = unrelatierter Distraktor (<i>Möhre</i>)).	13
5	Schematische Darstellung der EEG und EKP. Links: Elektroden, die an der Kopfoberfläche über eine spezielle Mütze angebracht sind. Mitte: Roh-EEG-Daten mit markiertem Beginn eines Stimulus („Hund“) und markierter Epoche. Rechts: Mehrere EEG-Epochen gemittelt ergeben EKP, in denen unterschiedliche Komponenten erkennbar sind.	14
6	Schematische Darstellung der semantischen N400 in einem Erwartungs-Paradigma. Graduelle Abstufung der N400-Amplitude in Abhängigkeit von dem satz-finalen Wort (schwarz = erwartetes Wort; grün = semantisch relatives Wort; rot = unrelatives Wort). Schattierung des N400-Zeitfensters und der Lokalisierung über posteriore Bereiche.	16
7	Schematische Darstellung für die Bottom-Up Aktivierung des Wortes „Korb“ nach der Präsentation des Wortes „Sack“ innerhalb des Idioms „jemand ließ die Katze aus dem Sack“.	25
8	Schematische Darstellung für die Top-Down Aktivierung des Wortes „Korb“ nach der Erwartung des Wortes „Sack“ innerhalb des Idioms „jemand ließ die Katze aus dem Sack“.	27
9	Eye-Tracking-Experiment: Versuchsperson hört Idiomfragment als auditiven Wortstrom während sie vier Wörter auf dem Bildschirm sieht.	33

10	EKP-Experimente in den drei Bedingungen (schwarz = korrekt; grün = related; rot = unrelatiert): Versuchsperson (A) hört Varianten eines Idioms als auditiven Wortstrom bzw. (B) liest Varianten eines Idioms Wort-für-Wort auf dem Bildschirm geschrieben (zwischen den einzelnen Wörtern erscheint jeweils ein leerer Bildschirm).	33
11	Bauen erwachsene Muttersprachler:innen eine Erwartung für das idiom-finale Wort auf und aktivieren sie währenddessen dessen Einzelwort-Bedeutung?	34
12	Example with times indicating the duration of the respective displays.	42
13	Panel (A) Fixation percentage for correct completions (black), related distractors (green) and mean of unrelated distractors (red); black vertical line = offset of spoken stimuli (0 ms); blue vertical, dashed line = start of the anticipation (-464 ms); grey background = time window for GCA. Panel (B) Fixation percentage for semantically related and unrelated distractors (points = mean; error bars = standard error) with fit of the growth curve model (line).	43
14	Electrode configuration used in the experiment. Anterior-Left, Anterior-Right, and Posterior-Central ROIs are highlighted in light gray. Anterior-Central, Posterior-Left, and Posterior-Right ROIs are highlighted in dark grey.	51
15	Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right. (B) Difference topographies for the time windows 100–200 ms, 300–500 ms, and 700–1000 ms.	52
16	Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right. (B) Difference topographies for the time windows 300–400 ms and 500–700 ms.	59
17	Schematische Darstellung der kurzzeitigen Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomem und deren Messbarkeit in Abhängigkeit von dem Verarbeitungsmechanismus ((a) und (b); additive Effekte von Erwartung und Präsentation sind nicht dargestellt), jeweils: Oberer Teil : mentale Repräsentation von Idiomem in Anlehnung an das Superlemma-Modell von Sprenger et al. (2006). Idiome sind als Mehrwort-Repräsentationen gespeichert die Verbindungen zu den einzelnen Wörtern besitzen (durchgehende Pfeile = gemeinsame Elemente; gestrichelte Pfeile = Relationen). Mittlerer Teil : Zeitverlauf der Aktivierung der idiom-finalen Einzelwort-Bedeutung (dunkelgrüne Linie); Zeitpunkt der Aktivierung des idiom-finalen Wortes aufgrund (a) der Erwartung oder (b) der Präsentation (graue Pfeile); Zeitpunkt der Messung der Aktivierung der Einzelwort-Bedeutung (hellgrüne Pfeile). Unterer Teil : Schematische Darstellung unterschiedlich schneller Präsentationsarten (z.B. schnell: auditiver Wortstrom; langsam: Wort-für-Wort, visuelle Präsentation).	76

