
WHAT MAKES A NEOLOGISM A SUCCESS STORY?

An Empirical Study of the Diffusion of Recent
English Blends and German Compounds

Sabrina Vanessa Link

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophie
in englischer Sprachwissenschaft
der Ludwig-Maximilians-Universität
München

vorgelegt von
Sabrina Vanessa Link
aus München

2021

Erstgutachter: Prof. Dr. Hans-Jörg Schmid

Zweitgutachter: PD Dr. habil. Wolfgang Schindler

Tag der mündlichen Prüfung: 21.05.2021

Zusammenfassung

Die vorliegende Arbeit untersucht empirisch die Diffusion von englischen Schachtelwort- und deutschen Kompositaneologismen. Im Rahmen des Projekts wurden vier Fragebogenstudien durchgeführt, anhand derer die Diffusion der jeweiligen Neologismen unter Muttersprachlern untersucht wurde.

Um ein besseres Verständnis für die durchgeführten Studien zu erlangen, widmet sich der erste Teil der Arbeit ausschließlich der zugrundeliegenden Theorie. Nach einer kurzen Einführung sowie der Vorstellung der Forschungsfragen wird der Theorieteil mit der Definition des Terms *Neologismus*, seine Einordnung in das englische Lexikon, die Unterscheidung zu *Ad-hoc-Bildungen* sowie die Darstellung des Lebenszyklus eines (neuen) Wortes eingeleitet.

Da die Arbeit sich der Diffusion von Neologismen widmet, ist es essenziell die Prozesse, die zur Etablierung neuer Wörter beitragen, genauer zu benennen und zu beschreiben. Namentlich handelt es sich hierbei um die eng miteinander verbundenen Prozesse der *Konventionalisierung* und *Verankerung (entrenchment)*. Während ersteres die soziale Komponente in der Etablierung neuer Wörter beschreibt, handelt es sich bei zweiterem um einen individuellen Prozess, der sich im Kopf einzelner Sprecher abspielt.

Der Vorgang der *Konventionalisierung* wird von zwei Katalysatoren angetrieben: der Diffusion und der Usualisierung. Da Diffusion und damit verbunden auch Usualisierung den Hauptgegenstand der vorliegenden Arbeit darstellen, werden sie im Theorieteil detailliert erläutert und definiert. Die Diffusion bezieht sich auf die Verbreitung des neuen Wortes, während es sich bei Usualisierung um einen Vorgang handelt, der dazu beiträgt, das wiederkehrende Worte sich etablieren und langfristig von den Sprechern als

Mittel angesehen werden, um kommunikative Ziele zu erreichen. Je höher die Diffusion und Usualisierung eines neuen Wortes, desto besser sind die Chancen, dass das Wort langfristig in Gebrauch kommt.

Vier verschiedene Aspekte sind bei der Usualisierung involviert: *Symbolisierung*, *Paradigmatikalisierung*, *Syntagmatikalisierung* und *Kontextualisierung*. Jeder dieser Vorgänge hat eine unterschiedliche Wirkung auf Usualisierung. Um dies genauer zu exemplifizieren: *Symbolisierung* ist für die Etablierung, Aufrechterhaltung und Anpassung onomasiologischer und semasiologischer Regularitäten verantwortlich, in Bezug auf die zur Erreichung eines kommunikativen Ziels ausgewählte Form, und die Bedeutung und pragmatischen Funktionen, die mit dieser Form verbunden sind.

Diffusion kann sozial und geographisch stattfinden. Es gibt verschiedene Modelle, die die Verbreitung neuer Innovationen im Allgemeinen, und sprachliche Innovationen im Speziellen, beschreiben. Diffusion wird von verschiedenen Faktoren begünstigt, wie zum Beispiel durch *Co-Semiosis*, eine Anpassung der SprecherInnen an die RezipientInnen, um schneller und leichter das intendierte kommunikative Ziel zu erreichen. Da Konventionalisierung ein Prozess ist, der nicht nur Wortneuschöpfungen betrifft, sondern auch andere Innovationen, werden zudem die beiden für Lexis spezifischen Unterprozesse – *Lexikalisierung* und *Institutionalisierung* – eingeführt und erklärt.

Während Usualisierung und Diffusion als Antrieb für Konventionalisierung dienen, gibt es auch äußere Kräfte, die den Konventionalisierungsprozess beeinflussen. Diese sind zum Beispiel Sprechersolidarität, das Prestige des Erfinders der Wortneuschöpfung, die Extravaganz des Wortes selbst, seine Anwendungsbereiche und seine Frequenz.

Eng verbunden mit der sozialen Komponente der Konventionalisierung ist der mentale Vorgang der Verankerung. Nur wenn sich neue Wörter im Kopf der SprecherInnen gespeichert werden, werden diese auch abgerufen und benutzt. Angetrieben wird dieser Prozess durch

die Katalysatoren *Routinisierung* und *Schematisierung*. Je routinierter und schematisierter ein Wort im Kopf der ProduzentInnen, desto höher die Wahrscheinlichkeit, dass es gut verankert ist und bei Bedarf nicht nur abgerufen wird, sondern sich auch gegen konkurrierende Formen durchsetzt. Während diese beiden Prozesse die Verankerung neuer Wörter, beziehungsweise neuer Innovationen im Allgemeinen, antreiben, gibt es auch einige zugrundeliegende psychologische Faktoren, die das Fundament für eine erfolgreiche Verankerung bilden. Dazu gehören statistisches Lernen sowie die Art und Weise wie Worte im Gedächtnis verfestigt werden – expliziter gesagt, ob sie ganzheitlich oder analytisch gelagert, abgerufen und verarbeitet werden. Neben psychologischen Faktoren wirken, ähnlich wie im Falle der Konventionalisierung, auch äußere Kräfte auf den Verankerungsprozess neuer Wörter. Dazu gehört zum Beispiel die Wortfrequenz, die Größe der Wortfamilie oder Analogien zu anderen, bereits existierenden Wörtern.

Konventionalisierung und Verankerung stehen in Wechselwirkung mit dem *Gebrauch (use)* von Innovationen. Je öfter eine Innovation verwendet wird, desto konventionalisierter und besser verankert wird sie. Gleichzeitig führen eine höhere Konventionalisierung und Verankerung dazu, dass ein Wort mit höherer Wahrscheinlichkeit verwendet wird. Abgesehen von Konventionalisierung und Verankerung haben auch noch andere Faktoren eine Auswirkung auf den Gebrauch. Dazu gehören zum Beispiel die Effizienz eines Wortes, seine Extravaganz sowie Sprechersolidarität und Machtgefüge.

Nachdem die Prozesse, die zur Etablierung neuer Wörter beitragen, genauer erläutert wurden, definiert die Arbeit die Termini *Schachtelwort (Blend)* und *Kompositum*, da im Rahmen des praktischen Teils englischen Schachtelwörter und deutsche Komposita untersucht werden. Schachtelwörter wurden in der Sprachwissenschaft lange marginalisiert. Aus diesem Grund ist die Forschung zu diesem Thema überschaubar, vor allem in Bezug auf Neologismen. Linguisten haben verschiedene Ansätze entwickelt, um Schachtelwörter strukturell voneinander zu unterscheiden. Es gibt zum Beispiel welche, die sich durch das Weglassen von Wortteilen

definieren lassen oder solchen, bei denen sich Wortteile überlappen. Abgesehen von Versuchen der Kategorisierung von Schachtelwörtern gibt es in jüngerer Zeit auch vermehrt Untersuchungen, die sich der Verarbeitung von Schachtelwörtern sowie deren Erkennung durch MuttersprachlerInnen widmen. Diese Arbeit zitiert einige Studien und ihre Ergebnisse, die sich jedoch oftmals nicht auf Neologismen, sondern auf bereits etablierte Schachtelwörter beziehen.

Im Gegensatz zu Schachtelwörtern stehen Komposita schon lange im Fokus der Forschung. Nichtsdestotrotz finden sich widersprüchliche Aussagen im Diskurs, angefangen von unterschiedlichen Auffassungen wie man Komposita definiert. Diese Unstimmigkeit basiert unter anderem darauf, dass ForscherInnen verschiedene Meinungen vertreten, ob Komposita in den Bereich der Morphologie oder Syntax fallen. Divergierende Terminologie lässt sich auch in der semantischen und strukturellen Einteilung von Komposita finden sowie in Bezug auf deren mental Produktion und Verarbeitung.

Der theoretische Teil dieser Arbeit schließt mit einem Kapitel, das sich mit der Methodik zur Neologismenaufzeichnung und -analyse beschäftigt, inklusive der Veränderungen, die das Internet mit sich gebracht hat. Da es Korpora oftmals nicht gelingt, lexikalische Veränderungen in Echtzeit darzustellen, da sich viele von Ihnen zum einen nur auf bestimmte Textarten stützen und zum anderen oftmals nicht in kurzen Zeitabständen auf den neusten Stand gebracht werden, spielt das Internet als Korpus eine immer größer werdende Rolle in der Neologismenforschung. Da auch dies einige Probleme mit sich bringt, haben LinguistInnen verschiedene Werkzeuge entwickelt, die dazu dienen, das Internet als Korpus zu benutzen. Des Weiteren werden in diesem Kapitel die methodischen Vor- und Nachteile von Onlinefragebogenstudien, welche die Grundlage des empirischen Teils dieser Arbeit darstellen, erörtert.

Der empirische Teil stellt die vier durchgeführten Fragebogenstudien detailliert vor. Die erste Studie dient dazu, mehrere Faktoren, die die Diffusion, Usualisierung und letztlich den Gebrauch von neuen Wörtern

beeinflussen können, zu untersuchen. Basierend auf den bereits gewonnenen Erkenntnissen, die im theoretischen Teil diskutiert wurden, erforscht die erste Studie mehrere Faktoren, die die Konventionalisierung eines Wortes antreiben. Dazu gehört zum Beispiel die Wortfrequenz, das ‚Alter‘ des Wortes sowie das Prestige der ersten ‚NutzerInnen‘. Des Weiteren werden auch die Eigenheiten des Wortes, wie seine semantische und formale Transparenz, seine Erkennbarkeit und sein *Appeal* - wie gut SprecherInnen ein neues Wort formal als auch konzeptionell finden - untersucht. Demografische Einflüsse auf das Kennen und Nutzen neuer Wörter werden ebenfalls in Betracht gezogen. All diese Faktoren werden anhand von verschiedenen Fragen in einen Fragebogen eingewoben. Bevor dies geschehen kann, müssen allerdings erst passende Neologismen extrahiert werden. Der methodische Teil der ersten Studie gibt detailliert Auskunft darüber, woher und anhand welcher Kriterien Neologismenschachtelwörter ausgesucht wurden. Das wichtigste Ergebnis der quantitativen Datenauswertung ist, dass der Faktor *Appeal* die signifikanteste Auswirkung auf die Variable *Gebrauch eines Neologismus* hat. Bei genauerer Betrachtung wurde zudem festgestellt, dass die vorgenommene Unterscheidung zwischen formalen und konzeptionellen *Appeal* von den Daten statistisch nicht unterstützt wird. Aufgrund des großen Einflusses von *Appeal* auf die Diffusion der untersuchten Neologismen, sollte dieser Faktor in einer Folgestudie genauer untersucht werden.

Die zweite Studie widmet sich dementsprechend einer genaueren Definition des Terms *Appeal*, vor allem unter Berücksichtigung der Tatsache, dass formaler und konzeptioneller *Appeal* statistisch nicht unterschieden werden können. Basierend auf der existierenden Forschung und den Daten des qualitativen Teils der vorausgehenden Studie, wurden Unterkategorien für *Appeal* erstellt, die eine genauere Beschreibung ermöglichen sollten. Die Subkategorien sind *Effizienz*, *Extravaganz* und *außersprachliche Relevanz*. Diese sind wiederum genauer definiert, so dass *Effizienz* Wörter beschreibt die präzise sind, eine klare Bedeutung haben und Analogie aufweisen. *Extravagante* Wörter sind davon geprägt, dass sie lustig und kreativ sind oder ein Wortspiel darstellen.

Außersprachliche Relevanz bezieht sich darauf, ob das Wort einen Trend, ein aktuelles Event, eine Innovation beschreibt oder von zeitgenössischer Relevanz ist. Mit Hilfe dieser drei Subkategorien und ihren jeweiligen Charakteristika wurde ein statistisches Modell erstellt, das auf die zweiten Fragenbogenstudie angewandt wurde. Das Ergebnis zeigt, dass das erstellte Modell *Appeal* sehr gut beschreibt. Darüber hinaus konnte festgestellt werden, dass die drei gefundenen Faktoren voneinander unterscheidbar sind. Während in der ersten Studie keine genauere Beschreibung oder Unterteilung von *Appeal* gelang, zeigten die Daten der zweiten Studie, dass eine genauere Definition von *Appeal* möglich ist, indem man den Faktor in *Effizienz*, *Extravaganz* und *außersprachliche Relevanz* unterteilt. Zudem ging aus den Daten hervor, dass alle drei Subkategorien von *Appeal* einen signifikanten Einfluss auf den Gebrauch und somit die Diffusion und Usualisierung von Neologismen haben.

Ausgehend von der zweiten Studie, sollte die nächste Studie herausfinden, ob sich das gefundene Modell für die Einteilung von *Appeal* auch auf eine andere Sprache übertragen lässt, in diesem Fall Deutsch. Dementsprechend wurden in einem ersten Schritt deutsche Neologismen gesammelt, die untersucht werden sollten. Da im Deutschen Schachtelwörter nur eine geringfügige Rolle spielen, befasst sich diese Studie mit Kompositaneologismen. Nachdem geeignet Wörter gefunden wurden, wurden diese in einem Fragebogen, der eine exakte Übertragung des vorherigen Fragebogens vom Englischen ins Deutsche darstellt, integriert und untersucht. Das Ergebnis der deutschen Fragebogenstudie zeigt, dass auch im Deutschen das dreigeteilte Modell zur Beschreibung von *Appeal* geeignet ist. Dementsprechend bestätigt diese Studie, dass eine Dreiteilung zwischen *Effizienz*, *Extravaganz* und *außersprachlicher Relevanz* gut geeignet ist, um *Appeal* genauer zu beschreiben. Darüber hinaus wurde untersucht, ob alle dieser drei Subkategorien auch im Deutschen einen signifikanten Einfluss auf den Gebrauch von Neologismen haben. Das Resultat zeigt, dass Effizienz und Extravaganz die zwei Hauptfaktoren sind, die dazu führen, dass PartizipantInnen ein neues Wort

benutzen. Außersprachlich Relevanz spielt in den deutschen Daten keine signifikante Rolle.

Nachdem die letzten zwei Studien gezeigt haben, dass eine Dreiteilung von *Appeal* in den Daten vorliegt und die drei gefundenen Subkategorien *Appeal* gut beschreiben – sowohl im Englischen als auch im Deutschen - dient eine letzte, kleinere Untersuchung dazu, herauszufinden, inwiefern *Appeal* ein guter Prädiktor für den zukünftigen Erfolg eines Neologismus ist. Hierzu wurden zwei verschiedene kleine explorative Studien durchgeführt. Die erste Untersuchung sollte herausfinden, ob *Appeal* langfristig ein guter Indikator ist und den Erfolg eines Neologismus über einen längeren Zeitraum begünstigen kann. Dazu wurden Neologismen, die bereits in der ersten Studie verwendet wurden, nochmals untersucht. Einige der Neologismen schienen zur Zeit der Durchführung der ersten Studie Erfolgskandidaten, während andere eher untauglich schienen. Zwei Jahre nach der ersten Studie wurde die Frequenz dieser Wörter nochmals betrachtet und es konnte festgestellt werden, dass *Appeal* sich nicht als langfristiger Prädiktor eignet, da die Frequenz der untersuchten Neologismen zwei Jahre später nicht unbedingt den Erwartungen entsprach. Im zweiten Teil der kleinangelegten explorativen Studie sollte der unmittelbare Effekt von *Appeal* untersucht werden. Dazu wurden aktuelle Neologismen extrahiert und von einer kleinen Zahl an TeilnehmerInnen in Bezug auf ihren *Appeal* bewertet. Die Entwicklung der Wörter wurde dann über die nächsten Wochen beobachtet und man konnte sehen, dass die Wörter, die von den TeilnehmerInnen als *more appealing* angesehen wurden, auch eine höhere Frequenz und demnach Diffusion aufwiesen. Dementsprechend scheint es, als sei *Appeal* ein wichtiger Faktor am Anfang des Lebenskreises eines neuen Wortes, da er dem Wort einen extra Antrieb verschaffen kann. Langfristig kann aber die Diffusion nicht nur von *Appeal* aufrechterhalten werden.

Abschließend kann man sagen, dass es der vorliegenden Arbeit gelungen ist, einen Faktor zu isolieren und genauer zu beschreiben, der die Diffusion von Neologismen in Schwung bringen kann und ihnen somit eventuell zum Erfolg verhelfen kann. Um allerdings langfristig erfolgreich zu

sein, müssen auch noch andere Kriterien in Betracht gezogen werden. Dementsprechend bieten die Ergebnisse dieser Arbeit einen guten Ausgangspunkt für weitere Forschung in diesem Bereich.

Contents

List of figures, tables, and graphs	16
Acknowledgements.....	18
1. Introduction	19
1.1 General introduction.....	19
1.2 Research questions	22
1.3 Outline.....	27
i. Theoretical background.....	31
2. Neologisms	31
2.1 What is a neologism?.....	31
2.2 Neologisms and the English lexicon.....	34
2.3 Neologisms versus nonce-formations	36
2.4 The life of a neologism.....	38
3. From nonce-formation to an established word: social, socio-pragmatic, structural, and cognitive processes.....	43
3.1 Conventionalisation	43
3.1.1 Catalysators for conventionalisation	46
3.1.2 Forces affecting conventionalisation	64
3.1.3 Conventionalisation of words: Institutionalisation and Lexicalisation as subprocesses of Conventionalisation.....	73
3.1.4 Conventionalisation summed up	83
3.2 Entrenchment	84
3.2.1 Catalysators for entrenchment.....	85
3.2.2 Psychological foundations for entrenchment.....	90
3.2.3 Forces affecting entrenchment.....	102
3.2.4 Entrenchment of lexical items	111
3.3 What's it all about: usage.....	113

4.	Neologisms and the word-formation processes of blending and compounding	124
4.1	Blending.....	124
4.1.1	Defining blends	124
4.1.2	The structure of blends.....	126
4.1.3	Identifying and processing blends	130
4.2	Compounding.....	133
4.2.1	Defining compounds	133
4.2.2	The structure of compounds.....	136
4.2.3	Accessing and processing compounds.....	139
5.	Keeping up with the pace: neologisms on the internet– changes in methodology	142
ii.	Empirical part	151
6.	Questionnaire study I	151
6.1	Method	151
6.1.1	Materials - retrieving neologisms	151
6.1.2	Questionnaire design and participants.....	156
6.1.3	Statistical analysis	161
6.2	Quantitative results.....	162
6.2.1	Demographic data: Age, Education, Location, Internet, and Social Media Usage	162
6.2.2	Frequency, first occurrence, and quality of media type/influence of the early user group.....	165
6.2.3	Participants' interest.....	169
6.2.4	Recognisability, transparency, and appeal	170
6.2.5	Further analysis – recognisability, transparency, and appeal.....	174
6.3	Qualitative results	178
6.4	Discussion	180

7. Questionnaire study II	190
7.1 Research object and questions.....	190
7.2 Method	193
7.2.1 Materials – retrieving neologisms.....	193
7.2.2 Questionnaire design and participants.....	194
7.2.3 Confirmatory Factor Analysis (CFA)	196
7.3 Quantitative results for appeal.....	197
7.4 Qualitative results	205
7.5 Discussion	208
8. Questionnaire study III	212
8.1 Research object and questions.....	212
8.2 Method	214
8.2.1 Materials – retrieving neologisms.....	214
8.2.2 Questionnaire design and participants.....	216
8.3 Quantitative Results.....	220
8.3.1 Demographic data: Age, Education, Location, and Internet usage	220
8.3.2 Frequency.....	220
8.3.3 Appeal	221
8.4 Discussion	224
9. Explorative follow-up Investigation	228
9.1 Research questions	228
9.2 Method	229
9.2.1 Materials – retrieving neologisms.....	229
9.2.2 Assessment and questionnaire design.....	232
9.3 Results.....	233
9.4 Discussion	236
10. Summary and conclusion	239
10.1 Overall summary of the results	239
10.2 Outlook	242

Appendices	247
Appendix 1 Questionnaire study I – Materials and data	247
Appendix 1.1 Preliminary neologism list for study I	247
Appendix 1.2 Definitions of the chosen neologisms for study I	248
Appendix 1.3 Neologisms according to their division into private and professional and their first appearance	250
Appendix 1.4 Questionnaire study I - sample	255
Appendix 1.5 Neologisms and distractors split in questionnaire I & II within study I	264
Appendix 1.6 Complete overview of the demographics of all participants in study I	265
Appendix 1.7 Overview of all outcomes regarding the demographics of study I	267
Appendix 1.8 Different social media platforms used within different age groups.....	267
Appendix 1.9 Time spent on the internet in relation to knowing a neologism – graph. 268	
Appendix 1.10 Media types and early user group of neologisms.....	269
Appendix 1.11 Media types and early user groups ranked by participants according to their quality and influence respectively.....	270
Appendix 1.12 Statistical details for categorisation of quality of media and influence of early user groups	271
Appendix 1.13 Graphs for quality and influence ratings in relation to use	272
Appendix 1.14 Participants’ perception of media the neologisms mainly first occurred in	273
Appendix 1.15 Categorisation of neologisms for interest	274
Appendix 1.16 All overall values for recognisability, transparency, appeal, and use.....	275
Appendix 1.17 Data sample <i>glampsite</i>	277
Appendix 1.18 Absolute values of motives for usage	278
Appendix 2 Questionnaire study II – Materials and data	279
Appendix 2.1 Questionnaire study II - sample	279
Appendix 2.2 Complete overview of the demographics of questionnaire study II	285
Appendix 2.3 Overview of non-significant values regarding demography study II	286
Appendix 2.4 Standardised factor loadings CFA II	286
Appendix 2.5 Average rating use and appeal per word.....	286
Appendix 3 Questionnaire study III - Materials and data	287
Appendix 3.1 Preliminary word list for the third study	287
Appendix 3.2 Definitions.....	288
Appendix 3.3 Questionnaire study III - sample	289

Appendix 3.4 Complete overview of the demographics for questionnaire III.....	293
Appendix 3.5 Overview of non-significant values regarding demography study III	295
Appendix 3.6 Model fit CFA I versus CFA II for questionnaire study III	296
Appendix 3.7 Standardised factors loadings for CFA I for questionnaire study III	297
Appendix 3.8 Standardised factor loadings for CFA II for questionnaire study III	298
Appendix 4 Follow-up experiment.....	299
Appendix 4.1 Preliminary word list for the fourth study	299
Appendix 4.2 Definitions.....	299
Appendix 4.3 Questionnaire study IV - sample.....	300
Appendix 4.4 Complete overview of the demographics for questionnaire IV	304
Appendix 4.5 Efficiency, extravagance and extralinguistic relevance rating - study IV..	305
Appendix 5 Forces affecting usage, conventionalisation and entrenchment taken from Schmid, 2020: 300.....	305
Bibliography	306

List of figures, tables, and graphs

Figure 1 Modified version of the English vocabulary based on the Shorter Oxford English Dictionary (Lipka, 1992: 10)	35
Figure 2 Life cycle of a neologism (Behera & Mischra, 2013: 27)	41
Figure 3 Idealised conformity profiles for five kinds of utterance types and their dominant conformity types (Schmid, 2020: 96)	46
Figure 4 Schmid's language as a simple Tinguely machine illustrating the EC-model (Schmid, 2020: 4)	47
Figure 5 Adaptor categorisation on the basis of innovativeness by Rogers, 1995: 262	54
Figure 6 Spread of new lexical invention (Würschinger, 2019)	55
Figure 7 Three level model of accommodation and language change (Auer & Hinskens, 1996: 22)	58
Figure 8 Factors affecting token frequency (Fernández-Domínguez, 2010: 202)	81
Figure 9 Relation between conformity profiles and types of associations (Schmid, 2020: 298)	89
Figure 10 Regression result for use, formal, and semantic transparency	175
Figure 11 Mediation model transparency - appeal - use	177
Figure 12 Mediation model recognisability - appeal – use	178
Figure 13 Regression result for use, formal, and conceptual appeal from study II	198
Table 1 Morphological matrices according to Tournier, 1985: 51	40
Table 2 Excerpt from Schmid's Survey of modules of change - for full detail see Schmid, 2020: 330-331	62
Table 3 Different ways to depict the impact of string frequency on mental representation (Blumenthal-Dramé, 2017: 131)	99
Table 4 The three perspectives and three stages of the establishment of new words by Schmid (2008), adapted by Kerremans, 2015: 40	113
Table 5 Chosen neologisms for study I	152
Table 6 Frequency averages of the chosen words with purple marking high-frequency words, and red marking low-frequency words	154
Table 7 Distractors for questionnaire study I	158
Table 8 Distribution amongst participants in questionnaire study I	160
Table 9 Participants per category	161
Table 10 Subcategories of appeal based on the previous study and research	192
Table 11 Neologisms chosen for the questionnaire study II	194
Table 12 Distribution amongst participants in questionnaire study II	196
Table 13 Values indicating model fit for CFA (I)	200
Table 14 Standardised factor loadings for CFA (I) depicted in graph 17	201
Table 15 Values indicating model fit CFA I in comparison to CFA II	202
Table 16 Values indicating model fit for one factor CFA	204
Table 17 Percentages of how neologisms are perceived by participants	207
Table 18 Chosen neologisms for study III	215
Table 19 Frequency values for the chosen neologisms for study III	216
Table 20 Distractors study III	217
Table 21 Distribution amongst participants in questionnaire study III	219
Table 22 Values indicating model fit for German data for CFA (II)	223
Table 23 New neologisms chosen for study IV	232
Table 24 Average frequency score for neologisms from study I & II during their first occurrence and in 2020	233
Table 25 Average frequency score in 2020 for recent neologisms	235

Graph 1 Influence of participants' age on renown of neologisms	164
Graph 2 Social media use graded by age.....	164
Graph 3 Frequency of the neologisms on the internet in relation to renown of neologisms amongst participants	166
Graph 4 Frequency of the neologisms on the internet in relation to participants using them.....	166
Graph 5 First occurrence in relation to use and renown of neologisms	167
Graph 6 Rated quality of media types and use of the neologisms from the respective categories.....	168
Graph 7 Rated influence of early user groups and use of the neologisms from the respective categories.....	169
Graph 8 Correlation between recognisability and use	171
Graph 9 Correlation between formal transparency and use	172
Graph 10 Correlation between semantic transparency and use	172
Graph 11 Correlation between formal appeal and use	173
Graph 12 Correlation between conceptual appeal and use.....	174
Graph 13 Correlation between general appeal and use.....	174
Graph 14 Correlation between formal and semantic transparency	175
Graph 15 Correlation between formal and conceptual appeal	176
Graph 16 Correlation between formal and conceptual appeal in study II	198
Graph 17 CFA (I).....	199
Graph 18 CFA (II).....	202
Graph 19 One factor CFA.....	203
Graph 20 Impact of efficiency, extravagance and extralinguistic relevance on use for questionnaire study II	204
Graph 21 Frequency of the neologisms on the internet in relation to renown of neologisms amongst participants.....	221
Graph 22 CFA II adapted to the German questionnaire study.....	222
Graph 23 Impact of efficiency, extravagance and extralinguistic relevance on use for questionnaire study III	223
Graph 24 Appeal rating and use for selected high- and low-frequency words from study II	231
Graph 25 Appeal rating for recent neologisms	235

Acknowledgements

Already after the first day of my Introduction into English Linguistics seminar, I was fascinated by the different fields that linguistics offer, and I soon aspired to research linguistic phenomena myself. Thus, a long journey began, and this thesis depicts the outcome of several years of intensive work, which would not have been possible without the profound support and advice I received from various people.

First and foremost, I would like to thank Prof. Dr. Hans-Jörg Schmid, who supervised my dissertation project over the years at Ludwig-Maximilians-University in Munich. He enabled me to freely develop my own ideas while at the same time steering me into the right direction by supporting me with his vast expertise. Further, I owe special thanks to Quirin Würschinger and Dr. Daphné Kerremans who always took their time for me and patiently answered all my questions. They came forward with many great ideas and contributions to this project. Many thanks go to PD Dr. habil. Wolfgang Schindler who kindly took on this project as a second assessor and Prof. Dr. Stephanie Hackert who agreed to be the third member of the examination commission.

Special thanks also go to Dr. Devin Ray, who was extremely forbearing with me and offered me such great support, both as a friend as well as professionally. He taught me so many new things – foremost about statistics - without which this project could have not been realised in the way it was.

I am also grateful to all my colleagues from the German Department and the School of Languages, Literature, Music and Visual Culture at the University in Aberdeen, who supported me throughout my project in various ways, from distributing and taking part in my studies to proofreading the thesis.

Special thanks also go to my parents, Max and Hedwig Link, who have been supportive and encouraging from the very beginning. They always believed in me and taught me that I can become whatever I choose to be. I would also like to thank my sister Verena Link, who always had an open ear for me and gave me advice whenever needed. I would further like to thank my friends who always encouraged me and lifted me up and who constantly sent me new words they found online, as well as patiently helped me by participating in some of my studies and distributing them. Lastly, I want to express deep gratitude to Massimo Prosdocimi for patiently listening to all my linguistic elaborations and for his love, encouragement, and support.

1. Introduction

1.1 General introduction

Since January 2020, the whole world has been dictated by a pandemic. A major event as such, which affects the whole globe and the life of each and every one of its inhabitants, also impacts and changes the language we use. Within days, the words *corona virus* and *COVID-19* were used in the media all over the world and soon spread and diffused amongst speakers. No matter in which part of the planet you are right now, everyone will understand these terms. *Corona viruses* denote a group of “any of a family (*Coronaviridae*) of large single-stranded RNA viruses that have a lipid envelope studded with club-shaped spike proteins, infect birds and many mammals including humans, and include the causative agents of MERS, SARS, and COVID-19” (Merriam-Webster Online Dictionary). While the term *corona virus* has been well-known and established amongst scientists within the field of virology, just now, with a raging pandemic going on, it has diffused amongst laymen. In contrast to this pre-existing, scientific term *Corona*, *COVID* is a blend of *co(rona)vi(rus)* and *d(isease)* and was coined by the WHO in February 2020 (BBC News Asia, 2020). Thus, *COVID-19* is a man-made term that, due to its importance for our daily lives, spread within no time.

However, it did not stop there. Some pre-existing words have suddenly been used with an increased frequency and many new words have come into being since early 2020. As a reaction to the sudden outbreak of the virus, governments have imposed *lockdowns* on their country. The word’s roots go back to Old English *loc* and Middle English *doun* (MacMillan Dictionary Blog), and *lockdown* is defined as “an occasion or time when access to a place is restricted because of some danger” (MacMillian Dictionary Blog). In case of the recent use though, it refers to whole countries being locked down, including people not being allowed to leave their homes. Thus, it could be argued that a new meaning, or at least a new facet has evolved. Other examples of words with a recent cumulative

frequency are *self-isolation*, *quarantine*, *keyworkers/essential workers*, *face masks* (especially since the current usage does not have anything to do with a spa day) and the term *pandemic* itself (Lawson, 2020). Along with the imposed lockdown went the requirement of *social distancing*, which implies “measures that can help to slow the spread of an infectious disease by avoiding close contact between people” (MacMillan Dictionary Blog). The WHO suggested and favoured the term *physical distancing*, as this indicates that people can still socialise by using technology (Tangermann, 2020; Kumar, 2020). However, this term never properly caught on.

Besides these rather official terms that are often imposed by governments and spread by mass media, many creative new words came into existence. Once in quarantine for a while, people started to enjoy *quarantinis* or *locktails* (Mahdawi, 2020). If someone does not stick to the public health advice, they can easily be denoted as a *covidiot*. Due to *social distancing*, people decided to have *covideo parties* via Zoom or Skype. Others spent their time *doomscrolling*, and thus obsessively follow the depressing pandemic news (Mahdawi, 2020). In the meantime, politicians tried to come up with *covexit* strategies, to make sure lockdown can be left sooner rather than later (Lawson, 2020). The sheer number of new words invented in connection with the pandemic has started to attract the interest of linguists (cf. Asif et al., 2020) and was even incorporated in form of a COVID glossary at the museum exhibition ‘Now accepting contactless’ at the V&A, Dundee.¹

The stated instances exemplify how creative language, and its speakers can be. When it comes to the invention of new words “no external driving force of any kind has any influence “on language”, without going through the freedom and the intelligence of the speakers” (Coseriu, 1958/1973:196). The speaker’s creativity, however, is not enough. For lexical changes to happen, a new word needs to be picked up and used by the majority of speakers. The case of *physical distancing* shows, that not every innovation is adapted and sometimes competing expressions, such

¹ <https://www.vam.ac.uk/dundee/exhibitions/now-accepting-contactless>

as in this case *social distancing*, are more successful. New words usually come into being under tough conditions and need to fight their way into the language in a Darwinian struggle for survival (Metcalf, 2002: 2). In a pre-internet, dictionary-based study, Algeo found that out of the 3565 terms he studied, only 1515 (=42%) survived, while all others sank back into obscurity (Algeo, 1993: 283). Nowadays, the figures might even be worse. With the internet being present in most parts of our lives, we are flooded with news, pictures, and posts. In order to get noticed, people try to come up with the most eye-catching captions, posts or tweets. It seems to be a recent trend to coin “catchy new expressions to appealingly characterise a topical scenario” (Moseley, 2016).

A relatively recent ‘flagship model’ for a catchy new expression that succeeded is *Brexit*. Having been modelled on *Grexit* and competing initially with *Brixit* - *The Economist* and the *Daily Mail* titled in 2012: “A Brixit looms” (Wooldridge, 2012), “Bring on the ‘Brixit’” (Murray, 2012) - *Brexit* soon started its victory parade. The word skyrocketed, especially after the referendum in 2016, and within a short while it has been conventionalised, institutionalised, and lexicalised and was added to reference dictionaries (Lalić-Krstin & Silaški: 2018: 3). The media cover of the word further helped its popularity and soon other terms based on *Brexit* were coined, such as *Calexit* (California leaving the US), *Scotxit* (Scotland leaving the UK) and *Megxit* (Megan and Harry leaving the royal family) (Lalić-Krstin & Silaški: 2018: 4). It seems as if the ‘-xit’ part has become productive and started acting like a suffix (Moseley, 2016). Time will show which of the newly coined words based on *Brexit* will survive, but one thing is clear, *Brexit* is a well-established term in our current vocabulary.

While *Brexit* was found as a word for an increasingly important political phenomenon (Lalić-Krstin & Silaški: 2018: 3), other new words suddenly emerge without describing topical scenarios. An example is *hangry* which denotes ‘reacting angrily because you are hungry’, a scientifically proven phenomenon (Salis, 2015). While the phenomenon itself surely did not just recently come into being, it seems to have received an increased public interest lately so that *hangry* was added to the Oxford

Dictionary in 2018. This shows the productive capacity language has to generate a new, well-formed linguistic sign by means of its productive word-formation rules whenever the need arises (Štekauer, 2002: 101). While *Brexit* and *hangry* are two successful examples, the question arises as to what distinguishes them from the majority of words that do *not* make it.

Thus, the question arises whether there are factors that could determine a successful diffusion and conventionalisation of a word? Factors that could give us information about what makes a neologism the ‘fittest’ for survival? Within the scope of this thesis, I will try to shed some light onto these questions. I will investigate different new words and, by using questionnaire studies, I will examine what factors might contribute or hinder their diffusion, amongst others the participants’ background, the influence of the nature of the word as well as effects of the extralinguistic environment of speakers.

1.2 Research questions

While it is argued that “[p]redicting the linguistic future is always a dangerous activity” (Crystal, 2003: 76), several authors have tried to get to the bottom of foretelling lexical change. Allan Metcalf describes in *Predicting New Words* (2002) the FUDGE scale, which entails factors that, according to his experience, influence the success of new coinages. FUDGE stands for **F**requency, **U**nobtrusiveness, **D**iversity of users and situations, **G**eneration of other forms and meanings, and **E**ndurance of the concept. Göran Kjellmer wrote an article entitled *Potential Words* (2000), elaborating which factors might limit either a new word from coming into existence or its success after being coined.

In the framework of this thesis, I will test several independent variables. They include

- the neologisms’ appeal, transparency, frequency and first occurrence,
- the coiner’s prestige and influence,
- the participants’ personal socio-demographic background.

These independent variables might allow to make predictions in regard to the dependent variable, the diffusion and with it also the usualisation of a new word (cf. sections 3.1 ff. & 3.3), and thus in a way a neologism's success.

While a word's potential as well as its realisation and existence might be seen as rather relative, there are several linguistic conditions that make a word possible or impossible (Kjellmer, 2000: 206). However, whether a word is existing or non-existing depends on changes in society (Kjellmer, 2000: 207). The linguistic conditions that make a word feasible can be of semantic, phonological, morphological and graphemic nature (Kjellmer, 2000: 206). The two factors considered here refer to the semantic and morphological conditions of neologisms. To start off with semantics, *semantic transparency* is an important factor for potential words. The easier it is to comprehend a new word, the more successful it will be when competing for lexicalisation (cf. section 3.1.3) (Kjellmer, 2000: 207). The acronym *BSE*, for instance, stands for the rather complex "Bovine spongiform encephalopathy". It was soon known as *mad cow disease* in colloquial speech, giving a much better insight into what it is (Ayto, 1990) and consequently easing and speeding up the comprehension of the word (Schmid, 2008: 13) (cf. section 3.2 ff.). Apart from semantic transparency, morphological parallels and analogies like *active – activity, pensive – *pensivity* create *formal transparency* (Schmid, 2008: 13) and might be preferred (Kjellmer, 2000: 212, 221). Taking this into account I assume that:

H1: Words that are formally and semantically transparent to speakers are more likely to be used.