18	Bauen zehnjährige Kinder eine Erwartung für das idiom-finale Wort auf und aktivieren sie währenddessen dessen Einzelwort-Bedeutung?	82
19	Example with times indicating the duration of the respective displays.	90
20	Panel (A) Fixation percentage for correct completions (black), related distractors (green) and mean of unrelated distractors (red); black vertical line = offset of spoken stimuli (0 ms); blue vertical, dashed line = start of the anticipation (-112 ms); grey background = time window for GCA. Panel (B) Fixation percentage for semantically related and unrelated distractors (points = mean; error bars = standard error) with fit of the growth curve model (line).	91
21	Electrode configuration used in the experiment. Anterior-Left, Anterior-Right, and Posterior-Central ROIs are highlighted in light gray. Anterior-Central, Posterior-Left, and Posterior-Right ROIs are highlighted in dark grey.	96
22	Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right; time windows for statistical analysis in boxes. (B) Difference topographies for the time windows 100–300 ms, 300–500 ms, and 600–1000 ms	99
23	Grand-Averaged ERPs (A) ERP-waveforms for the ROIs Anterior-Left, Anterior-Central, Anterior-Right, Posterior-Left, Posterior-Central, and Posterior-Right; time windows for statistical analysis in boxes. (B) Difference topographies for the time windows 100–300 ms, 300–500 ms, and 600–1000 ms	106
24	Bauen dreizehnjährige Kinder (L1) und erwachsene Zweitsprachler:innen (L2) eine Erwartung für das idiom-finale Wort auf und aktivieren sie währenddessen dessen Einzelwort-Bedeutung?	121
25	Example for visual display and auditory stimuli used in the experiment.	130
26	(A) Fixation patterns for L1 Children: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line).	132
27	(A) Fixation patterns for L2 Adults: Fixation Percentage for Correct Completions (black), Related Distractors (green) and Mean of Unrelated Distractors (red); eye fixations are aligned to the offset of the acoustic stimulus (0 ms); start of anticipation (dashed, blue); time window for GCA (grey). (B) Fixation Percentage for Related and Unrelated Distractors (points = mean; error bars = standard error) with fit of growth curve model (line).	134
28	Schematische Darstellung: Dreizehnjährige Kinder repräsentieren MWE als eine Mehrwort-Repräsentation; Zweitsprachler:innen lernen MWE als Kombination aus den einzelnen Wörtern.	143

29	Vergleich der Fixationen für die drei Altersgruppen im Eye-Tracking-Experiment. Schwarze Linie = idiomatisch korrektes Wort; grüne Linie = semantisch relatives Wort; rote Linie = unrelatives Wort; vertikale, blau gestrichelte Linie = Beginn der Vorhersage.	147
30	Schematische Darstellung des Modells, wie sich die mentale Repräsentation von MWE mit steigendem Alter verstärkt. Dunklere Farbe der Mehrwort-Repräsentation stellt eine stärkere Repräsentation dar. Anhand der vorliegenden Daten können keine Rückschlüsse auf die Verknüpfung von Phrasen-Bedeutung und Wort-Konfiguration gezogen werden.	149

Tabellenverzeichnis

1	Zusammenfassung über Studien, die die Aktivierung der Einzelwort-Bedeutung innerhalb von Idiomen gemessen haben.	30
2	German Example Sentence for Types 1-4 with English Equivalent	42
3	Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).	44
4	RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions.	51
5	RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions.	57
6	German Example Sentence for Types 1-4 with English Equivalent	89
7	Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).	91
8	RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions	97
9	RM-ANOVAs. C–Condition, R–Region, H–Hemisphere. * for significant main effects and interactions	104
10	German example sentence for types 1-4 with english equivalent.	130
11	Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).	132
12	Parameter Estimates for the Model including Distractor Type (Related vs. Unrelated).	135
13	Durchschnittliche Prozentzahl der idiomatisch korrekten Antworten für die drei Altersgruppen in den einzelnen Experimenten.	146
14	Vergleich der EKP-Effekte für die Verarbeitung relatierter und unrelatierter Vollständigungen von Idiomfragmenten in den Studien I und II	148

Literaturverzeichnis

- Abbot-Smith, K. & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review*, 23 (3).
- Ackerman, B. (1982). On comprehending idioms: Do children get the picture? *Journal of Experimental Child Psychology*, 33 (3), 439-454.
- Aitchison, J. (2012). *Words in the Mind: An Introduction to the Mental Lexicon*. John Wiley & Sons.
- Altmann, G. T. M. & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73 (3), 247-264.
- Altmann, G. T. M. & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language*, 57 (4), 502-518.
- Arnon, I. & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in Cognitive Science*, 9 (3), 621-636.
- Arnon, I. & Clark, E. V. (2011). Why brush your teeth is better than teeth – children’s word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7 (2), 107-129.
- Arnon, I. & Cohen Priva, U. (2013). More than words: the effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, 56 (3), 349-371.
- Arnon, I. & Cohen Priva, U. (2014). Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, 9 (3), 377-400.
- Arnon, I., McCauley, S. M. & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 265-280.
- Arnon, I. & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, 62 (1), 67-82.

- Atchley, R. A., Rice, M. L., Betz, S. K., Kwasny, K. M., Sereno, J. A. & Jongman, A. (2006). A comparison of semantic and syntactic event related potentials generated by children and adults. *Brain and Language*, 99 (3), 236-246.
- Bannard, C. & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, 19 (3), 241-248.
- Beck, S. D. & Weber, A. (2016a). Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology*, 7, 1305.
- Beck, S. D. & Weber, A. (2016b). L2 idiom processing: Figurative attunement in highly idiomatic contexts. *Conference Proceedings Paper at CogSci*.
- Beck, S. D. & Weber, A. (eingereicht). Phrasal learning is a horse apiece: No memory advantages for idioms in L1 and L2 adult learners.
- Bentin, S., McCarthy, G. & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60 (4), 343-355.
- Bernicot, J., Laval, V. & Chaminaud, S. (2007). Nonliteral language forms in children: In what order are they acquired in pragmatics and metapragmatics? *Journal of Pragmatics*, 39 (12), 2115-2132.
- Bobrow, S. & Bell, S. (1973). On catching on to idiomatic expressions. *Memory & Cognition*, 1 (3), 343-346.
- Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review*, 23 (3), 291-320.
- Bonin, P., Meot, A. & Bugaiska, A. (2013). Norms and comprehension times for 305 french idiomatic expressions. *Behavior Research Methods*, 45 (4), 1259-1271.
- Borovsky, A., Elman, J. L. & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112 (4), 417-436.
- Boudewyn, M. A., Long, D. L. & Swaab, T. Y. (2015). Graded expectations: Predictive processing and the adjustment of expectations during spoken language comprehension. *Cognitive, Affective, & Behavioral Neuroscience*, 15 (3), 607-624.
- Brown, C. & Hagoort, P. (1993). The processing nature of the N400: Evidence from masked priming. *Journal of Cognitive Neuroscience*, 5 (1), 34-44.

- Bybee, J. & McClelland, J. L. (2005). Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review*, 22, 381-410.
- Cacciari, C. & Corradini, P. (2015). Literal analysis and idiom retrieval in ambiguous idioms processing: a reading-time study. *Journal of Cognitive Psychology*, 27 (7), 797-811.
- Cacciari, C. & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language*, 27 (6), 668-683.
- Caillies, S. & Declercq, C. (2011). Kill the song-steal the show: what does distinguish predicative metaphors from decomposable idioms? *Journal of Psycholinguistic Research*, 40 (3), 205-223.
- Caillies, S. & Le Sourn-Bissaoui, S. (2006). Idiom comprehension in french children: A cock-and-bull story. *European Journal of Developmental Psychology*, 3 (2), 189-206.
- Cain, K., Oakhill, J. & Lemmon, K. (2005). The relation between children's reading comprehension level and their comprehension of idioms. *Journal of Experimental Child Psychology*, 90 (1), 65-87.
- Cain, K., Towse, A. S. & Knight, R. S. (2009). The development of idiom comprehension: an investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, 102 (3), 280-298.
- Carrol, G. & Conklin, K. (2015). Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research*, 7, 1-11.
- Carrol, G. & Conklin, K. (2020). Is all formulaic language created equal? unpacking the processing advantage for different types of formulaic sequences. *Language and Speech*, 63 (1), 95-122.
- Carrol, G. & Littlemore, J. (2020). Resolving figurative expressions during reading: The role of familiarity, transparency, and context. *Discourse Processes*, 57 (7), 1-18.
- Collins, A. M. & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, 82 (6), 407-428.
- Conklin, K. & Carrol, G. (2020). Words go together like 'bread and butter': The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*.
- Conklin, K., Pellicer-Sánchez, A. & Carrol, G. (2018). *Eye-tracking: A Guide for Applied Linguistics Research*. Cambridge University Press.