Apart from the clear structure and meaning of a neologism, it is assumed that familiarity has an impact on whether it will be used. Since even formally and semantically transparent words might not be recognised and/or understood.

H2: Words that are recognised by speakers are more likely to be used.

Another factor that could play a role in the diffusion of neologisms is appeal. Two types of appeal are distinguished in this thesis: *conceptual* and *formal appeal*. Since both terms are rather subjective, I will briefly state how I define them in the framework of this thesis. I denote conceptual appeal as ‘speakers liking the concept behind the word and finding the word/concept useful in the current society’. This bears similarities to the ‘E’ in Metcalf’s FUDGE scale (cf. section 1.1). ‘E’ stands for *endurance of the concept*. *Typewrite*, for instance, has a relatively low endurance level, as it is not very relevant nowadays, but is still widely known. In contrast to this, a well-know, widely spread and useful word like *icebox* has a higher endurance level and, in my terminology, a higher conceptual appeal (Metcalf, 2002: 162 ff.). It should be noted though that, while Metcalf’s *endurance* refers to a longer period, my definition and understanding of conceptual appeal refers to a certain point in time.

Formal appeal combines several different aspects. One of them is humour: funny words like *bookfairy* (‘a person that sells or buys antique books from bookfairs’) are often more likely to catch on (Kjellmer, 2000: 215). Furthermore, formal appeal can also refer to the phonological or orthographical form of a word. A word’s success could be boosted if it bears phonological parallels to well-established sound patterns and can be easily pronounced (Kjellmer, 2000: 209-210). Orthographically, words that start or end with non-English syllables and letters might cause the speaker to find the word less attractive as there is a lack of information on whether pronunciation and spelling agree with each other according to established conventions. It could, however, also be the other way around. A word could be formally especially appealing if it is not very English as it indicates some prestige or exoticness (Kjellmer, 2000: 2012). One last subfactor of formal appeal is how concise a word is. English favours short words and forms, as the incredible number of acronyms and shortenings, like *poll tax* for *community charge*, show (Kjellmer, 2000: 214). While the notion of formal appeal is rather complex, it is considered as an important factor. A general appeal, which could be seen as a mixture of these two types of appeal, will also be investigated. Thus, I hypothesise that:

H3: Words that are formally and conceptually appealing to speakers are more likely to be used.

H4: Words that are generally more appealing to speakers are more likely to be used.

F on Metcalf's FUDGE scale stands for frequency (Metcalf, 2002: 152 f.). Even easily predictable and thus transparent words (cf. H1) do not seem to catch on, if they are not used frequently enough (Algeo, 1993: 286). Frequency is not a straight-forward notion (cf. chapter 3.2.3). In the framework of this thesis frequency of words on the internet will play an important role in this thesis regarding two different aspects. On the one hand, frequency could have a positive effect on usage, as the more frequent words already are, the more likely they are used by speakers (H5). On the other hand, frequency is often treated as the measurable effect of conventionalisation and hence diffusion (Schmid, 2020: 6) (cf. chapter 3.1.1). This implies that neologisms that are more frequent on the internet are potentially also more diffused offline (H6). Another factor that might influence the use of neologisms is not only their frequency and diffusion, but also how long they have been around for:

H5: Words that are more frequent and diffused on the internet are more likely to be used by speakers.

H6: Neologisms that are more frequent and diffused on the internet are also more likely to be frequent and diffused offline.

H7: Words that have been used for a longer period online are more likely to be known to and used by speakers.

Another factor that needs to be considered is the person that invented or coined the new word. Prestige acts as a driving force for processes such as diffusion, usualisation and borrowing (Lutz, 2013). Despite its subjectivity – the distinction between overt and covert prestige already aims to show how relative prestige is regarding its context – it is an important factor that has to be considered (Schmid, 2020: 114). In regard to neologisms, especially the prestige of the *innovator* (Labov, 1980) or *coiner* (Kerremans, 2015)

plays a major role. Since the internet is used as a source, it can be very difficult to find a specific coiner, therefore it is instead assumed that the early user group and media type have an impact on a word's diffusion:

H8: A word that is initially primarily used by a more influential early user group and a high-quality media type is more likely to be used by speakers.

However, not only prestige can have an influence on a neologism's chance to be taken up but also its domain. This becomes clear when looking at jargon. A term like *phoneme* is frequent and widely used and thus usualised to a high degree amongst linguists. However, it is hardly diffused or conventionalised outside this field (Schmid, 2020: 93 ff.). This probably does not only apply to professional fields, but also private interests. Therefore, I assume that:

H9: A word that comes from an area a person is interested in is more likely to be known to and used by the speaker.

Besides all the language related factors that lay the foundation for this thesis and that will be investigated throughout, I will also look at whether the background of the participants has an impact on their usage of new words. I assume that, as young people generally manage change better than older people, they are also more receptive to new lexical innovations. Furthermore, people who live in cities and are therefore constantly surrounded by many other human beings as well as by change, might also be more responsive to new words. Education could play a role, too. Highly educated people might be more flexible regarding their language and adapt quicker, also to new lexical items. Lastly, as all the researched words are taken from the internet, people who spend more time on the internet are maybe also more familiar with the new words and lean more towards using them. Social media platforms are a particularly rich source for neologisms and different types of social media are also used for as web-based corpora for academic neologism research (Kerremans, 2015, Kerremans et al. 2011, Würschinger et al., 2016) (cf. section 6). Hence, I assume that the use of

social media also influences the diffusion of new words. All in all, I hypothesise:

H10: Younger people, people in urban areas, highly educated people, and people who spend more time on the internet or use more social media platforms will know and use more words in contrast to older people, people in rural areas, less educated people, and people who spend little time on the internet and use a limited number of social media.

With these hypotheses in mind, I created and performed questionnaire studies that will be outlined in more detail in the empirical part of this thesis (cf. part II).

1.3 Outline

The thesis is divided into two major parts: a theoretical and an empirical part. The theoretical part starts off by providing a detailed discussion of the nature of neologisms, giving a definition of the term itself (section 2.1), locating them within the English lexicon (section 2.2), explaining how they can be distinguished from *nonce-formations* (section 2.3) as well as presenting the life circle of new words (section 2.4).

Chapter 3 outlines the establishment processes of new words. First, the social process of conventionalisation is explained in detail (section 3.1). Since it forms the fundament for this thesis, various aspects of conventionalisation are elaborated. It will be shown which catalysators propel conventionalisation (section 3.1.1) and which external forces impact it (section 3.1.2). Further, as conventionalisation in general does not only apply to words but all sorts of linguistic innovations, section 3.1.3 illustrates the subprocesses of conventionalisation specific to lexis: institutionalisation and lexicalisation. All the different aspects of conventionalisation are then summed up in section 3.1.4.

Since the establishment of words does not only call for social processes but also cognitive ones, the second part of chapter 3 sheds light on the entrenchment process (section 3.2). Although this thesis is mainly

based on social processes, it is important to understand the cognitive ones, as there is a strong connection and interplay between the two of them. Therefore, similarly to the first part of chapter 3, the second part provides information about the catalysators for entrenchment (section 3.2.1) as well as the forces affecting it (section 3.2.3). Additionally, the psychological foundations for entrenchment are outlined (section 3.2.2). Lastly, it is shown how entrenchment works when tailored to lexical items (section 3.2.4). As usage constitutes the connecting wheel between conventionalisation and entrenchment, it is elaborated upon in section 3.3. A description of various forces affecting usage is given, and factors that are of importance when it comes to the usage of new words are introduced.

As the empirical part (part II) is based on English blend neologisms and German compound neologism, it is important to lay a theoretical framework for the connection between neologisms, blends, and compound words. This includes defining what blends (section 4.1.1) and compounds are (section 4.2.1), introducing the different forms and structures they can take on (section 4.1.2; section 4.2.2), as well as explaining how blends and compounds are identified, accessed, and consequently processed by speakers (section 4.1.3; section 4.2.3). Before moving on to the empirical part, chapter 5 elucidates the connection between neologisms and the internet and how researching neologisms has changed with the increasing dominance of the internet in our daily lives. The chapter describes how the methods for retrieving, collecting, and documenting neologisms have changed over the last decades as well as the advantages and limitations of using online questionnaires, which were used as a tool to investigate neologisms in the empirical studies conducted.

The empirical part of the thesis is divided into four sections, which all follow a relatively similar pattern. Each section outlines one of the questionnaire studies that were implemented. Chapter 6 outlines the first study. In a first step the methods used are explained in more detail (section 6.1), complementing the more general discussion of methodology in section 5 for this particular survey. This includes how and according to which criteria the neologisms in question were retrieved (section 6.1.1), how the

questionnaire was designed and, once it was conducted, what the demographic distribution amongst the participants looked like (section 6.1.2). It is also elaborated which statistical analyses were used in evaluating the study (section 6.1.3). After having established the methodological groundwork, the quantitative results are presented (section 6.2). The results section starts off with the demographic outcomes (section 6.2.1), followed by the results regarding frequency, first occurrence and quality of media (section 6.2.2), the participants' interest (section 6.2.3) and the factors recognisability, transparency, and appeal (section 6.2.4). Based on the outcomes of section 6.2.4, a further analysis of the three aforementioned factors has been conducted and the results are presented in section 6.2.5. After the presentation of the quantitative results, the qualitative results are introduced (section 6.3). Chapter 6 concludes with a discussion about the conduction, analysis, and evaluation of the first questionnaire study and by identifying the research objects for the next study (section 6.4).

Chapter 7 follows a similar pattern. Based on the outcomes of the previous study, the newly assessed research questions are presented (section 7.1). This is followed by an overview of the methods used for this particular survey (section 7.2), including the criteria that formed the basis for choosing the neologisms (section 7.2.1), the design of the questionnaire and the participants' demographic distribution (section 7.2.2) and a short introduction into Confirmatory Factor Analysis (CFA), the method used to analyse the data (cf. section 7.2.3). The quantitative results for appeal are presented in a next step (section 7.3). This is complemented by the qualitative results (section 7.4). The overall outcomes are summarised in a discussion and their impact on the next study is described (section 7.5).

Succeeding that, chapter 8 introduces the third questionnaire study. This study was conducted on German rather than on English. The chapter starts off with the research object and questions (section 8.1). Since German neologisms had to be retrieved for this study, more details are given on the particular method used in this study (section 8.2), including the criteria for retrieving the German neologisms (section 8.2.1), the design of

the questionnaire and the demographic distribution amongst participants (section 8.2.2). Section 8.3 presents the quantitative results, starting with what has been gained from the demographic data (section 8.3.1), followed by the outcomes of the assessment of the role of the word's frequency (section 8.3.2) and appeal (section 8.3.3). The chapter is concluded by discussing the results (section 8.4).

Chapter 9 presents the last study, an explorative follow-up investigation which tries – on a very small scale - to put in practice what has been found in the three previous studies. The chapter opens with stating the research questions (section 9.1), followed by the introduction of the methods used (section 9.2), including the retrieval of materials (section 9.2.1) and the questionnaire design (section 9.2.2). Lastly, the results of this small project are presented (section 9.3) and discussed (section 9.4).

The thesis concludes with chapter 10, which gives an overall summary of what has been found, pointing out the most prominent and promising results (section 10.1), as well as an outlook towards future research that could be conducted, based on the results of the studies presented (section 10.2).

I. Theoretical background

2. Neologisms

2.1 What is a neologism?

Defining a neologism is quite a fuzzy business and it is even said that the terminological confusion is at its worst here (Hohenhaus, 2005: 363). What can be said with certainty is that the term itself derives from Greek *neo* (=new) and *logos* (=word) and made its way into English in 1772 via the French expression *nèologisme* denoting the “practice of innovation in language, the use of new words or old words in new sense” (Online Etymological Dictionary). Prior to this, *neological* was used in 1754, when Lord Chesterfield wanted “a genteel, *neological* dictionary, containing those polite [...] words and phrases, commonly used by the beau mode” (Clauzure, 2003: 208). Ever since, both lexicographers and linguists have undertaken attempts to define the term, partially with different outcomes.

Most dictionaries agree that a neologism is “a newly coined word or expression” (Soanes & Stevenson, 2008), “a new word, usage, or expression” (Merriam-Webster Online Dictionary) or “a new word or expression in a language, or a new meaning of an existing word or expression” (Collins English Dictionary). Lexicographers have the difficult task to decide which ones of the many newly coined terms are added to the dictionary. This is particularly challenging, as much more terms are invented than can be added to reference dictionaries. According to the *Concise Oxford English Dictionary (COED)*, 70% of new words do not make it into a dictionary (Soanes & Stevenson, 2008: 5). In order to come to a decision, three criteria are often taken into account: the frequency of the word over a longer period, its usefulness, and its uptake in the common, globally used English vocabulary. This entails that the word has been adapted to a wider range of different contexts (Kerremans, 2015: 28-29), and that “the general public was made aware of the word or sense” (Tulloch, 1993: V). Therefore,

to be considered for a dictionary entry, a word must have made a lasting contribution to the vocabulary and have an established foundation (Kerremans, 2015: 29).

Amongst linguistics there is a lack of sufficient definitions in literature at large, as the term neologism is seemingly often regarded as self-explanatory (Kerremans, 2015: 29). Many agree that “a new word is “new” if it (or a particular use of it) does not appear in general dictionaries [yet]” (Algeo, 1991: 2). This view resonated across languages, such as French where “il appert que la néologie (ou le néologisme) ne devient visible et palpable que dans l’orbite du dictionnaire” [it seems neology (or neologisms) only become visible and tangible orbiting around the dictionary] (Boulanger, 2010: 68), in Italian, where a neologism is “[una] parola o un’espressione nuova, non ancora registrata nei dizionari” [a word or a new expression, not yet recorded in the dictionaries] (Adamo & Della Valle, 2005: V) and Catalan “paraules [...], que no apareix en uns determinats diccionaris de referència” [words (...), that do not appear in certain reference dictionaries] (Observatori de Neologia, 1998: 9).

However, this criterion alone is not sufficient for defining neologisms. Cabré states that apart from the exclusion from dictionaries, other important parameters for identifying neologisms include the date of appearance in a reference dictionary, formal or semantic instability and the perception speakers have of an item’s novelty (cf. Cabré, 1993). Thus, like lexicographers, also linguists (across languages) regard novelty as one of the core criteria, as a neologism depicts “una palabra, una expresión pluriverbal [...] o un sentido nuevo que surge en una lengua determinada” [a word, an expression or a new meaning that arises in a certain language] (Moliner, 2013: 11) and refers to words that “sean procedentes de idiomas extranjeros, [...] sean nuevos términos ‘inventados’ ex novo o [...] y también palabras en novedosas acepciones” [are coming from foreign languages, are newly ‘invented’ terms ex novo, and also words with novel meanings] (Fernández Fernández, 2004: 9-10). It is “une unité nouvelle, de nature lexicale, dans un code linguistique défini” [a new unit, of a lexical nature, in a defined linguistic code] (Rey, 1976: 4). However, novelty is rather

subjective and in order to make it measurable, *new* has to be defined in connection with a certain time and given period (Plag, 2003: 52):

“[E]ls neologismes en tant que objectes de coneixement són unitats relatives que només es poden reconèixer si ens situem en un període de temps, situació discursiva o perspectiva enunciativa precisos”

[As objects of knowledge, neologisms are relative units that can only be identified when placed in a specific time period, discursive context and enunciative perspective] (Cabré, 2015: 133-134).

“Ein Neologismus ist eine lexikalische Einheit bzw. eine Bedeutung, die in einem bestimmten Abschnitt der Sprachentwicklung in einer Kommunikationsgemeinschaft aufkommt, sich ausbreitet, als sprachliche Norm allgemein akzeptiert und in diesem Entwicklungsabschnitt von der Mehrheit der Sprachnutzer über eine gewisse Zeit hin als neu empfunden wird”

[a neologism is a lexical unit or a meaning which comes up in a speech community in a specific stage in the development of a language, which spreads and gets generally accepted as language norm and which, in this particular time span, is perceived as new by the majority of speakers] (Herberg, 2004 et al.: XII).

Thus, these definitions show that novelty needs to be set in context (Cabré et al. 2014: 15). This is hardly surprising, since changes in lexis can mirror changes in the world and society (Algeo, 1991: 1). It is argued that one of the main driving forces for new words are new concepts that enter our lives (Behera & Mishra, 2013: 25). However, linking novelty to time is very subjective to individual speakers, and does not depict an objective temporal phenomenon (Fischer, 1998: 3). Some speakers might already be familiar with a term, while others will experience the same term as novel.

This shows that novelty and no dictionary entry are two criteria that need to be accompanied by others to define neologisms adequately. Hence, apart from time, also frequency plays a role in defining neologisms. In contrast to an established word, a neologism is usually defined as being still relatively infrequent and not subject to socio-pragmatic diffusion in different text types and semantic contexts (Fischer, 1998: 4, Schmid, 2008: 1-2; 2011: 75, 77; Hohenhaus, 1996: 19,29; 2006: 17; 2007: 17-18). Therefore, it can be summarised that neologisms are “form-meaning pairings (in one of the three possible combinations), i.e. lexical units, that have been

manifested in use [...], but have not yet occurred frequently and are not widespread enough in a given period to have become part and parcel of the lexicon of the speech community and the majority of its members” (Kerremans, 2015: 30). Therefore, a neologism is characterised by its novelty, the (relatively subjective) involvement of time, its infrequency, its low degree of socio-pragmatic diffusion and the lack of a dictionary entry.

2.2 Neologisms and the English lexicon

While the numerous definitions of neologisms in the previous section extend over various languages, the structure of the vocabulary differs between languages. The vocabulary of the English lexicon is a unique mixture of Germanic and Romance elements and therefore rich in nuance distinctions (Leisi, 1985: 68)², like the example of German *Wagen*, and its English translations *car*, *cart*, *carriage*, and *chariot*, shows (Lipka, 2002: 14). Despite the many nuances in English, a traditional, synchronic survey of the structure of the English vocabulary based on a diagram given in the introduction to the *Shorter Oxford English Dictionary* (SOED, 1973: x) shows that the English vocabulary contains a large central area. *Figure 1* displays a model in which this central area is indicated as COMMON. Most media, styles, and social classes are familiar with the vocabulary of this area (Lipka, 1992: 10). Hence, it contains words like *come*, *father*, *chair*, *good*, *bad*, and *very*.