- Conklin, K. & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29 (1), 72-89.
- Conklin, K. & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, 32, 45-61.
- Connolly, J. F., D'Arcy, R. C., Kujala, A. & Alho, K. (2001). Phonological aspects of word recognition as revealed by high-resolution spatio-temporal brain mapping. *NeuroReport*, 12 (2), 237-243.
- Connolly, J. F. & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience*, 6 (3), 256-266.
- Connolly, J. F., Phillips, N. A. & Forbes, K. A. (1995). The effects of phonological and semantic features of sentence-ending words on visual event-related brain potentials. *Electroencephalography and Clinical Neurophysiology*, 94 (4), 276-287.
- Cortés-Monter, D. R., Angulo-Chavira, A. Q. & Arias-Trejo, N. (2017). Differences between skilled and less-skilled young readers in the retrieval of semantic, phonological, and shape information. *Journal of Research in Reading*, 40, S170-S189.
- Crutchley, A. (2007). Comprehension of idiomatic verb + particle constructions in 6- to 11-year-old children. *First Language*, 27 (3), 203-226.
- Cruttenden, A. (1981). Item-learning and system-learning. *Journal of Psycholinguistic Research*, 10 (1), 79-88.
- DeLong, K. A., Chan, W. H. & Kutas, M. (2018). Similar time courses for word form and meaning preactivation during sentence comprehension. *Psychophysiology*, e13312.
- DeLong, K. A., Quante, L. & Kutas, M. (2014). Predictability, plausibility, and two late ERP positivities during written sentence comprehension. *Neuropsychologia*, 61, 150-162.
- DeLong, K. A., Urbach, T. P. & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8 (8), 1117-1121.
- Diaz, M. T. & Swaab, T. Y. (2007). Electrophysiological differentiation of phonological and semantic integration in word and sentence contexts. *Brain Research*, 1146, 85-100.
- Diependaele, K., Lemhöfer, K. & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: a lexical entrenchment account. *Quarterly Journal of Experimental Psychology*, 66 (5), 843-863.

- Duchowski, A. T. (2017). *Eye tracking methodology: Theory and practice*. Springer.
- Fanari, R., Cacciari, C. & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology, 22* (3), 321-334.
- Federmeier, K. D. & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language, 41* (4), 469-495.
- Federmeier, K. D., McLennan, D. B., De Ochoa, E. & Kutas, M. (2002). The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study. *Psychophysiology, 39* (2), 133-146.
- Findlay, H. & Carrol, G. (2018). Contributions of semantic richness to the processing of idioms. *The Mental Lexicon, 13* (3), 311-332.
- Friederici, A. D. & Hahne, A. (2001). Development patterns of brain activity reflecting semantic and syntactic processes. In J. Weissenborn & B. Houle (Hrsg.), *Approaches to Bootstrapping: Phonological, Lexical, Syntactic, and Neurophysiological Aspects of Early Language Acquisition* (S. 231–246). Amsterdam/Philadelphia: John Benjamin.
- Geeraert, K., Baayen, R. H. & Newman, J. (2017). Idiom variation: Experimental data and a blueprint of a computational model. *Topics in Cognitive Science, 9* (3), 653-669.
- Geeraert, K., Baayen, R. H. & Newman, J. (2018). “Spilling the bag” on idiomatic variation. In S. Markantonatou, C. Ramisch, A. Savary & V. Vincze (Hrsg.), *Multiword expressions at length and in depth: Extended papers from the MWE 2017 workshop* (Bd. 2, S. 1-33). Berlin: Language Science Press.
- Grimm, R., Cassani, G., Gillis, S. & Daelemans, W. (2017). Facilitatory effects of multi-word units in lexical processing and word learning: A computational investigation. *Frontiers in Psychology, 8*, 555.
- Hagoort, P., Baggio, G. & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Hrsg.), *The Cognitive Neurosciences, 4th ed.* (S. pp. 819-836). Cambridge, MA: MIT Press.
- Hagoort, P. & Brown, C. M. . (2000). ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia, 38* (11), 1518-1530.
- Heil, M., Rolke, B. & Pecchinenda, A. (2004). Automatic semantic activation is no myth: Semantic context effects on the N400 in the letter-search task in the absence of response time effects. *Psychological Science, 15* (12), 852-857.
- Holcomb, P. J., Coffey, S. A. & Neville, H. J. (1992). Visual and auditory sentence processing: A developmental analysis using event-related brain potentials. *Developmental Neuropsychology, 8* (2-3), 203-241.

- Holcomb, P. J. & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, 5 (4), 281-312.
- Holsinger, E. (2013). Representing idioms: syntactic and contextual effects on idiom processing. *Language and Speech*, 56 (3), 373-394.
- Huang, Y. T. & Snedeker, J. (2011). Cascading activation across levels of representation in children's lexical processing. *Journal of Child Language*, 38 (3), 644-661.
- Huettig, F. (2015). Four central questions about prediction in language processing. *Brain Research*, 1626, 118-135.
- Huettig, F. & Altmann, G. T. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, 96 (1), B23-B32.
- Huettig, F. & Pickering, M. (2019). Literacy advantages beyond reading: Prediction of spoken language. *Trends in Cognitive Sciences*, 23 (6), 464-475.
- Huettig, F., Rommers, J. & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: a review and critical evaluation. *Acta Psychologica*, 137 (2), 151-171.
- Ito, A., Corley, M., Pickering, M. J., Martin, A. E. & Nieuwland, M. S. (2016). Predicting form and meaning: Evidence from brain potentials. *Journal of Memory and Language*, 86, 157-171.
- Ito, A. & Husband, E. M. (2017). How robust are effects of semantic and phonological prediction during language comprehension? A visual world eye-tracking study.
- Jiang, N. & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, 91 (3), 433-445.
- Joseph, H. S., Nation, K. & Liversedge, S. P. (2013). Using eye movements to investigate word frequency effects in children's sentence reading. *School Psychology Review*, 42 (2), 207-222.
- Juottonen, K., Revonsuo, A. & Lang, H. (1996). Dissimilar age influences on two ERP waveforms (LPC and N400) reflecting semantic context effect. *Cognitive Brain Research*, 4 (2), 99-107.
- Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, 20 (4), 647-670.
- Kamide, Y., Altmann, G. T. M. & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49 (1), 133-156.

- Kapatsinski, V. & Radicke, J. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. In R. L. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Hrsg.), *Formulaic Language* (Bd. 2, S. 499-520). John Benjamins.
- Kessler, R. & Beck, S. D. (eingereicht). L1 and L2 learners keep their eyes on the prize: Eye-tracking evidence during idiom recognition. *Proceedings of Linguistic Evidence 2020*.
- Kessler, R. & Friedrich, C. K. (eingereicht). Delayed prediction of idiom constituent meaning points to weak holistic multi-word representation in children. *Language, Cognition and Neuroscience*.
- Kessler, R., Weber, A. & Friedrich, C. K. (2020). Activation of literal word meanings in idioms: Evidence from eye-tracking and ERP experiments. *Language and Speech*.
- Konopka, A. E. & Bock, K. (2009). Lexical or syntactic control of sentence formulation? structural generalizations from idiom production. *Cognitive Psychology*, 58 (1), 68-101.
- Kuperberg, G. & Wlotko, E. (2020). A tale of two positivities and the N400: Distinct neural signatures are evoked by confirmed and violated predictions at different levels of representation. *Journal of Cognitive Neuroscience*, 32 (1), 12-35.
- Kutas, M. (1993). In the company of other words: Electrophysiological evidence for single-word and sentence context effects. *Language and Cognitive Processes*, 8 (4), 533-572.
- Kutas, M. & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, 4 (12), 463-470.
- Kutas, M. & Federmeier, K. D. (2011). Thirty years and counting: finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62 (1), 621-647.
- Kutas, M. & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207 (4427), 203-205.
- Kutas, M. & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307 (5947), 161-163.
- Kyriacou, M., Conklin, K. & Thompson, D. (2020). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech*, 63 (2), 404 -435.
- Langlotz, A. (2006). Occasional adnominal idiom modification-a cognitive linguistic approach. *International Journal of English Studies*, 6 (1), 85-108.
- la Roi, A., Sprenger, S. A. & Hendriks, P. (2019). Event-related potentials reveal increased dependency on linguistic context due to cognitive aging. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46 (7), 1226-1257.