While the editors of the SOED call this *Common English*, Quirk and Greenbaum denote this as the *common core* in their *University Grammar of English*, stating that it is not only present amongst the majority of speakers but also in all varieties of English (Little et al., 1973: x; Quirk & Greenbaum 1973 :1 ff.). Above the common area there is *literary*, containing the radially connected subcategories of *scientific*, *foreign*, and *archaic words* such as *Weltanschauung* or *blasé*. (Lipka, 1992: 10). In opposition to this stands the

² Ernst Leisi’s “Das heutige Englisch” (1985) gives a detailed picture of the structure of the English vocabulary. Leisi undertakes a synchronic description on a historical basis. Only a few aspects will be discussed in this thesis due to their relevance.

colloquial section, containing *dialectal* and *vulgar* elements as well as *slang* words and *technical* language.

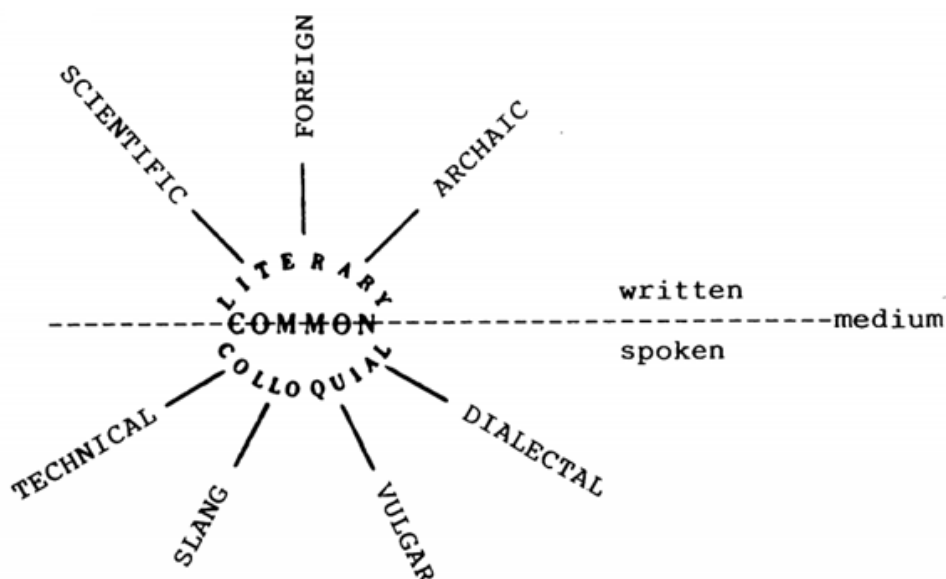


Figure 1 Modified version of the English vocabulary based on the Shorter Oxford English Dictionary (Lipka, 1992: 10)

As slang refers to special languages of specific groups (e.g., army slang, navy slang), it is difficult to distinguish it from technical language (Lipka, 1992: 10). Hence, there are no sharp boundaries between the categories, and the areas often merge and interpenetrate. The further away a word is from the common centre, the more peripheral and unknown the word will be. The radial lines of the model connect the peripheral and the central areas of the vocabulary (Lipka, 1992: 10). Leisi adds that, when starting from *archaic* and turning right in a circle until *foreign*, there is always an inherent connection between the labels. For instance, archaic words are often alive in dialects and dialectal, and vulgar expressions both belong to popular language and so on (Leisi, 1985: 187f.). Moreover, drawing a line through the common core depicts the division between written English above the line and spoken English below (Leisi, 1985: 187f.). Vocabulary is thus assigned in equal parts to the two media (Lipka, 1992: 11).

Neologisms can usually be found in the periphery of the model. They can occur in various peripheral categories, both in spoken and written language. Some neologisms are taken from scientific areas or loan words

from foreign languages, others develop in slang or dialect (Lipka, 1992). Eventually, only very few of them will make it into the common core of the English vocabulary (cf. section 2.4).

2.3 Neologisms versus nonce-formations

Alongside with neologism, the term *nonce-formation* is often used, sometimes even synonymously (Kerremans, 2015: 28). However, these terms need to be differentiated from each other and their terminological distinction goes all the way back to 1906, when Henry W. and Francis G. Fowler, the founders of the *Concise Oxford English Dictionary* (COED), dedicated a short chapter to it in their prescriptive style book *The King's English*. The book deals with the appropriate use of neologisms in writing and claims that, while neologisms are words that describe new things and ideas, *nonce words*, *nonce-phrase*, or *nonce-sense* (Fowler, 1954: 29) are words that “serve a need of the moment” (Fowler, 1954: 29). They are “[n]ew complex word[s] created by a speaker/writer on the spur of the moment to cover some immediate need” (Bauer, 1983: 45) and are unlikely to become permanent (Algeo, 1991: 3). They are actively formed by speakers and not retrieved from storage (Hohenhaus, 2005: 264; 2007: 17). When speakers scan their mental lexicon without finding a suitable lexeme for the required meaning (Fernández-Domínguez, 2010: 199), they might come up with a nonce-formation, which is per definition not lexicalised (Lipka, 1977: 162).

Part of the nature of nonce-formation is their contextuality, and thus they are ambiguous without context (Bauer, 1983: 45). Since they are invented in a certain situation, serving a certain need or function, they might not be understood under other circumstances (Hohenhaus, 2007: 19 ff.). An example is the term *apple juice seat* which refers to a chair in front of which a glass of apple juice has been placed. As the expression only relates to one deictic reference, without implying the existence of a noteworthy subcategory of ‘seat’, it is an example for a nonce-formation (Downing, 1977: 819, 823). Thus, in contrast to neologisms, which reoccur and eventually possibly diffuse, they tend to be one-offs that are not necessarily

meant to spread (Crystal, 2002: 132; Kerremans, 2005: 30). Therefore, it can be helpful to distinguish between *absolute* new words (nonce-formations) and *relatively* new words (neologisms) (Bauer, 2001: 38ff.).

The second a word is coined it can never be predicted whether it will 'only' be a nonce-formation or whether it will evolve into a neologism. Their perception is also very relative and subjective, depending on whether a speaker has encountered the word beforehand, and perceives it as a neologism or not (Hohenhaus, 2005: 365). Being attention seekers (Lipka, 2002: 147, 189), nonce-formations are often used in a rather informal style like the yellow press, advertisements and text types that want to achieve humour (Hohenhaus, 1996: 296-317). In order to catch the reader's eye, they try to be witty or stick out by breaking grammar rules like the German example of *unkaputtbar* (Crystal, 1998: 93 ff.). The German suffix *-bar* requires a transitive verb base. *Kaputt*, however, is an adjective (Hohenhaus 2005: 369). Thus, nonce-formations can serve a speaker's need in a certain situation but they can also occur in written language.

In sum, neologisms are not only new words, but words that have lost their nonce-formation status and are in the process of becoming, or already have become, part of the norm of the language (in the sense of Coseriu's notion of norm) (cf. chapter 3.1.3). They are still considered new by most members of the speech community (Schmid, 2008: 1). This implies that neologisms are located in a transitional state between not being a nonce-formation anymore, but also not being an established word yet. When a word is coined, its fate is unclear. It is possible that a word has been coined in the spur of a moment but then 'sticks around' and evolves from a nonce-formation to a neologism (Bauer, 2001:38 f.). This implies that "nonce-formations are somewhat 'in-between' actual words and possible words" (Hohenhaus, 2005: 364). They have a physical reality, but do not exist in such a way that they are part of the lexicon (Hohenhaus, 2005: 364). However, especially with the internet, the pure numerical amount of nonce-formations should neither be underestimated nor ignored. Nonce-formations may well be a much more important phenomenon of language

use than they are often given credit for (Herbst, 2010: 118) and it is important to distinguish them from neologisms.

2.4 The life of a neologism

Like living organisms, languages undergo maintenance, shift, decline, death, and revival (Edwards, 2009: 61). Neologisms can develop or enter a language either as the result of loan words, word-formation, or layering (Lipka, 2010: 97).³ While productive word-formation processes include compounding, affixation, blending, acronyms and conversion (Ayto, 1990), layering describes the process by which an established word develops a new meaning (Lipka, 2010: 97). Thus, word-formation and layering depict the “Unterschied zwischen neuen Semen oder nur neue Gebrauchsweise einer lexikalisch etablierten Einheit” [difference between a new sem or a new usage of an established unit] (Kinne, 1996: 345), “zwischen Neulexemen und Neubedeutungen” [between a new lexeme and a new meaning] (Herberg, 2002: 195). Algeo (1991: 3) expanded on this and identified six potential sources of new words:

1. **combining** such as compounding and affixation
2. **shift in meaning** (e.g. *web* - applies to 15% of the researched words)
3. **shortening words** (e.g. *bus* from *autobus*) or acronyms (e.g. *laser* from ‘Light Amplification by Stimulated Emission of Radiation’)
4. **blending** (only applies to 5% of researched words)
5. **borrowing** (e.g. *salsa*)
6. **creating** (e.g. *blurb* - least common group as completely newly created words are hard to understand without context)

³ Lipka also mentions a rather minor side-category, so-called *buzzwords*, which are “words or phrases from special areas of knowledge that people suddenly think very important” (Definition Longman Dictionary) like *sexting* (Lipka, 2010: 97)

Thus, neologisms are

“un novo significante ou un novo significado que aparece nunha lingua por procedementos formais internos (derivación, composición, sintagmación, abreviación) ou externos (empréstitos) ou por procedementos semánticos internos (ampliacións semánticas) ou externos (calcos semánticos)”

[a new signifier or a new signified appearing in a language by means of internal formal procedures (derivation, composition, syntagmation, abbreviation) or external ones (loanwords), or by means of internal semantic procedures (semantic extensions) or external ones (semantic calques)] (López Fernández, 2005: 8).

This means that novelty can either affect *signifiant* and *signifié* simultaneously or separately (Kerremans, 2015: 31-32). *Table 1* shows that the productive patterns for potential words can be subdivided in accordance with Saussure's linguistic sign. In a morphological neologism only the *signifiant* is affected, (e.g. *snownado*). This does not only apply to blends, but also to clippings and acronyms (Lipka, 1992: 92). In case of a change in the *signifié* and hence a semantic transfer, the neologisms are denoted as semantic neologisms. An example is *subprime*, which used to bear a positive connotation before the financial crisis. Lastly, if both - *signifiant* and *signifié* - are concerned, the neologism is of morpho-semantic nature such as *detweet*, a word that is completely new in form and meaning (Tournier, 1985: 47-50). New compounds also fall into this category (Lipka, 1992: 92). The external process of adopting loanwords remains outside of the three categories.

morpho-semantic neologism	construction:		internal matrices
	affixation	prefixation suffixation inverse derivation	
	composition	juxtaposition amalgam	
	phonological motivation	onomatopoeia and ideophonic elements	
semantic neologism	transfer of class	conversion	
	“métasémie”	metaphor metonymy	
morphological neologism	reduction of the <i>signifiant</i>	aphaeresis apocope acronyms	
loan words		morpho-semantic loan/ semantic loan and loan translation/ morphological loan	external matrices

Table 1 Morphological matrices according to Tournier, 1985: 51

For a new word to be formed, innovation is needed, and speakers have to get productive and creative.⁴ While change is observed within the system and can be depicted in graphs and models, innovation is an act of the speaker, who can influence the system. This makes a successful innovation a speaker-act that leads to a socially fuelled change in the language system

⁴ It should be noted that while productivity and creativity were regarded as denoting more or less the same thing until not too long ago, they are seen as two different processes these days (Hohenhaus, 2007: 16). Productivity allows speakers to produce an infinite large number of sentences accounted for by the rule of grammar, creativity is the speakers' ability to extend language in a motivated but unpredictable way, hence not, or at least not completely, rule governed (Lyons, 1977: 549; Hohenhaus, 2007: 16). The word *head-hunter*, for instance, in its literal meaning is a proof of productivity, whereas its metaphorical use is a sign of creativity (Lipka, 1992: 92). Although the distinction seems clear, productivity itself is still quite a fuzzy concept. Fully unrestricted productivity constitutes the exception rather than the norm, which makes restrictions to productivity a common occurrence (Hohenhaus, 2007: 16). At the same time, creativity ex nihilo is extremely rare which means that creative innovations tend to be at least rule related. This makes a clear-cut distinction between creativity and productivity even harder. Therefore, it is better to regard the difference between creativity and productivity as a cline (Hohenhaus, 2007: 16).

(Jones, 1993: 222; Croft, 2000: 60). Although we can see and observe innovation, we usually do not know whether it really depicts the beginning of a change in the system, as innovation is often not successful (Milroy, 1992: 165 ff.). The ideal life circle of a new word consists of five steps (cf. fig. 2) (Behera & Mishra, 2013: 26):

1. **Unstable**: very new or being used only by a small sub-culture
2. **Diffused**: having attained a noteworthy incidence of use, but not yet having gained pervasive acceptance
3. **Stable**: having gained recognisable, being in vogue, or perhaps, gaining lasting acceptance
4. **Dated**: the point where the word has ceased being novel, entered formal linguistic acceptance and even may have passed into becoming a cliché
5. **Passé**: when a neologism becomes so culturally dated that the use of it is avoided because its use is seen as a stigma, a sign of being out of step with the norms of a changed cultural tradition, perhaps, with the neologism dropping from the lexicon altogether

GRAPHICAL PRESENTATION OF NEOLOGISM LIFE CYCLE

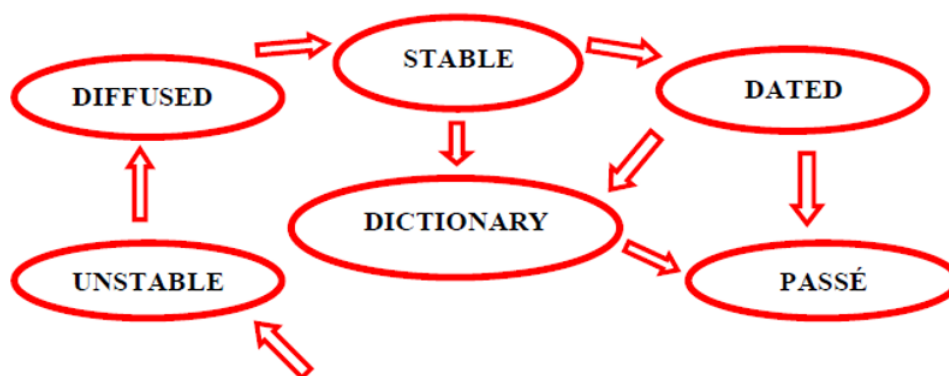


Figure 2 Life cycle of a neologism (Behera & Mischra, 2013: 27)

I would like to exemplify this life circle with the aid of the example *diskette*, a synonym for floppy disk (Advanced Oxford Learner's Dictionary). The word depicts a combination of the word *disk* and the diminutive suffix *-ette* (Fowler, 2015: 231). When computing developed more and more in the early 1970s, there was soon a need to describe that little disk on which data was saved. The word *diskette* was first attested in 1973 (Merriam-Webster

Online Dictionary). As the disk gained importance, the invented word was regarded as useful by most speakers (cf. section 2.1), which led to an increased endurance of the word – a factor that is regarded as an indicator to ensure long lasting success (cf. section 1.2) (Metcalf, 2002: 162). The analogy to the already existing *cassette* might have helped (cf. section 3.4). The transition from diffusion to usualisation and stabilisation goes hand in hand with the word moving from the periphery to the centre of the common vocabulary by means of institutionalisation and lexicalisation (cf. section 3.1.3). This is the moment in which a word might enter a dictionary. In the case of *diskette*, it has indeed been added and has been in there ever since. However, considering how our society has changed over the last 50 years, and how *diskettes* are not in use anymore, it is plausible that the word is at the end of its life circle and will soon be passé. It was found that “names for referents, once newsworthy, that have ceased to be topics of current discourse” (Algeo, 1993: 284) are unlikely to survive. Whereas *diskette* has been a rather successful word for a long time, most words do not get past the unstable stage.

This thesis will focus on the diffusion process within the life circle of new words (cf. section 1.2). While the distinction between a possible and non-possible word is a linguistic one, the distinction between an existing vs non-existing word depends on the ever-changing society (Kjellmer, 2000: 207). Hence, “a neologism stays new until people start to use it without thinking, or alternatively until it falls out of fashion, and they stop using it altogether. But there is never any way of telling which neologisms will stay and which will go” (Crystal, 1995: 132).

3. From nonce-formation to an established word: social, socio-pragmatic, structural, and cognitive processes

3.1 Conventionalisation

The previous section showed that there are several different factors that influence whether a new word survives or not. So far, researching innovations (cf. Labov, 1972, 1980, 1994, 2001; Trudgill, 2008; Brinton & Traugott, 2005; Clark, 1996) has mainly been restricted to rather easily observable phonological innovations (Kerremans, 2015: 60). The social process involved in the establishment of (new) words, namely conventionalisation, with its socio-cognitive and socio-pragmatic subprocesses, has been neglected (Kerremans, 2015: 60). While lexical innovations form the centrepiece of this thesis, conventionalisation can be applied to all sorts of utterance types and therefore it is necessary to briefly define the term *utterance types*, since it will be mentioned throughout the next few chapters.

“Utterance types are more or less conventionalized probabilistic connections between recurrent communicative goals and linguistic forms and between recurrent linguistic forms and recurrent communicative goals, which are contingent with regard to their immediate syntagmatic environment, and the wider linguistic as well as situational, social, and cultural context.” (Schmid, 2020: 23)

Neologisms are lexical innovations which, in order to be conventionalised, need to become accepted by the majority of members of a speech community as a new convention. Linguistic conventions can be described as collective habits that are required for successful communication (de Saussure, 1916: 100). Since the two sides of the linguistic sign are arbitrary and since the linguistic sign is unmotivated, such conventions are needed (de Saussure, 1916: 101–102). However, it can also be argued that conventionalisation is independent of motivation and arbitrariness and rather depicts a process that affects the degree to which the members of a community conform to regularities in use, irrespective of the motivation of the term (Keller, 1995: 156). Letting aside the impact of power and

normative forces, a linguistic convention can be defined as “a mutually known regularity of behaviour which the members of a community conform to because they mutually expect each other to conform to it” (Schmid, 2020: 88). For instance, while ‘half a year’ and ‘six months’ denote the same thing, corpus data show that the latter is more conventionalised amongst native speakers (Schmid, 2020: 88).

Normation can lead to a convention, with convention being defined as a regularity in behaviour (1) that is partly arbitrary (2) but counts as a common ground in a given community (3). It depicts a coordination device (4) for a recurrent coordination problem (5) (Clark, 1996: 71). To give an example, to solve the naming problem for ‘an object with a dark coloured bark which carries leaves’ (5), speakers agreed on the phonological and orthographical form of *tree* (3 & 4). There is no clear and evident reason as to why this particular string of sounds has been chosen (2) and an alternative expression would have been possible if the majority of a speech community had been conformed to it. Hence, *tree* is the convention to denote ‘a dark coloured bark which carries leaves’ (1) (Croft, 2000: 97). However, the level of conventionalisation is not set in stone but marks a dynamic process.

How dynamic conventionality is, can be seen when looking at the six different types of conformities that are involved, namely *onomasiological*, *semasiological*, *syntagmatic*, *cotextual*, *contextual* and *social/community conformities* (Schmid, 2020: 89 ff.). The degree of conventionalisation of an utterance type can be derived from how it behaves in regard to these six types of conformity. For the words investigated in this thesis, especially the former two are relevant since, in an idealised conformity profile, onomasiological and semasiological conformity are predominantly present in content words (cf. fig. 3). *Conventionality in terms of onomasiological conformity* refers to the conditional probability of a specific competing form being chosen to encode a given meaning or function. *Conventionality in terms of semasiological conformity* is the conditional probability of specific competing meanings or functions that are connected to a given form (Schmid, 2020: 89).

Despite their predominance in content words, also the four other types of conformity influence the conventionalisation of such content words. When putting them all together, it can be said that a lexical item is conventionalised to the extent that members of communities are in conformity (*community related conformity*) with respect to how to reach a communicative goal by using a linguistic form (*onomasiological conformity*), how to interpret this linguistic form (*semasiological conformity*), how to combine form and meaning (*syntagmatic conformity*) in a specific cotext (*cotextual conformity*) and context (*contextual conformity*). Thus, all six dimensions of conformity are involved in conventionalisation and their trade-off relation implies that changes on one dimension will have an impact on all others (Schmid, 2020: 200). This further means that conventionalised lexical items are not ready-made units, but co-semiotic potentialities competing for dominance in a multidimensional probability space (Schmid, 2020: 92). Because of this, conventions have to be seen as subjective and are in constant need of support from usage activities and ongoing conventionalisation processes to stay in place or adapt to changing needs (Paul 1920: 24–32, Schiffer, 1972: 148–149). The more a speaker's behaviour conforms to the current conventions, the smaller the contingency of a lexical item. Therefore, conventionalisation is not static but a process with various degrees, entailing that there is a spectrum from more to less conventionalised (Eckert, 2000:45). In sum, it can be said:

“Conventionalization is the continual process of establishing and readapting regularities of communicative behaviour among the members of a speech community, which is achieved by repeated usage activities in usage events and subject to the exigencies of the entrenchment processes taking place in the minds of speakers.” (Schmid, 2020: 2)

Given this dynamic, several catalysators and forces impact the level of conventionalisation.

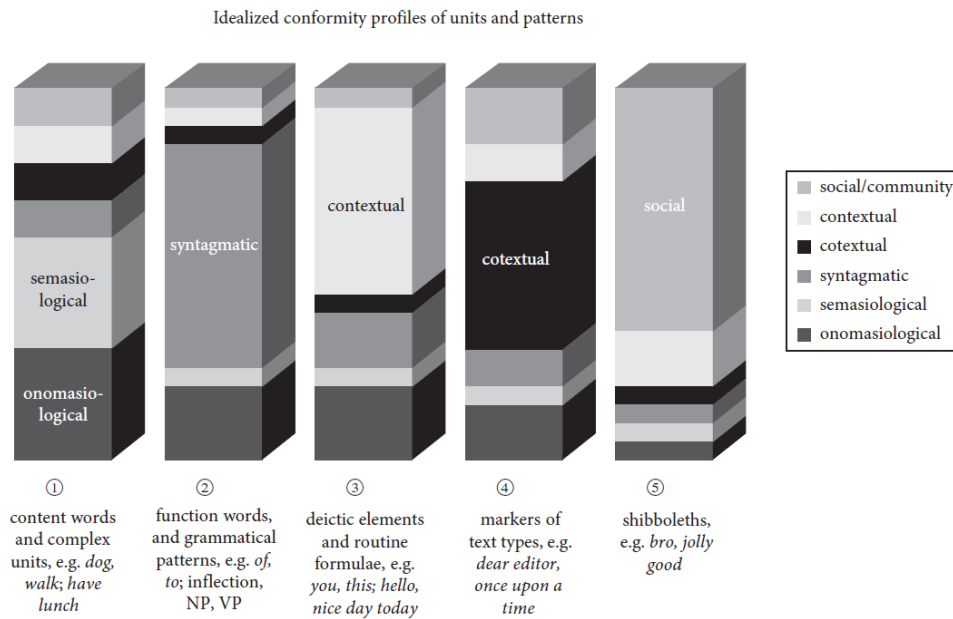


Figure 3 Idealised conformity profiles for five kinds of utterance types and their dominant conformity types (Schmid, 2020: 96)

3.1.1 Catalysators for conventionalisation

The social process of conventionalisation is embedded in a bigger network of interplaying factors (cf. fig. 4). The *Entrenchment-and-Conventionalization Model* (EC-Model), depicted in form of a simple Tinguely machine in figure 4, indicates how usage, conventionalisation and entrenchment are connected by two intertwining feedback cycles. The entrenchment and conventionalisation feedback loops are cause and effect for usage. The feedback loops can be put into motion by usage, at the same time they are further pushed and accelerated, in case of the conventionalisation feedback loop by usualisation and diffusion and in case of the entrenchment feedback loop by routinisation and schematisation. This can further push usage. An increased movement of the cycles also means that there is a repeated co-semiosis as well as a potential for licensing when it comes to the conventionalisation process. For entrenchment it implies that the patterns of associations are repeatedly activated, and the activation will be eased.

The model points out the nature of the conventionalisation feedback-loop, illustrating that the catalysators of diffusion and usualisation fuel the

conventionalisation process. Usualisation mainly affects and sustains the semasiological, onomasiological, and syntagmatic dimensions of conformity, diffusion the community related, cotextual and contextual dimensions. The two catalysators can be defined as follows:

“Usualization (Blank 2001, Paul 1920: 31–4) contributes to the effect that recurrent utterance types become established and are sustained and continually adapted as agreed-upon means of reaching recurrent communicative goals in specific contexts and can therefore function as tacit norms.” (Schmid, 2020: 5)

“Diffusion (Bailey et al. 1993, Labov 2003, Trudgill 1974a) causes usualised utterance types to become and remain conventional in specific contexts and in certain parts of a speech community.” (Schmid, 2020: 5-6)

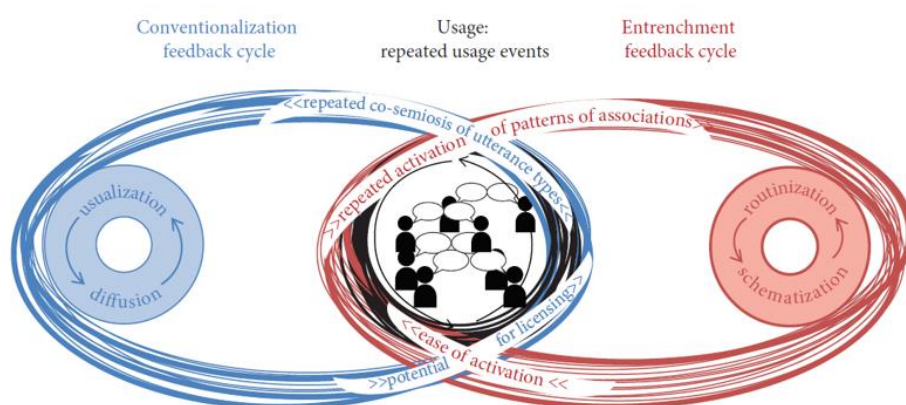


Figure 4 Schmid's language as a simple Tinguely machine illustrating the EC-model (Schmid, 2020: 4)

(Lexical) innovations can be anchored within the process of conventionalisation as they depict partly licensed utterances that take place in concrete language use (Paul, 1920: 32; Weinreich et al., 1968). These innovations can either happen by a deliberate or non-deliberate decision of the speaker (Schmid, 2020: 101) and thus can either be the result of a spur of the moment creation (cf. section 3.2) or more strategical planning, as it is the case in advertisements, newspaper headlines or even word contests (Metcalf, 2002: 47).

Innovations can also evolve through the hearer, who interprets a lexical item in a non-conventionalised way or through one of the participants for whom an item might not be conventionalised and lexicalised while it is for the other speakers (Schmid, 2020: 101). Due to the fact that all new

words are non-conventionalised upon their creation, even if they make use of conventionalised material (Croft, 2000: 98), it could be argued that innovation is on the opposite end of the spectrum from conventionalisation. The more a word gets propelled by usualisation and diffusion, the higher its chances to 'catch on' and thus to get conventionalised.

There are four different aspects involved in usualisation: *symbolisation*, *paradigmatisation*, *syntagmaticalisation* and *contextualisation*. *Symbolisation* is responsible for establishing, sustaining, and adapting onomasiological regularities, regarding the form that has been chosen to reach a certain communicative goal, and semasiological regularities, regarding the meanings and pragmatic functions connected with the forms. Symbolisation controls how the members of a community collectively begin to develop and sustain regularities concerning form-meaning and meaning-form relations. In case of lexical items, which combine forms and meanings, the effects of symbolisation are quite obvious (Schmid, 2020: 127). In relation to lexis, symbolisation corresponds to the process of lexicalisation, which deals with the meaning that has been established for a certain lexical item (cf. section 3.1.3). Early lexicalisation is marked by a relatively high degree of different spellings, pronunciation, and meanings. Speakers that are confronted with a new lexical innovation rely on word formation patterns, licensing the word, as well as on syntagmatic, context and contextual cues. In order to reach co-semiosis (cf. section 3.1.2), speakers often make use of inverted commas as a means to mark an innovation (Svanlund, 2018). Through reoccurring use, the word's spelling and pronunciation stabilises and its semasiological contingency decreases as a result of symbolisation. An example of the effect of increased symbolisation is the term *selfie*. Once it has reached a certain degree of usualisation, it was considered as unusual to say *she used her mobile to take a picture of herself* rather than *she took a selfie* (Schmid, 2020: 129).

However, not only symbolisation plays a role in the example of *selfie*. The decrease of semasiological contingency can also be attributed to *paradigmatisation* which establishes, sustains, and adapts regularities concerning onomasiological and semasiological oppositions between

utterance types/lexical items. It concerns the competitive aspects residing in the similarities and oppositions between lexical items. This process allows speakers to develop and adapt regularities regarding the ways onomasiological and semasiological contingency spaces are opened up. Paradigmatic relations in the lexicon emerge from this process (e.g., synonymy, antonymy, hyponymy, lexical sets and word-fields). As long as these relations are stable, the effect of paradigmatisation is hardly noticed. However, as soon as things begin to change, paradigmatisation becomes a major player. For example, when a competing loanword enters a language, the onomasiological space tends to be reorganised to keep oppositions intact (Schmid, 2020: 127 f.).

Syntagmaticalisation is about establishing, sustaining, and adapting regularities regarding linearisation. It corresponds to the structural principles of linearity and combination and brings about and sustains conformity among speakers regarding the serial order within patterns, and in combinations of units. It is of great importance for syntax and the lexicon-syntax interface and thus of less importance in the framework of this thesis. Lastly, *contextualisation* refers to establishing, sustaining, and adapting regularities in using utterance types/lexical items in certain cotexts (genres, registers) and situational contexts (Schmid, 2020: 128). Conventionalised lexical items must be regarded as records of their own rich usage history. These records come complete with contextual information regarding goals, typical users, and settings, as the process of contextualisation takes care of this. It controls the speakers' regularities of behaviour regarding correlations between the forms and meanings of the lexical item, on the one hand, and characteristics of genres and text types, types of situations, and types and groups of speakers on the other hand (Schmid, 2020: 128).

Hand in hand with a more stable form of the lexical item goes a decreasing need of cotext, context and word formation pattern to access the word. While innovations usually rely on cotext and context to be understood (Traugott & Dasher, 2002), usualisation allows new lexical items to emancipate. This can lead to the word becoming a direct carrier of co-semiosis rather than a generation of patterns (Lipka, 2002: 112). Syntagmaticalisation and contextualisation can also take place during this,

which makes speakers conform to collocations, complementation patterns and existing cotextual and contextual tendencies and restrictions (Schmid, 2020: 131). Eventually, a strong usuality and stable conformity profile will lead to a conserving effect. Frequent irregular forms that resist analogical levelling (e.g., *drank, ate, went*) are good examples of this (Schmid, 2020: 169). At the same time, it should be noted that frequent lexical items are more likely to be subject to semantic extension due to their larger range of cotexts and contexts (Schmid, 2020: 170).

The example of irregular verb forms as well as the possibility for semantic extension of frequent lexical items shows that the process of usualisation integrates normation. Although conventionalised lexical items function as norms in speech communities (Goldberg, 2019: 9-10), the norm is not simply a set agreement but is constantly negotiated (Rickford, 1993). Thus, even established norms are part of a tacit sense of mutually shared identity amongst the members of a speech community (of any size) (Schmid, 2020: 97). An innovation needs to undergo normation in order to be conventionalised. This can happen implicitly through the speakers (Blank, 2001: 1596), and explicitly in form of a dictionary entry (Holmes, 2008: 110 ff.). The more an innovation appears in metalinguistic discussions (Kerremans, 2015: 165 ff.) and the sooner it is listed in a dictionary, the higher its chances for the innovation's promotion and eventually its conventionalisation (Bauer, 1988: 246 ff.). Thus, normation allows an innovation to diffuse beyond a group to the whole society (Lewis, 1969: 99).

In sum, usualisation plays a key role in the establishment and adaption of structure on the macro-level of speech and controls the four corner stones of structure (linearity, meaning, opposition, context) (Schmid, 2020: 93). It has a great impact on the early stage of conventionalisation of an innovation. Usualisation is not the vehicle for the establishment of innovation but a motor for change (Paul, 1920: 32) and involves an alteration in the speaker's regularity of behaviour on all dimensions of the conformity profile. Its component processes of symbolisation, paradigmatisation, syntagmaticalisation, and pragmaticalisation become apparent in various phonological, morphological, semantic, lexical, morphosyntactic, and syntactic changes (cf. *selfie*). Thus, for innovations to

become conventionalised, usualisation is required. It depicts an abstract social process that affects the community at large and is integrated in a complex interplay with entrenchment (cf. fig. 4), as it partially depends on the cognitive process of routinisation (cf. section 3.2.1). The origin of usualisation is the activity of the speaker, which locates it in concrete usage events. Upon being repeated, the activity will become more and more routinised, implying that it can be produced more effortlessly, which leads to an increased likeliness to be performed. Higher usualisation involves the normation of the lexical item and an increased stability. This can also lead to predictions and more understanding amongst speakers (Berger & Luckmann, 1966: 53, 72, 73, 74).

The fact that new words are initially not conventionalised and only develop a norm after a certain degree of diffusion (Schmid, 2014; Schmid, 2020) makes diffusion the second catalysator for conventionalisation. It mainly affects the community related aspects of conventionalisation. There is a strong interaction with cotext and context conformity since communities are established by means of communication in concrete situation in specific activity types. Diffusion can affect different number of speakers, size and structure of communities that partake in cotext and context dependent conventionalisation. Thus, a lexical item might be more or less conventionalised, depending on how many speakers conform to a convention in bigger or lesser numbers of cotext and contexts. For instance, when looking at linguistics, terms such as *sentence* and *word* are widely diffused amongst most speakers, while *phoneme* and *clitic* are not so much diffused. Thus, diffusion of a term differs depending on the members within a community, the activity types as well as the types of situations (Schmid, 2020: 93).

Innovations, like neologisms, usually start off within a location, a community (politicians, artists), a discourse domain (sports, politics) and a discourse type (spoken, written). Successful neologisms often spread through different regional and social sections and conquer new discourse domains, taking the usual path from specialised to general. Computing gives several examples of new words that started off as jargon but are nowadays widely diffused and used by the general public (e.g., *google*,

scan, tweet) (Schmid, 2020: 94). Diffusion stabilises the cotext, context and community-related conformity profiles. It contributes to the spread of a lexical innovation across communities of speakers and the contexts of usage (Schmid, 2020: 179).

Three different types of diffusion can be distinguished: social, spatial, and stylistic diffusion. While social and spatial diffusion are community-related, stylistic diffusion is cotext and context related.⁵ Spatial diffusion is explained by the *wave* or *contagion* model, which can be traced back to Schmidt (1872: 27), who explained language change as the geographical spread of a feature on concentric rings, analogous to the waves caused when an object is dropped into water. Alternatively, the so-called *gravity* models (Hägerstrand 1952, Trudgill 1974b) or *urban hierarchical* or *cascade* models (Labov, 2003) of diffusion claim that innovations spread from large cities via smaller cities and towns to villages. The main predictors for these models are population density and distance.

Diffusion can also take place in the opposite direction, i.e. from rural areas to more urban ones. Amongst others, this has been shown by Trudgill (1986: 47–49). His findings have led to the formulation of *contra-hierarchical* models (Trudgill 1986: 47–49; Bailey et al. 1993). Since all of these models contradict each other, Bailey et al. (1993: 386) claim “that linguistic diffusion is far more complex than previous work might suggest” and that the patterns summarised above “probably do not exhaust the range of patterns involved in linguistic diffusion”. The authors conclude by stating that: “[a]lthough the diversity of patterns may seem surprising, they simply reflect the variety of demographic processes at work in a complex society and the complex motives people have for using the variety that they use” (Bailey et al. 1993: 386). One of the major issues with spatial diffusion models is that they do not take social factors into account interaction.

⁵ Even though the spread of neologisms into different co- and contexts is of major importance for their establishment (cf. section 2.1), stylistic diffusion is only remotely relevant for this thesis. Therefore, it should only be noted that it refers to changes in the way in which members of a community conform in regard to the use of utterance types in specific cotexts and contexts of use (Schmid, 2020: 192).

The S-curve model (Aitchison, 2001; Chambers, 2002: 361; Croft 2000: 183; Milroy & Milroy, 1985: 367) is commonly regarded as the normal distribution when it comes to social diffusion.⁶ Initially, an innovation is subject to a low diffusion as only innovators or early adopters embrace it (16% of people). After a while, the early majority (32%) takes on the innovation, followed by the late majority (32%), so that the slope flattens. Lastly, the innovation also spreads amongst the laggards (16%) (cf. fig. 5) (Rogers, 2003: 282-283). While this model shows the normal distribution of social diffusion⁷, the speed with which new words are adopted depends on the structure of the network. Especially networks with very dense ties often tend to resist innovation and maintain conservative forms (Milroy & Milroy, 1985: 355). Hence, closed groups need much longer to take on change, but once the change started it moves very quickly (Auer & Hinskens, 2005: 351). Often, an innovation starts off to diffuse slowly via weak ties and then enters the adopter phase in which the diffusion speeds up due to the strong ties (cf. fig. 6) (Croft, 2000: 4, 98). The S-curve stops once a lexical item is conventionalised. This, however, would imply that once conventionalisation has taken place, no change in state will occur. This does not agree with the fact that the conventionalisation process is in constant flux (Schmid, 2020).