- Lau, E. F., Phillips, C. & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the N400. *Nature Reviews Neuroscience*, 9 (12), 920-933.
- Laurent, J. P., Denhieres, G., Passerieux, C., Iakimova, G. & Hardy-Bayle, M. C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research*, 1068 (1), 151-160.
- Levorato, M. C. (1999). Idiom comprehension in children: Are the effects of semantic analysability and context separable? *European Journal of Cognitive Psychology*, 11 (1), 51-66.
- Levorato, M. C. & Cacciari, C. (1992). Children's comprehension and production of idioms: the role of context and familiarity. *Journal of Child Language*, 19 (2), 415-433.
- Levorato, M. C. & Cacciari, C. (1995). The effects of different tasks on the comprehension and production of idioms in children. *Journal of Experimental Child Psychology*, 60 (2), 261-283.
- Libben, M. R. & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition*, 36 (6), 1103-1121.
- Lin, P. M. S. (2018). Formulaic language and speech prosody. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Hrsg.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (S. 78-94). London: Routledge.
- Luck, S. J. (2014). *An Introduction to the Event-Related potential Technique*. MIT press.
- Mahler, N. A. & Chenery, H. J. (2018). A developmental perspective on processing semantic context: Preliminary evidence from sentential auditory word repetition in school-aged children. *Journal of Psycholinguistic Research*, 48 (1), 81-105.
- Mancuso, A., Elia, A., Laudanna, A. & Vietri, S. (2020). The role of syntactic variability and literal interpretation plausibility in idiom comprehension. *Journal of Psycholinguistic Research*, 49 (1), 99-124.
- Mani, N. & Huettig, F. (2012). Prediction during language processing is a piece of cake-but only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance*, 38 (4), 843-847.
- Mani, N. & Huettig, F. (2014). Word reading skill predicts anticipation of upcoming spoken language input: a study of children developing proficiency in reading. *Journal of Experimental Child Psychology*, 126, 264-279.
- McCallum, W. C., Farmer, S. F. & Pocock, P. V. (1984). The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 59 (6), 477-488.

- Meng, X., Tian, X., Jian, J. & Zhou, X. (2007). Orthographic and phonological processing in chinese dyslexic children: an ERP study on sentence reading. *Brain Research*, 1179, 119-130.
- Meyer, D. E. & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90 (2), 227.
- Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F. & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? a reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes*, 28 (6), 762-788.
- Molinaro, N. & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology*, 83 (3), 176-190.
- Moon, R. (1988). *Fixed Expressions and Idioms in English: A Corpus-Based Approach*. Oxford, UK: Clarendon Press.
- Newman, R. L. & Connolly, J. F. (2009). Electrophysiological markers of pre-lexical speech processing: evidence for bottom-up and top-down effects on spoken word processing. *Biological Psychology*, 80 (1), 114-121.
- Nieuwland, M. S. (2019). Do 'early' brain responses reveal word form prediction during language comprehension? a critical review. *Neuroscience and Biobehavioral Reviews*, 96, 367-400.
- Nippold, M. A. & Rudzinski, M. (1993). Familiarity and transparency in idiom explanation: A developmental study of children and adolescents. *Journal of Speech, Language, and Hearing Research*, 36 (4), 728-737.
- Nippold, M. A. & Taylor, C. (2002). Judgments of idiom familiarity and transparency: A comparison of children and adolescents. *Journal of Speech, Language, and Hearing Research*, 45, 384-391.
- Nippold, M. A. & Taylor, C. L. (1995). Idiom understanding in youth: Further examination of familiarity and transparency. *Journal of Speech, Language, and Hearing Research*, 38 (2), 426-433.
- Ortony, A., Turner, T. J. & Larson-Shapiro, N. (1985). Cultural and instructional influences on figurative language comprehension by inner city children. *Research in the Teaching of English*, 19 (1), 25-36.
- Peterson, R., Burgess, C., Dell, G. & Eberhard, K. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 27 (5), 1223-1237.