⁶ For further reading: there is a vast number of studies taking into account various social factors and their impact on language change, such as class (Labov 1964; 1994), gender (Ochs 1992; Bucholtz 1999; Cameron & Kulick 2003), age (Sankoff & Blondeau, 2007), ethnicity (Cukor-Avila & Bailey 2001), and social network structure (Labov, 1972; Milroy & Milroy, 1985; Eckert, 2000; Paolillo, 2001).

⁷ It should be noted that while the comparison between the sociological and sociolinguistic S-curve models of diffusion stand to reason, Maybaum states that they differ substantially since one measures the diffusion through a *linguistic* system, while the other measures diffusion through a *social* system (Maybaum, 2013: 154).

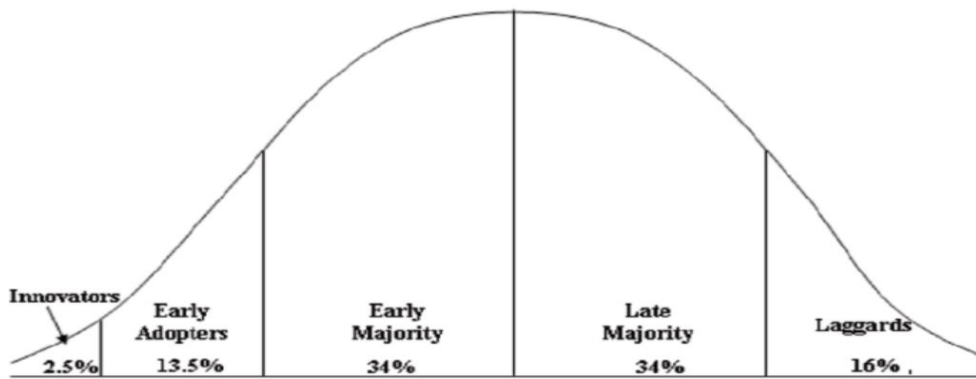


Figure 5 Adaptor categorisation on the basis of innovativeness by Rogers, 1995: 262

Another issue of the s-curve model is that it is not clear who leads the change. The fact that there is change from above (e.g. via external sources like borrowing) and below (transmission from generation to generation) (Labov, 1972: 123) as well as the fact that the leaders of change are in the centre for Labov, while in Milroy's definition they are peripheral members, shows the difficulties in grasping this parameter. Regarding social networks, it was found that usually people from higher socio-economic classes have more weak tied relations in larger numbers of communities in contrast to lower classes (Michael, 2014; Milroy & Milroy, 1992). However, in other studies, conducted in Reading and Milton Keynes, social class and networks show independent effects (Kerswill & Williams, 1999). Hence, we can only make abstract generalisations and postulate factors such as network structure, overt and covert/local prestige, power, ideology, solidarity, social meaning, social order, and so on (Kiesling 2011: 88–89). How these forces play out in specific cases and what other forces may play a role must be determined in detailed case-by-case analyses (Schmid, 2020: 191).

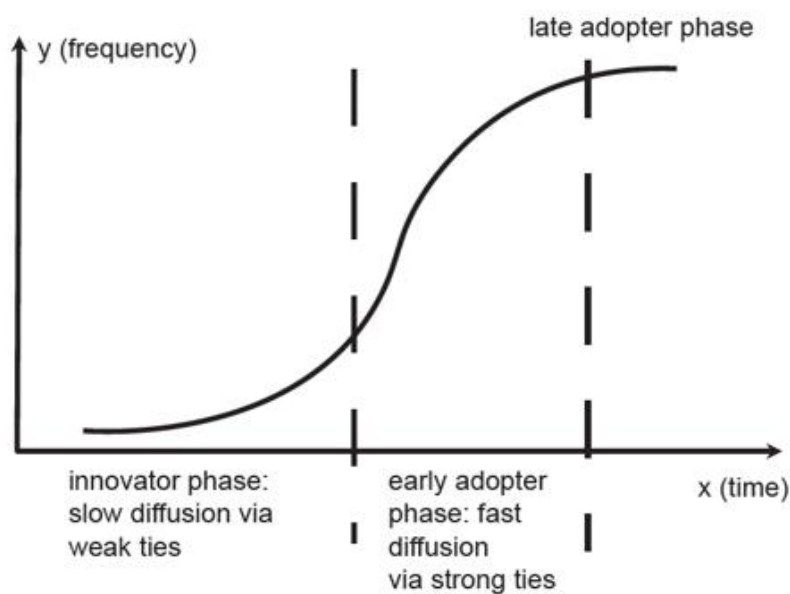


Figure 6 Spread of new lexical invention (Würschinger, 2019)

Looking at all the contradicting research in relation to spatial and social diffusion, it has to be concluded that it is hard to make generalisations about specific patterns of and effects on diffusion (Schmid, 2020: 191). All three types of diffusion interact and can influence each other, so that, for instance social and spatial patterns can change course when speakers from a certain peer group realise that items that are marking membership, have spread beyond their group and consequently abandon using these markers (Schmid, 2020: 195). Another example of the complex link between the different types of diffusion are Americanisms in British English. These do not only indicate a spatial spread, but are also driven by cultural factors, ideology, and power (Culpeper & Nevala, 2012: 380-381; Schneider 2011: 51-52).

While diffusion is in general an important catalysator for conventionalisation, it particularly promotes innovations, as they offer new choices and introduce a competition between an old and a new version, usually revealing only one winner (Schmid, 2020: 179). Although diffusion is constantly at work, stabilisation is usually more common than change due to the continuous self-reflexive dynamic homeostasis (Labov, 2001: 75,85). Change itself can take on different forms: it can be of a sporadic character, beginning and ending abruptly at times that cannot be predicted by any

universal principles. It can be a stable, long-term variation that persists over many centuries in much the same form which is perhaps even more common than changes which go to completion. Furthermore, it can also be retrograde, where the direction of change reverses or opposes the directions of movement in parallel communities (Labov 2001: 75).

One of the most important mechanism supporting diffusion – and usualisation - are *speech chains* (Agha, 2003) which indicate “the repetition of commonalities of utterance types in different situations” (Schmid, 2020: 95). The key components of this snowball effect are *co-semiosis* and *co-adaption* as well as *accommodation* (Glies et al. 1991; Trudgill, 1986, Labov, 1990, Schmid, 2020). *Co-semiosis* can be defined as “the activity of negotiating and establishing mutual beliefs of the mutual understanding of an utterance in a given context” (Schmid, 2020: 30). This can only be reached if a speaker believes that a hearer assumes that the speaker produces an utterance in order to be understood as well as to try to solve a communicative task by producing an utterance with a meaning that is intended by the speaker. Further, it implies that the hearer believes that the speaker assumes the hearer understands the utterance as an attempt to solve a communicative task (Schmid, 2020: 30). This means that co-semiosis is a joint attempt between speaker and hearer. If the participants are confronted with an original or innovative item, cotext and context are usually supportive in order to overcome communication issues, and to establish a successful co-semiosis, which then again can lead to future conventionalisation (Schmid, 2020: 31). This is of particular interest in terms of neologisms. While they are usually unlicensed utterance types, communication and co-semiosis can still be successful through cotext and context. Once co-semiosis with a new term was successful, it is possible that the neologism conventionalises further and, the more conventionalised it becomes for a specific co- and context, the higher its potential for being used.⁸

⁸ Co-semiosis mainly takes place in spoken, face-to-face interactions but it can also happen in asynchronous communication, albeit not quite to the same extend. Writers can use *recipient design*, *audience design* or *intersubjectivity* (Bell, 1984; Clark, 1996; Pickering & Garrot, 2004) to adapt to their readers. However, even then there is a lack of the same situational and cultural context and no

Co-adaptation implies the replication of features through speakers, which should, however, not be mistaken for a simple imitation but rather a mutual adaptation process (Schmid, 2020: 33). There are contradictory views about whether co-adaptive repetitions have an effect on memory and learning and whether it can have a long-lasting impact that goes beyond an explicit speech event (e.g., Du Bois, 2014; Pickering & Ferreira, 2008). Closely linked to co-adaptation is the concept of *accommodation* (cf. section 3.2.1).

Speakers often accommodate as an act of identity and by adapting certain dialects and accents as well as words, in-group connections are strengthened (Le Page & Tabouret-Keller, 1985: 181). When an innovative speaker meets a traditional one, the latter usually adapts to the former (Auer & Hinskens, 2005: 335). This can either lead to the traditional speaker adopting new features or abandon old ones. However, mainly the former leads to a social and geographical expansion, while interpersonal accommodation often does not have a lasting effect. Nevertheless, interpersonal accommodation is an important factor, being the root of any structural convergence or advergence (Auer & Hinskens, 2005: 335). Once the accommodating speaker starts using the innovation also in contexts without the 'model speaker' being present, the innovation becomes part of the speaker's individual speech habit (Auer & Hinskens, 2005: 336). For the word not to 'end up' being idiosyncratic and restricted to one user, which would prevent any noteworthy change, frequent use and exposure is needed. Thus, long-term change is usually encouraged if the innovator and his adopters are part of the same dense network group (Auer & Hinskens,

boundaries in regard to social factors, culture, distance or time. Besides, spoken language is spontaneous while written language is much more planned, making spoken utterances often incomplete and in need to rely on cotext, context and co-semiosis. In written language, the main aim is to conform to language norms and stick out through forces such as extravagance (Schmid, 2020: 76). Thus, some forms of written language can mainly be seen as an attempt to access a co-semiotic engagement rather than properly engaging in the joint activity of negotiating co-semiosis (Schmid, 2020: 32).

1996: 22; 2005: 336). Therefore, for a word to move from being an innovation to causing a long-lasting change by diffusion, it has to go through three phases (cf. fig. 7):

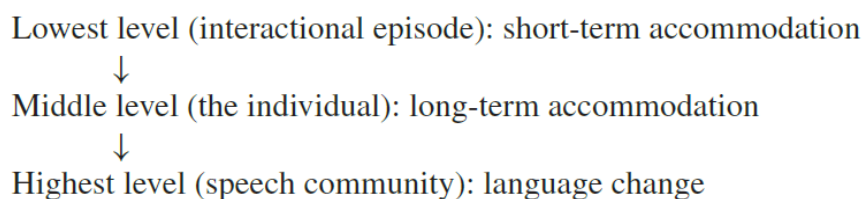


Figure 7 Three level model of accommodation and language change (Auer & Hinskens, 1996: 22)

Bell even argues that one can not only accommodate to another person but also to a third person or a listener that is further away in the sense of addressee > auditor > overhearer (Bell, 1984: 160, 163-167, 170-178), which would imply that accommodation cannot only take place between two speakers but even bystanders.

Labov calls change-through-accommodation the ‘principle of intimate diversification’ which describes “[e]ach act of communication between speakers [that] is accompanied by a transfer of linguistic influence that makes their speech patterns more alike” (Labov, 1990: 207). Trudgill states that “speakers accommodate to each other linguistically by reducing the dissimilarities between their speech patterns and adopting features from each other’s speech [...]. If a speaker accommodates frequently enough to a particular accent or dialect [...] then the accommodation may in time become permanent, particularly if attitudinal factors are favourable” (Trudgill, 1986: 39).

There is plenty of supporting empirical evidence for accommodation (mainly in the field of phonology), from Labov’s New York study (1966), via Trudgill’s Norwich study (1972) to Coupland’s Cardiff study (1984)⁹, just to name some of the most famous examples. While speakers use convergence as an attempt to reduce dissimilarities in social images (Coupland, 1984: 65), some experiments show that it is often mediated by

⁹ Despite all of these three studies being interesting contributions to the field of sociolinguistics in particular and to linguistics in general, they will not be explained in detail here as this would go beyond the scope of the thesis, especially since they all focus on phonological not on lexical changes.

stereotypes (Giles, Coupland & Coupland 1991: 16). It has been observed that some speakers do not accommodate to other speakers' real language use but rather to how they think the other person would stereotypically speak. Selting found that a German TV presenter accommodated to his callers by always imitating the same (stereotypical) dialect/accent rather than reacting to how the other person actually spoke (Selting, 1983: 42). Thus, in this case only an accommodation towards a stereotype persona took place (Auer & Hinskens, 2005: 343).

While all of the above research assumes that language change is socially fuelled, Keller regards change as an unintentional neutral evolution (Blythe & Croft, 2009: 49). His invisible hand theory "tries to explain structures and reveal processes, namely those structures which are produced by human beings who do not intend or even notice them, as if they were 'led by an invisible hand'" (Keller, 1994: 65). Accordingly, if several people adapt a certain new pattern, language change will happen. Keller illustrates this with beaten paths at his university. There are several shortcuts off the actual sidewalks. Only the fact that a lot of people use them, however, will eventually make them visible (Keller, 1994: 69 ff.). Thus, "speakers change their language neither intentionally, nor to a plan, nor consciously. This is generally true, and there is nothing more to it" (Keller, 1994: 13). This approach, however, still makes use of the ideal speaker and listener in Chomsky's sense. The truth lies probably somewhere in between, with language change being both, an intentional, social change as well as a subconscious process that eventually gets visible. Diffusion leads to levelling differences between speakers and communities. It is constrained by spatial and social boundaries as well as by other forces such as prestige and identity (Schmid, 2020: 199). The bigger collective is always needed to make change visible:

"Diffusion models portray society as a huge learning system where individuals are continually behaving and making decisions through time but not independently of one another. They watch one another and talk to one another about one another's behaviour and the experienced consequences [...]. Thus, the collective process involves some direct learning but mostly observational and symbolic learning" (Hamblin et al., 1979: 809)

Due to its relevance for this thesis, the role of mass media and diffusion should briefly be mentioned. Mass media seems to be a good tool for diffusion, as it reaches a lot of people. However, already Hägerstrand (1952) saw that one-directional communication renders co-adaptation as more or less impossible (Trudgill, 1974a: 223). It should be noted though, that it might work for lexical diffusion as the influence of American English on British English shows. Strang (1970) collected Americanisms in British English and claims that TV and radio have fuelled their occurrences. Regarding neologisms, diffusion-driven change is usually characterised by rapidity (as it is denoted today as 'going viral'). Thus, by using mass media, such as social media, a quick diffusion might be achieved, but it does not necessarily lead to usualisation and entrenchment in the same speed. Even if a neologism goes viral, not all people who read the new word will also partake in its usualisation. A lack of usualisation and hence also routinisation will lead to the word not becoming conventionalised and entrenched (Schmid, 2020: 314).

Table 2 shows the dominant forces that can drive change by diffusion (Schmid, 2020: 315). The forces include the speakers' attitudes and notions of prestige, solidarity and distance, identity, the spatial proximity and topology, the spatial and social mobility of speakers, high communication frequency and intensity, the structure and density of social networks and communities of practice, the salience of the new item, and the onomasiological and semasiological competition (cf. section 3.1.2). While diffusion is the leading process, it is triggered by innovation and mainly affects lexical, phrasal and semantic neologisms (cf. tbl 2). Side effects of the process are the institutionalisation and lexicalisation of the neologisms (cf. section 3.1.3).

Thus, diffusion can, like usualisation, trigger change. However, the two processes differ in regard to their contributions to change (Schmid, 2020: 201). While usualisation affects forms, meanings, and combinatorial options and constraints in cotext and context and is therefore mainly involved in various types of reanalysis, including semantic change,

syntagmaticalisation and chunking, grammaticalization, and pragmaticalisation, diffusion is conducive to change with regard to spatial, social, and stylistic aspects of conventionality, e.g. spread from urban centres to rural areas or vice versa, from higher and middle to lower social classes or vice versa, from colloquial to formal contexts or vice versa, or from technical genres to everyday life language or the other way around (Schmid, 2020: 201). This implies that usualisation affects form and meaning and is related to the dimension of onomasiological, semasiological, and syntagmatic conformity.

In contrast to this, diffusion affects the situational and community-related dimensions of conventionalisation. They overlap and meet in the situational dimension of conformity, where form and meaning are negotiated against the backdrop of the social character of the speakers and the interpersonal and social relations between them (Schmid, 2020: 179). Both processes are largely unpredictable, erratic, and subject to sudden interruptions or even reversals. They also rely on the same mechanisms: the repetition of commonalities of utterance types in different situations (Schmid, 2020: 201), hence *speech chains* (Agha, 2003). They depict two dependent, co-operating processes that share characteristics, use similar mechanisms, and influence each other. Both processes are continually at

	Trigger	Leading process	Utterance types mainly affected	Specific variants and side effects	Forces
Diffusion of complete novelty and salient innovation	Innovation: complete novelty or salient innovation; borrowing or creative coining or semantic innovation	Diffusion	Lexical, phrasal, and semantic neologisms	Institutionalisation and lexicalisation of neologisms and borrowed words and expressions; chunk diffusion	Mainly social forces: salience and extravagance of innovation; social order, network structures, prestige, power, social, and geographic mobility, media, language contact and multilingualism

Table 2 Excerpt from Schmid's Survey of modules of change - for full detail see Schmid, 2020: 330-331

work as part of the ongoing conventionalisation feedback loop, which not only drives change but is also required to sustain persistence (Schmid, 2020: 199).

Furthermore, diffusion can have an effect on usuality, as the example of the word *gay* shows. In this case an usualised meaning ('bright, lively-looking' and 'light-hearted, carefree' (*OED3* s.v. *gay*)) has changed from a new meaning ('homosexual') has diffused. This impacted the meaning in the wider speech community (Schmid, 2020: 95). Therefore, diffusion is often a trigger or motor for reorganisation driven by usualisation (Schmid, 2020: 198). However, it is also possible that innovations are diffused but not usualised. For instance, while in jargon language words might be widely usualised within a certain group of speakers – as in the case of linguistics terms such as *morpheme*, *phoneme* etc. - the same words might not be generally diffused (Schmid, 2020: 95).

Nevertheless, the study of diffusion processes must integrate the interaction between diffusion and usualisation. Changes affecting diffusion are likely to affect not only the cotextual, contextual, and community-related components of the conformity profiles, but also the onomasiological, semasiological, and syntagmatic ones. Hence, utterance types cannot only change their contextual features and social meanings under the influence of diffusion, but also their propositional meanings and forms (Schmid, 2020: 198). Diffusion brings collective and individual change and is likely to take the lead in change triggered by novelty and salient innovations rather than repetition driven change, since speakers are, per definition, unfamiliar with completely new utterances and are therefore unable to usualise and entrench them before they encounter them for the first time. While innovations usually originate in a concrete usage event (cf. section 3.3), they have to spread geographically, socially and situationally in order to become fully conventionalised, which happens by means of diffusion. The more speakers are exposed to an innovation in different places, situations, groups, social classes, cotexts, and contexts, the more it allows speakers to usualise it, establishing a regularity on all dimensions of conformity

such as its concrete forms, meanings, and combinatorial possibilities and restrictions (Schmid, 2020: 313). The processes are closely linked, and work together, but nonetheless must be regarded as two different entities.

3.1.2 Forces affecting conventionalisation

The previous section showed that usualisation and diffusion function as catalysators for conventionalisation. In order to bring the usualisation and diffusion wheel into swing, several (outside) forces impact them and thus influence the process of conventionalisation. While some of these forces have been touched upon in the previous section, due to their influence on usualisation and diffusion, they will be elaborated in the following.¹⁰

With conventionalisation being a social process, the behaviour of the involved speakers proves to be an important force. Co-semiosis, at least when repeated, can have a positive effect on usualisation. Similarly, repeated co-adaption, albeit not being an absolutely necessary force, facilitates usualisation of lexical items (Schmid, 2020: 104). Interpersonal activities, such as co-semiosis and co-adaption can be affected by different forces, such as *distance versus solidarity*, *power*, *status*, *extravagance*, *subjectivity* and *intersubjectivity*. As mentioned, network structures have an impact on conventionalisation (cf. section 3.1.1). Communication accommodation theory (Giles et al. 1991, Giles & Ogay, 2007) shows that, depending on how much solidarity the speakers feel towards each other, they converge or diverge more from their interlocutor. The felt solidarity can motivate interlocutors either to try to intentionally reduce or increase the distance between themselves. While closeness amongst speakers usually leads to a higher degree of usualisation, since they are keen to create solidarity, distance amongst speakers, such as when the speakers are from different communities, will increase the diffusion of more words due to a lack of co-adaption (Schmid, 2020: 105).

¹⁰ Only the relevant ones for this thesis shall be outlined, for a more elaborate overview compare Schmid, 2020: 104 ff.

Solidarity and the idea of an in-group feeling is linked to *power*. While there is a high degree of usualisation between the members of a powerful elite in regard to the markers they use, aiming at establishing and sustaining the language, it is not in their interest to diffuse their markers into other groups of society. Hence, while usualisation is high, diffusion is attempted to be kept low (Schmid, 2020: 105). This shows that the identity of a speaker, the social structure and order as well as the network have a great impact on conventionalisation. In the early stages, Labov and Trudgill mainly relied on static models of society, focusing on the socio-economic strata as well as social categories such as gender, age, and ethnicity (Schmid, 2020: 112).

The *2nd wave of variation studies* (Eckert, 2012) came with a change towards a more flexible and local understanding of social order in terms of social networks, inspired by insights from ethnography (Schmid, 2020: 112). This approach allowed more detailed descriptions of networks, regarding size, density, distribution, strength of connection between members and therefore enabled socially richer and much more dynamic explanations for variation and change. This second wave also yielded the term *community of practice* (Lave & Wenger, 1991). It is defined as

“an aggregate of people who, united by a common enterprise, develop and share ways of doing things, ways of talking, beliefs, and values—in short, practices. A [community of practice] can develop out of a formally or informally constituted enterprise: a choir, a gang, a secretarial pool, a family, a garage band, a friendship group, or an academic department.”

(Eckert & McConnell-Ginet, 1992: 186)

Developing shared ways of talking implies that, within these communities of practice, a specific form of usualisation takes place. Within the community and social network, central members have a stronger effect on the persistence of community-specific norms. In contrast to that, members in the periphery of the network are more likely to bring in outside influences, such as innovations, which contribute to diffusion. Thus, social order can facilitate, but also hinder usualisation and

diffusion. While members of a network or community of practice will stick to their internal, distinct forms of usualisation, communication outside of the boundaries of this network leads to an increased diffusion (Schmid, 2020: 113).

The fact that central and peripheral members of a community have a different impact on usualisation and diffusion shows that a person's or group's place in the social order, and hence their *status*, matters. Closely linked to status is the *prestige* of the speaker, which will be obtained by a member who achieves a status in a domain that is valued (Henrich & Gil-White, 2001: 167). Prestige, the positive interpretation of an individual, is also linked to influence. Vice versa, a member of the community who lacks prestige and influence might be *stigmatised* (Schmid, 2020: 114). In order to avoid this, speakers often adapt to and copy from a prestigious person. Especially social aspirers, who are placed in the middle of the social hierarchy, try to move up the social strata and are therefore more inclined to accommodate to a prestigious speaker and are more sensitive towards their innovations (Schmid, 2020: 115). In this context, the acts of self-presentation and positioning play an important role.

A particularly noticeable mechanism within these acts is *subjectivity*, which refers to the speakers' expression of their attitudes and beliefs (Lyons, 1982: 102). In contrast to this, *intersubjectivity* refers to the way in which speakers express "their awareness of the addressee's attitudes and beliefs" (Traugott, 2010: 33) and is therefore represented in the interpersonal foundation of co-semiosis and co-adaptation, implying that illocutionary and perlocutionary acts are inherently interpersonal activities, which are directed at addressees by taking their attitudes and beliefs into account (Schmid, 2020: 107). By doing so, the speakers can influence the way they are perceived.

Many studies give indications of how the *innovator's/coiner's prestige* impacts other speakers (Labov, 1980; Milroy, 1992, Kerremans, 2015). While Labov (1980) assumes that the most prestigious speaker, with the highest social status, most contacts within as well as outside their group (or neighbourhood), leads changes (Labov, 1980: 261), Milroy

states that the innovator has weak ties with several circles and is a marginal individual, whereas the early adaptor is central to a group and norm confirmatory (Milroy, 1992: 175ff.). Milroy's model is less based on prestige but on the strength of the network.

Although Milroy's and Labov's innovator creates changes on the phonological level, similar mechanisms can be seen in regard to lexical innovation, caused by the *coiner* (Kerremans, 2015: 21). In both cases, listeners might start adapting to a change and spreading it through accommodation (cf. section 3.1.1). Therefore, the coiner's authority and prominence could push the conventionalisation of a new word (cf. section 3.1) (Kerremans, 2015: 21). For instance, when Sarah Palin accidentally said *mandation*, instead of *mandate*, the word quickly diffused and became a topic in the media. It did not, however, diffuse due to its newness or due to describing a new idea or innovation but simply because a famous person, who is under constant scrutiny, uttered it (Kerremans, 2015: 150). Another example of the influence of the coiner or first user is *triphibian*, which seemingly became popular after it was used by Winston Churchill (Adams, 1973: 2). Furthermore, it is said that we can find early usage evidence for more than 1000 words we use today in Shakespeare's oeuvre. While it is argued that he probably did not invent all these words himself, his status helped in diffusing them successfully (Metcalf, 2002: 61).¹¹

While prestige is an important force for usualisation and diffusion (Lutz, 2013), the term is not clear-cut and rather subjective and relative, depending on the network, social group or community of practice (Schmid, 2020: 114). The differentiation between overt and covert prestige (Labov, 1966: 108; Trudgill, 1972) points out the fact that the perception of prestigious ways of speaking are distinct amongst different social circles and networks (Schmid, 2020: 114). Not only the status of a speaker within certain groups and networks can trigger conventionalisation, but also the mobility and outreach into other

¹¹ Other linguistic processes, such as grammaticalization, are also assumed to be favoured by the social influence of the founder(s) or early user groups (Haspelmath, 1999: 1057).

communities. Highly mobile speakers, both socially as well as geographically, increase the opportunities for co-adaptation, diffusion, and eventually usualisation by being in touch with members of different groups (Britain, 2010: 208, 210).

Whether or not co-semiosis and co-adaption takes place can also depend on *extravagance*. Extravagance as a term is also widely used in other linguistic areas. In grammar, it is assumed that speakers usually choose confirmatory grammar to be understood. While it is not clear in which situations extravagant innovations are made, it can be said that the opposite (anti-extravagant innovations) does not exist (Haspelmath, 1999: 1043). Once an extravagant grammatical construction becomes more frequently used, it becomes more predictable and less consciousness is needed to decode it and it will lose its extravagance (Haspelmath, 1999: 1057-1058). In contrast to lexis, extravagance within grammaticalization usually replaces existing concepts, while neologisms often denote something completely new. The field of semantics also makes use of the term. However, both in semantics and grammar, extravagance does not include components such as punning, which is essential for my definition (Traugott, 2001: 14). The area of morphology also makes use of the term *extravagance*. It is argued that extravagance is a trigger for language variation and change, even beyond creativity, as an integral part of language change (Eitelmann & Haumann, 2019). Examples of extravagance in morphology are the use of the suffix *-ish* in *everybody-ish*, *eight-ish* as well as reduplications like *a lot*, *a lot* (cf. Oltra-Massuet, 2017, Traugott & Trousdale 2013). Diachronic linguistics also investigate extravagance in grammar. Petré, for instance, has looked at grammatical changes from a historical point of view, namely the grammaticalization of the progressive [BE *Ving*] in present tense main clauses in comparison to SIMPLE PRESENT in the 17th century. He found that “[the] opportunity consisted of the transfer of [be *Ving*] from its preferred past-tense niche to present-tense main clause uses” (Petré. 2017: 233). Thus, the main clause use stood out in contrast to the past-tense equivalent, due to its novelty as well as its added quality of being

more isomorphic. Thus, the use of [be *Ving*] opened up an opportunity for individual speakers to be extravagant (Petré, 2017: 232-233).

Extravagant utterances are salient because of their lack of conformity (Schmid, 2020: 106). Since this does usually apply to innovations, such as neologisms, it is of particular interest for this thesis. However, it should be noted that amongst innovations there is a gradual scale from salient to non-salient innovations, which can take place on all levels of conformities. To give some examples, on an onomasiological level, a creative blend like *mansplaining* is more unconventional and salient than new products of regular word-formation processes, such as the compound *fake news*. Syntagmatically, an innovation can be salient in highly unconventional collocations like *commit social science* (Schmid, 2020: 79). It is also possible to use conventional forms in unconventional genres, registers, and situations, for instance, when the form *LOL* was transferred from e-communication to spoken language. Thus, this can be placed on the cotextual and contextual dimensions of conformity (Schmid, 2020: 103). Innovations regarding community-related conformity vary in terms of salience, depending on the distance between the donor and the recipient community. The use of terms associated with the repertoire of social elites by a member of a youth gang in a peer communication setting is probably more salient than the use of an expression associated with American English in a situation involving speakers of British English (Schmid, 2020: 103-104).

Although not all innovations are equally salient, extravagance can increase co-adaptation and prompt speakers to repeat these extravagant innovations in other situations, following the speech-chain mechanism which accordingly has a positive impact on conventionalisation. However, extravagance alone does not cut it. Other factors - such as whether speakers sympathise with the social significance of the extravagant innovation - are also important. An effect from an extravagant innovation being repeated in other situations is an increased usualisation and diffusion and thus an automatic decrease in extravagance (Schmid, 2020: 106). Thus, a once extravagant innovation cannot attract attention anymore and the speakers are forced to find a

new way to be extravagant, in the so-called *renewal* process (Hopper & Traugott, 2003: 121-123). This process is also relevant when it comes to the use of extravagant in-group markers, such as in youth language, slang, and jargon. Once this marker spreads across the wider speech communities, the members of the group will be less inclined to use it and, in order to find a different way to express their affiliation, *renewal* will take place. Thus, within in-groups, extravagant expressions or innovations experience a high usualisation but a low diffusion. Once they do diffuse though, they will be used less and less by the members of that group (Schmid, 2020: 106).

Since conventionalisation is a social process, taking place in the community, it is not surprising that all forces pointed out so far have been predominantly of a social nature, concerning the status, prestige, and mobility of the single members as well as the general fabric of social networks, groups, or communities of practice. However, social processes are not completely independent of what happens in the individual speaker's mind, and thus cognitive forces can also trigger and enhance conventionalisation. Usualisation can only happen if routinisation takes place (cf. section 3.1.1) (Schmid, 2020: 121). A speaker who wants to conform to the conventions and make use of a conventionalised utterance type needs to have a representation of this in his mind.

Furthermore, usualisation and entrenchment are also connected in so far that an entrenched lexical item is more likely to be activated, thus will be used more and therefore stands a better chance of being usualised. In contrast to this, less entrenched items will be harder to be activated, they will be used less and are less likely to usualise. This connection between entrenchment and conventionalisation is one example of where the conventionalisation feedback cycle and the entrenchment feedback cycle interact via repeated usage (cf. fig. 4) (Schmid, 2020: 121). This goes along with the fact that higher frequency¹²

¹² Frequency will be of great importance for this thesis. Thus, it should be mentioned briefly that it has been ignored by linguists for a great part of the 20th century. The assumption, that high-frequency words have other properties than low-frequency words was widely agreed on.

in repetition can promote conventionalisation. It stands to reason to assume that the more frequently a word is repeated, the more it becomes usualised, diffused and potentially also stable (Schmid, 2020: 122). It seems that for a word to become conventionalised, the word's frequency in general as well as in different discourses – both linguistically as well as metalinguistically – is essential (Kerremans, 2015: 115 ff.). In Metcalf's FUDGE scale (cf. section 1.2) 'D' refers to the diversity of users and situations, implying that a word which can only be used in a very specific situation might not survive (e.g. *amuse-bouche* – 'a single, bite-sized appetizer (hors d'œuvre)') (Metcalf, 2002: 158).

However, it is not only about the different contexts the word can be used in, but also its topicality. The stage of topical/transitional conventionalisation, which, for example, the neologisms *Burqini* experienced in France in 2016, is shaped by high frequency. From then onwards the word can either vanish or reoccur depending on real-life occasions. Hence, a higher nameworthiness conditions conventionalisation positively (Kerremans, 2015: 122 ff.). However, some terms just reoccur seasonally, if their topicality episodes are linked to extralinguistic events, which makes these words semi-conventionalised. *Cherpumple*, for instance, found a second peak in the next winter holiday season after its first creation (Kerremans, 2015: 129 ff.). In this context, it is sometimes argued that constant changes in our society constantly cause gaps in our language:

“Every year that passes throws up new ideas, experiences, and inventions for which no name has hitherto existed, and since names are indispensable cogs

However, the fact that frequency deals with the word level made linguists avoid it, as the linguistic trend in the past was focusing on structural changes, such as sound changes (Bybee, 2007: 5.). Chomsky and his followers completely ruled out frequency and repetition which becomes apparent in Chomsky's reaction to Skinner's verbal behaviour essay (Bybee, 2007: 6). The more recent interest in frequency (Bybee, 2007: 6) was started off by Greenberg, who investigated frequency in deriving the effects of underlying markedness and found that unmarked members of categories throughout the grammar are more frequent than marked members (Greenberg, 1966). The use of corpora was another important step forward (Bybee, 2007: 7).

in the machinery of communication, our natural human propensity for coining them soon plugs most gaps” (Ayto 2007: 1).

It seems, that “[n]ative speakers [...] have a mania for trying to fill lexical gaps. If a word does not exist to express a concept, there is no shortage of people very ready to invent one” (Crystal, 1995: 133). However, whether a word for a gap succeeds depends on how nameworthy it is to the majority of speakers. Well-established words are usually perceived as extremely nameworthy (Kerremans, 2015: 170 ff.). At the same time, innovations like *bagger* ‘person who bags your items in the supermarket’ or *cohab* ‘your partner you live together with, but you are not married’ (Kjellmer, 2000: 223) never took off, as they have seemingly not been regarded as nameworthy by society and thus have never conventionalised. At the same time, there are also gaps that simply never get filled even when trying, so that, for instance, a term for the 2000s has never been found (Metcalf, 2002: 74). Hence, an existing gap does not guarantee a word’s success. While new inventions often encourage creativity, vice versa productivity is restricted by the non-existence of things (Lipka, 1977: 161). It was found that “terms proposed for referents that did not come into existence” (Algeo, 1993: 284) such as *Airplane Conference* or *Airplane League* (a proposed football conference) struggled to survive. The same applies to very rare referents like *dartchery* (a sports event that combines archery and dart throwing) (Algeo, 1993: 284). Often “[t]he most salient type of neologism is a word which is new in its form and which refers to a concept which is new” (Mair, 2006: 38).

Although it can be argued that very advanced or complete conventionalisation is reached when a word achieves high frequency and comes up in different text types (Kerremans, 2015:136), frequency is much more complex than this and can cause contradiction and inconsistent effects. On the one hand, a high token frequency of content words such as *horse*, *book*, or *street* have proven to be quite stable. On the other hand, frequency of words was found to correlate with their polysemy (Zipf, 1949), implying that high frequency can also be the

engine for semantic change, such as in the case of the intensifier *fucking* (Schmid, 2020: 122).¹³

In sum, there are many different forces that affect conventionalisation. Most of them are of a social nature, such as distance and solidarity, power, extravagance, the speaker's identity, and their role in a network. However, the cognitive process of routinisation, as part of the entrenchment feedback loop, also influences and impacts conventionalisation. Linked to this is the frequency with which an item is used, which can depend on its nameworthiness (Kerremens, 2015: 148). Thus, while diffusion and usualisation are accelerating the conventionalisation process, many different forces influence whether a new word eventually gets conventionalised. Hence, conventionalisation is a gradual socio-historic and socio-cognitive process, accompanied by normation, by which innovative structures diffuse into the speech community on the basis of, for instance, accommodation theories and strategies.

3.1.3 Conventionalisation of words: Institutionalisation and Lexicalisation as subprocesses of Conventionalisation

The catalysators and forces promoting conventionalisation that have been introduced so far, do not only apply to lexical items but can affect and change the conventionalisation of all sorts of utterance types. Therefore, institutionalisation and lexicalisation should be mentioned separately as two subprocesses of conventionalisation that influence the establishment of (new) words in particular. Over the years, different

¹³ A similar observation can be made regarding grammar. While high token frequency exercises a preserving power with frequent irregular verb forms (e.g., *ate, went, kept*) and nouns (e.g., *women, men, feet*) resisting regularisation through analogical levelling (Bybee, 1985: 117–18; 2010: 24–32, Diessel 2007: 92), it is, on the other hand, also assumed that high frequency facilitates grammaticalization processes such as reduction, fusion, or coalescence (Narrog & Heine, 2011). At the same time, many complex prepositions (e.g., *by dint of or in conformity with*), that underwent grammaticalization, exhibit low frequency (Hoffmann, 2004).

scholars have used various approaches for these two terms¹⁴, since “both lexicalization and institutionalization are global notational terms, which may be further subcategorized [and] both notions must be made more precise in analysis” (Lipka et al., 2004: 12). They have been considered as two independent and coexisting processes (Lipka et al., 2004: 11; Blank, 2001: 1579-1599) as well as being successive, following the pattern:

nonce-formation > institutionalisation > lexicalisation

(Bauer, 1983: 45ff.)

Both of these ideas, however, do not take into account the dynamic nature of the processes. When anchoring them within the EC-model, institutionalisation concerns diffusion, since it deals with the establishment of new lexical items amongst a wider range of speakers, whereas lexicalisation can be regarded as a subprocess of usualisation (cf. section 3.1.1 *symbolisation*). They are two simultaneous, linked processes that both have an impact on a different conventionalisation catalysator. The following will show that due to the interplay between entrenchment and conventionalisation, lexicalisation and institutionalisation are respectively also subject to cognitive factors.