- Pinker, S. (1991). Rules of language. *Science*, 253 (5019), 530-535.
- Primativo, S., Reilly, J. & Crutch, S. J. (2017). Abstract conceptual feature ratings predict gaze within written word arrays: Evidence from a visual wor(1)d paradigm. *Cognitive Science*, 41 (3), 659-685.
- Qualls, C. D., Treaster, B., Blood, G. W. & Hammer, C. S. (2003). Lexicalization of idioms in urban fifth graders: a reaction time study. *Journal of Communication Disorders*, 36 (4), 245-261.
- Rabanus, S., Smolka, E., Streb, J. & Rösler, F. (2008). Die mentale Verarbeitung von Verben in idiomatischen Konstruktionen. *Zeitschrift für Germanistische Linguistik*, 36 (1), 27-47.
- Radeau, M. (1983). Semantic priming between spoken words in adults and children. *Canadian Journal of Psychology*, 37 (4), 547-556.
- Reuterskiöld, C. & Van Lancker Sidtis, D. (2013). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy*, 29 (2), 219-231.
- Rämä, P., Sirri, L. & Serres, J. (2013). Development of lexical-semantic language system: N400 priming effect for spoken words in 18- and 24-month old children. *Brain and Language*, 125 (1), 1-10.
- Rommers, J., Dijkstra, T. & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, 25 (5), 762-776.
- Rommers, J., Meyer, A. S., Praamstra, P. & Huettig, F. (2013). The contents of predictions in sentence comprehension: activation of the shape of objects before they are referred to. *Neuropsychologia*, 51 (3), 437-447.
- Schmitt, N. & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Hrsg.), *Formulaic Sequences: Acquisition, Processing, and Use* (S. 173-189). Amsterdam: John Benjamins.
- Shen, W., Qu, Q. & Li, X. (2016). Semantic information mediates visual attention during spoken word recognition in chinese: Evidence from the printed-word version of the visual-world paradigm. *Attention, Perception, & Psychophysics*, 78 (5), 1267-1284.
- Sivanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon*, 8 (2), 245-268.
- Sivanova-Chanturia, A. (2015). On the 'holistic' nature of formulaic language. *Corpus Linguistics and Linguistic Theory*, 11 (2), 285-301.

- Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E. & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language*, 175, 111-122.
- Siyanova-Chanturia, A., Conklin, K. & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research*, 27 (2), 251-272.
- Siyanova-Chanturia, A., Conklin, K. & van Heuven, W. J. (2011). Seeing a phrase "time and again" matters: the role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37 (3), 776-784.
- Siyanova-Chanturia, A. & Lin, P. M. S. (2018). Production of ambiguous idioms in English: A reading aloud study. *International Journal of Applied Linguistics*, 28 (1), 58-70.
- Siyanova-Chanturia, A. & Martinez, R. (2015). The idiom principle revisited. *Applied Linguistics*, 36 (5), 549-569.
- Smolka, E. & Eulitz, C. (2020). Can you reach for the planets or grasp at the stars?—modified noun, verb, or preposition constituents in idiom processing. In S. Schulte im Walde & E. Smolka (Hrsg.), *The Role of Constituents in Multiword Expressions: An Interdisciplinary, Cross-lingual Perspective* (S. 179-204). Berlin: Language Science Press.
- Smolka, E., Rabanus, S. & Rösler, F. (2007). Processing verbs in German idioms: Evidence against the configuration hypothesis. *Metaphor and Symbol*, 22 (3), 213-231.
- Snider, N. & Arnon, I. (2012). A unified lexicon and grammar? compositional and non-compositional phrases in the lexicon. *Frequency Effects in Language*, 127-163.
- Sosa, A. V. & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word of. *Brain and Language*, 83 (2), 227-236.
- Sprenger, S., la Roi, A. & van Rij, J. (2019). The development of idiom knowledge across the lifespan. *Frontiers in Communication*, 4, 29.
- Sprenger, S., Levelt, W. & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of Memory and Language*, 54 (2), 161-184.
- Strandburg, R., Marsh, J., Brown, W., Asarnow, R., Guthrie, D. & Higa, J. (1993). Event-related potentials in high-functioning adult autistics: linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia*, 31 (5), 412-434.
- Swinney, D. A. & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, 18 (5), 523-534.

- Tabossi, P., Arduino, L. & Fanari, R. (2011). Descriptive norms for 245 Italian idiomatic expressions. *Behavior Research Methods*, 43 (1), 110-123.
- Tabossi, P., Fanari, R. & Wolf, K. (2005). Spoken idiom recognition: meaning retrieval and word expectancy. *Journal of Psycholinguistic Research*, 34 (5), 465-495.
- Tabossi, P., Fanari, R. & Wolf, K. (2009). Why are idioms recognized fast? *Memory & Cognition*, 37 (4), 529-540.
- Taylor, W. L. (1953). "Cloze procedure": A new tool for measuring readability. *Journalism quarterly*, 30 (4), 415-433.
- Thornhill, D. E. & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83 (3), 382-392.
- Titone, D. A. & Connine, C. M. (1994). Descriptive norms for 171 idiomatic expressions: Familiarity, compositionality, predictability, and literality. *Metaphor and Symbol*, 9 (4), 247-270.
- Titone, D. A. & Connine, C. M. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics*, 31 (12), 1655-1674.
- Titone, D. A. & Libben, M. R. (2014). Time-dependent effects of decomposability, familiarity and literal plausibility on idiom meaning activation: A cross-modal priming investigation. *The Mental Lexicon*, 9 (3), 473-496.
- Titone, D. A., Lovseth, K., Kasparian, K. & Tiv, M. (2019). Are figurative interpretations of idioms directly retrieved, compositionally built, or both? *Canadian Journal of Experimental Psychology*, 73 (4), 216-230.
- Tremblay, A. & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Hrsg.), *Perspectives on Formulaic Language: Acquisition and Communication* (S. 151-173). London: Continuum.
- Tremblay, A., Derwing, B., Libben, G. & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, 61 (2), 569-613.
- Underwood, G., Schmitt, N. & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Hrsg.), *Formulaic Sequences* (S. 155-172). Amsterdam: John Benjamins.

- Van Den Brink, D., Brown, C. M. & Hagoort, P. (2001). Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. *Journal of Cognitive Neuroscience*, 13 (7), 967-985.
- van Ginkel, W. & Dijkstra, T. (2019). The tug of war between an idiom's figurative and literal meanings: Evidence from native and bilingual speakers. *Bilingualism: Language and Cognition*, 23 (1), 131-147.
- Van Lancker Sidtis, D. (2003). Auditory recognition of idioms by native and nonnative speakers of English: It takes one to know one. *Applied Psycholinguistics*, 24 (1), 45-57.
- Van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In *Advances in the Neural Substrates of Language: Toward a Synthesis of Basic Science and Clinical Research* (S. 342-367). London, UK: Blackwell Publishing.
- Van Lancker Sidtis, D., Cameron, K., Bridges, K. & Sidtis, J. J. (2015). The formulaic schema in the minds of two generations of native speakers. *Ampersand*, 2, 39-48.
- Van Petten, C., Coulson, S., Rubin, S., Plante, E. & Parks, M. (1999). Time course of word identification and semantic integration in spoken language. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25 (2), 394-417.
- Van Petten, C. & Luka, B. J. (2012). Prediction during language comprehension: benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83 (2), 176-190.
- Vespignani, F., Canal, P., Molinaro, N., Fonda, S. & Cacciari, C. (2010). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience*, 22 (8), 1682-1700.
- von Koss Torkildsen, J., Syversen, G., Simonsen, H. G., Moen, I. & Lindgren, M. (2007). Electrophysiological correlates of auditory semantic priming in 24-month-olds. *Journal of Neurolinguistics*, 20 (4), 332-351.
- Vulchanova, M. D., Vulchanov, V. & Stankova, M. (2011). Idiom comprehension in the first language: a developmental study. *Vigo International Journal of Applied Linguistics*, 8, 207-234.
- Wojcik, E. H. (2018). The development of lexical-semantic networks in infants and toddlers. *Child Development Perspectives*, 12 (1), 34-38.
- Wray, A. (1992). *The focusing hypothesis: the theory of left hemisphere lateralised language re-examined* (Bd. 3). John Benjamins Publishing.
- Wray, A. (2005). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

- Wray, A. (2012). What do we (think we) know about formulaic language? An evaluation of the current state of play. *Annual Review of Applied Linguistics*, 32, 231-254.
- Yang, S., Ahn, J. S. & Van Lancker Sidtis, D. (2010). The perception and acoustic features of Korean ditropic sentences. *The Journal of the Acoustical Society of America*, 127 (3), 1955-1955.