When speakers encounter a word for the first time, they might be able to draw conclusions on its composition and semantics by referring to familiar word-formation pattern and processes (Kerremans, 2015: 37). At this stage, the word is *type-familiar*. Once it is known and used more widely in society, the word reached the stage of being *item-familiar* (Schmid, 2011: 74). Item-familiarity means that the word’s ambiguity is

¹⁴ Due to the lack of consensus in literature, some approaches avoid the terms lexicalisation and institutionalisation altogether and try to find other denotations to grasp the concept. Leech, for example, suggests *petrification* as a term to express the ‘shrinkage’ of denotation” (Leech, 1974: 226). Another term that has been brought forward is *fossilisation* (Lyons, 1977: 547). Quirk et al. argue for following the succession: sentence/paraphrase > nominalisation > word, which entails that lexicalisation labels the transition from what previously could only be expressed in sentences or phrases and can now be expressed in a single lexical item (Quirk et al., 1985: 1526).

ignored and, by using only one possible form, the word becomes a known lexical item (Quirk et al., 1985: 1522ff.). Therefore, the progress of institutionalisation begins once “the nonce-formation starts to be accepted by other speakers as a known lexical item” (Bauer 1983: 48). Consequently, speakers stop seeing the single components or the construction of a word and regard it as a whole lexeme instead (Bauer, 1983: 48). As the new word takes on a form in the speakers’ minds, that it could not have done through the application of productivity rules, it begins to behave like a monomorphic entity (Bauer, 1983: 48). Therefore, institutionalisation is “the integration of a lexical item, with a particular form and meaning, into the existing stock of words as a generally acceptable and current lexeme” (Lipka, 2002: 112).

Institutionalisation and conventionalisation depict two entangled sociolinguistic processes (Lipka 2002: 112). Institutionalisation is “the result of collective cognitive entrenchment [(cf. section 3.2)] spreading over the mental lexicons of the members of a given speech community in a process of conventionalisation” (Langlotz, 2006: 99). In the process of institutionalisation, the speakers’ mental lexicons are affected by a new item getting listed with its full derivational history (Bauer, 1983: 48). Therefore, the way a new word-formation comes to be conventionalised and accepted in the vocabulary of the community is by means of institutionalisation (Brinton & Traugott, 2005: 45).

The entanglement of institutionalisation and conventionalisation also resonates in the fact that similar forces, such as nameworthiness/topicality and frequency promote them (Fischer, 1998: 176) (cf. section 3.1.2). Nameworthiness mainly depends on extralinguistic circumstances, such as either the speakers’ surroundings - a word like *snowman* is very unlikely to succeed in most parts of Africa due to its lack of nameworthiness (Lipka et al., 2004: 10) - or changes in society (Lipka et al., 2004: 11). Once a word is very topical, the use of contextual clues can be reduced (Fischer, 1998: 103-106). Nameworthiness, though, is relative, which makes institutionalisation not a one-way process but indicates that some words, after being

institutionalised, might de-institutionalise again, since they could be “used in connection with current affairs [only] for a short period of time” (Fischer, 1998: 16). *Millennium-bug* is an example of de-institutionalisation: while it was in frequent use at the end of 1999, it hardly occurred 15 years later (Kerremans, 2015: 38). Thus, nameworthiness often goes hand in hand with an increased frequency as well as an enhanced transparency (Bauer, 1983: 48).

While most research supports the assumption that institutionalisation is dependent on topicality, context, and situation, Hohenhaus found that the word *bouncebackability* does not quite follow this rule (Hohenhaus, 2006: 17ff.). Despite the fact that it is not clear whether the word is a *-able* derivative or a compound, and speakers did not have a great type-familiarity with the word, it turned out to be quite successful (Hohenhaus, 2006: 19). This could be the result of an ‘artificial’ institutionalisation by means of media promotion (Hohenhaus, 2006: 22). Thus, the possibility that words can be deliberately pushed should be kept in mind. In sum, institutionalisation denotes the process in which a word spreads and starts being perceived as a lexical item. Fuel for this process is nameworthiness, which often goes along with increased frequency in use. However, it can occasionally be overruled by the power and influence of the media.

Lexicalisation was neglected in linguistics for a long time (Lipka, 1977: 155) - Marchand’s first edition of *The Categories and Types of Present-Day English Word-Formation* does not mention it in the outline, only in the second edition did it become slightly more prominent (Marchand, 1960: 81-83, 1969). Although it is much more researched today, the approaches over the years differ significantly. In contrast to the recent perception that lexicalisation resembles the usualisation subprocess of symbolisation, Bauer’s approach puts it as a follow-up step to institutionalisation. A successive approach, however, would ignore the dynamics involved in conventionalisation.

The cognitive process of routinisation is needed for usualisation (cf. section 3.1.1), so that hypostatisation has often been seen as a motor for lexicalisation (Lipka, 1977: 155). Hypostatisation is regarded as “*die Erscheinung, daß die Existenz eines sprachlichen Zeichens auch die Existenz eines einzigen von diesem bezeichneten Dings suggeriert*” [the phenomenon, that the existence of a linguistic sign also suggests the existence of a single thing being denoted by the sign] (Lipka, 1977: 161). This idea signifies that every concept formation mirrors an existing entity in the extralinguistic reality (Lipka, 1975: 200). The term denotes the ‘concept-forming power of the word’ (Leech, 1981: 32) and therefore causes ‘*Vergegenständlichung*’ [objectification] (Lipka, 1977: 161) and evokes the impression that one single linguistic sign should also only denote one referent. The fact that this works for concrete objects like *raincoat* or extralinguistic phenomena like *holiday* supports this objectification (Lipka, 1977: 162). Thus, a word, which starts being accepted widely in a speech community, develops a limited, specialised, and fixed meaning through hypostatisation (Bussmann, 1996: s.v. “hypostatization”).

These specifications often go along with phonological, semantical, morphological, syntactical, and motivational changes (Lipka, 1977: 155). Each of these processes can take place on a continuum. Thus, phonological lexicalisation might only affect the reduction of the final vowel, such as in *postman* or *Monday*, or can be more pronounced as in *breakfast* or *prayer* (Lipka et al., 2004: 9). It can also refer to prosodic changes happening after institutionalisation, such as a change in word-stress in a way which would not have been possible if the word was the product of a productive process (Bauer, 1983: 50). Changes could also affect segmental features, like morphs (Bauer, 1983: 50). Morphological lexicalisation includes the use of linking elements such as the German *Fugen-S* or changes in affixes. Suffixes like *-ment*, *-ric*, *-dom*, for instance, are not productive anymore and thus lexicalised (Bauer, 1983: 53; Brinton & Traugott, 2005: 51). Sometimes, morphemes can lose their grammatical and semantic contribution to a word and become an

indistinguishable part of it, even though they keep their original phonological structure (e.g. *alone* < *all* + *one*). This process is also denoted as *demorphologisation*.

It is possible that grammatical inflection is preserved without actually being used, like in *whilst*, where the genitive case is kept or in *elder*, which still bears the comparative form but is not perceived as such any longer (Brinton & Traugott, 2005: 52). Hence, for some morphological lexicalisation processes it can be stated that “today’s morphology is yesterday’s syntax” (Givon, 1971: 413) and “today’s grammar might become tomorrow’s lexicon” (Ramat, 1992: 557). A word like *pickpocket*, which contains the verb and its direct object, a pattern that is not productive in English, is an example of syntactic lexicalisation. Verbs like *disbelieve* and *believe*, where the latter can take an accusative and the former cannot, are examples of word external syntactic lexicalisation (Bauer, 1983: 59-61). Finally, semantic lexicalisation usually involves a ‘*Bedeutungsverengung*’ [narrowing in meaning] (Lipka, 1977: 157). Words that contain only very general meaning elements (e.g. [+HABITUAL]) are not greatly lexicalised.

Vice versa, the more new meaning elements are added to a neologism, the more lexicalised it becomes. Thus, words that bring specific idiosyncratic characteristics with them, such as *wheelchair* and *pushchair*, which cannot only be defined as a ‘chair which has wheels’ and a ‘chair which one pushes’, but contain additional information, such as ‘for invalids’ or ‘for infants’, are highly lexicalised (Leech, 1974: 226 f.). Another example of such an idiosyncratic entity and interference with regular word-formation processes is *cobweb*, the first part being derived from Old English *coppe* meaning *spider* (Lipka, 2002: 11). Semantic lexicalisation can also be contextual. With a word like ‘*reader*’, the context will reveal whether it denotes ‘a person that reads’ or, for instance, a ‘university teacher ranking between professor and lecturer’ (Lipka, 1977: 159).

The loss of semantic criteria can sometimes be so strong that it can lead to complete idiomatisation and thus to a word being synchronically unanalysable, as in the case of *understand* (Lipka, 1977: 160). Idiomatisation is commonly identified with lexicalisation (Bauer, 1983, Lipka 1992, Bussmann 1996, Brinton & Traugott 2005). However, besides definition problems of the term idiom itself, there is also no accordance or clarity between linguists as to how idiomatisation and lexicalisation are related to each other. Some argue that idiomatisation is the best example of lexicalisation (Moreno Cabrera, 1998: 214), that both terms are closely linked to each other but cannot be used synonymously, since idiomatisation only affects a part of the changes that can happen during lexicalisation and usually depicts a consequence of lexicalisation (Lipka, 1977: 155).

Others claim that idiomatisation is the diachronic element¹⁵ of lexicalisation, denoting the state in which the original meaning can no longer be deduced from individual elements (Bussmann, 1996: s.v. “lexicalization”; also see “idiomatization”). In other words, idiomatisation concerns the semantic changes involved in the process of lexicalisation (Lipka 2002: 113). It can manifest itself “as the addition or loss of semantic features. Synchronically, the result of this process, various degrees of idiomaticity, form a continuous scale” (Lipka 2002: 113).¹⁶

Along with changes in phonology, semantics, syntax and morphology in hypostatisation goes the loss of motivation, hence

¹⁵ Despite the general agreement that lexicalisation is a diachronic process, Bauer adds that it “is essentially a diachronic process, but the traces it leaves in the form of lexicalized lexemes have to be dealt with in a synchronic grammar” (Bauer, 1983: 50). Lipka also argues that lexicalisation is both, a diachronic and synchronic process, which, however, can only be explained diachronically (Lipka, 1977: 155).

¹⁶ It is sometimes even argued that idiomatisation may include morphological changes (Brinton & Traugott, 2005: 98), as it is associated with routinisation which leads to univerbation, compacting and simplification (Brinton & Traugott, 2005: 54). This routinisation is lexicalisation in the sense of a word coming to belong to the inventory (Lehmann, 2002: 14).

demotivation. This can be caused by extralinguistic factors, as well as language-internal ones. An example is *blackboard*, which today, in most cases is green and therefore a change in denotatum has caused demotivation (Lipka et al., 2004: 10). On a language-internal level, *mincemeat* is defined as “a mixture of dried fruit, spices, etc. used especially for making pies” (Advanced Oxford Learner’s Dictionary), which has nothing to do with *meat*. While the motivation of the second component was still clear in Old English, when *mete* simply denoted ‘food for a human’, its meaning became more restricted in Modern English (Lipka, 1977: 157).

The term *motivation* goes back to Saussure and his pupil Bally and was later further developed by Ullmann (1962: 81ff.). Motivation can refer to phonological motivation such as is the case with onomatopoeia (*cuckoo*), morphological motivation like word-formation (*preacher*), semantic motivation such as metaphor or metonymy (*coat of paint*) or a mixture of several types of motivation (*bluebell*). When words lose various types of motivation they transfer from transparent to opaque words, the result of demotivation (Lipka et al., 2004: 3). Thus, hypostatisation as an engine for lexicalisation can affect all sorts of linguistic areas to various degrees. From this it follows that lexicalisation, as a subprocess of usualisation is responsible for which meaning of a word gets established, as well as how the word behaves in regard to losing motivation or phonological, morphological etc. criteria.

This implies that lexicalisation does not only take place in society but is dependent on cognitive events in the speakers’ minds. As the entrenchment feedback loop (cf. fig. 4) shows, repeated activation will lead to an ease in activation. Therefore, similar to institutionalisation, lexicalisation is also partially driven by frequency of occurrence. It is widely believed that frequency of usage represents a direct cause of lexicalisation (Aronoff & Anshen 2001: 240; Lipka, 2002: 111; Bakken 2006: 107).

When looking at the different steps from non-lexicalised to lexicalised, nonce-formations are defined by low frequency, while institutionalised words already have an increased frequency and lexicalised words are even more frequent (Fernández-Domínguez, 2010: 202). However, the connection between lexicalisation and frequency is not absolute. In section 2.1 it was mentioned that lexicographers often make decisions about which words they add to a dictionary based on their usefulness. This can also be applied here, considering that if a word is seen as useful by speakers, they will more commonly call for it, while vice versa less important and less useful words might only rarely be employed (Fernández-Domínguez, 2010: 202) (cf. fig. 8). Hence, the succession of the factors shows how token frequency and lexicalisation correlate, and that usefulness could be seen as a root for frequency.



Figure 8 Factors affecting token frequency (Fernández-Domínguez, 2010: 202)

Summing up all the above, lexicalisation can be defined as

“die Eingliederung eines Wortbildungs- oder syntaktischen Syntagmas in das Lexikon mit semantischen und/oder formalen Eigenschaften, die nicht vollständig aus den Konstituenten oder dem Bildungsmuster ableitbar sind.”
 [the incorporation of a word formation of syntactical syntagma into the lexicon with semantic and/or formal characteristics, which cannot be fully deduced from the constituents or the formation pattern] (Kastovsky, 1982: 164-165)

It illustrates “the process by which complex lexemes tend to become a single unit with a specific content, through frequent use. In this process, they lose their nature as a syntagma, or combination [of smaller units], to a greater or lesser extent” (Lipka, 1992: 107; 1981: 120; Lipka et al., 2004: 5) and it can thus be seen as “a gradual, historical process, involving phonological and semantic changes and the loss of motivation” (Lipka, 2002: 113). Due to these changes, the application of general

grammar rules to the expression becomes more and more restricted (Barkema, 1996: 135; Langlotz, 2006: 100). A lexicalised item becomes a listed entity as part of the lexicon and can no longer be generated by word-formation (Fernández-Domínguez, 2010: 200). Lexicalisation leaves idiosyncratic traces (Bauer, 1983: 50) and the more idiosyncratic a word becomes, the more it loses its regularity and, when being reproduced, is subject to direct stipulation, similar to the direct retrieval of words (Langlotz, 2006: 99). This implies that a result of lexicalisation is that a word is stored in the lexicon and will be recalled from there when needed (Bussmann, 1996: s.v. “lexicalization”).

Lexicalisation therefore goes along with weakening the mental activation-set, implying that lexicalised constructions are institutionalised by definition, as their production is no longer guided by general principles of linguistic composition. Therefore, the final output of lexicalisation is a lexical and content item in the inventory that needs to be learnt and which can be of any complexity. Lexicalisation results in an irregularity of the lexicon, which can only be explained historically (Lipka 2002: 113). As a whole it depicts a gradual rather than an either-or process (Bauer 2004: 19) and involves many changes such as a decrease in pattern productivity as well as in token productivity (Brinton & Traugott, 2005: 96 f.).

To sum this up, while “institutionalization is defined as the reasonably frequent occurrence of a word, lexicalization refers to the emergence of word-specific additional semantic content beyond what can be predicted from productive word-formation processes” (Mair, 2006: 37). Hence, whereas institutionalisation can be seen as a socio-pragmatic process of the language users’ interactive accommodation of their cognitive grammar, lexicalisation depicts a cognitive process of routinisation (including the word’s loss of transparency and idiomaticity) (Langlotz, 2006: 99). They are both of a more-or-less, rather than an all-or-none nature (Lipka, 1972: 76), as the number of changes can vary from little phonological and semantic deviations (e.g., *postman*, *blackboard*) to bigger, even combined, graphemic, phonological, or

semantic changes (e.g., *cupboard, prayer, holiday*) (Lipka et al., 2004: 9). Furthermore, both complex lexicalised and institutionalised lexemes stand in between *langue* and *parole*, on the level of the *norm* (Lipka, 1992: 96). The term *norm* was first introduced by Coseriu as an intermediate level in the Saussurean dichotomy (Coseriu, 1967: 11). It is not restricted to the lexicon, but “is particularly useful to apply the concept of norm, as the traditional, collective realization of the language system, to lexicology” (Lipka et al., 2004: 3).

The norm is responsible for alternative word-formation types, like the use of *nationalize* instead of **nationalify*, lexical gaps and habitual disambiguation such as using *sleeping pills* (FOR sleeping) but *headache tablet* (AGAINST headache) (Lipka et al., 2004: 3). Furthermore, to a certain extent, they are both positively influenced by frequency. In short, it can be said that these two processes depict a part of the conventionalisation feedback loop that is particularly tailored for new lexical items. While one of them is located within diffusion and hence affects the wider society, the other one is located within the usualisation process and incorporates cognitive aspects.

3.1.4 Conventionalisation summed up

Summarising the last few sections, it became apparent that conventionalisation depicts a rather complex process by which regularities of communicative behaviour get established and readapted in a wider speech community/society. The fact that these regularities need to be constantly readapted and updated hints towards how dynamic the process is. It shows that this is not a one-way mechanism with a set end point, implying that once utterance types are conventionalised, they do not necessarily stay like this but are subject to constant change. Affecting the whole speech community makes conventionalisation a social process that is urged by usualisation and diffusion. Although these two catalysators are distinct they share common features, such as the fact that they rely on similar mechanisms or that they are both constantly

at work, as part of the ongoing conventionalisation feedback loop. They also affect and influence each other.

Furthermore, various forces impact usualisation and diffusion, so that, for instance, interpersonal activities, usage frequency or newsworthiness can positively as well as negatively stimulate diffusion and usualisation and thus, conventionalisation. They also both have an impact on language change, even though they influence different areas of change. This is important since this thesis deals with lexical changes and therefore, while conventionalisation affects all sorts of utterance types, it can be applied to the establishment of new lexical items. In connection with this, lexicalisation and institutionalisation are two important processes that can be located within diffusion and usualisation, that are specific variants for lexis and thus serve to describe the conventionalisation process for neologisms in more detail. Although conventionalisation is per definition a social process, it is contingent to the requirements of entrenchment processes which are taking place in the speakers' minds.

3.2 Entrenchment

While conventionalisation is of major importance for the empirical part of this thesis, entrenchment will not play a predominant role. However, due to their entanglement (cf. fig. 4, section 3.1 ff.), it is important to outline entrenchment. Like the definition problems in previous sections, *entrenchment* also depicts a notational term that must be explicitly defined but differs depending on the author (Lipka, 2010: 96). While it is repeatedly argued that entrenchment only takes place in an individual's mind, Lipka distinguishes between entrenchment in a speaker's mind ('individual entrenchment') and in a whole speech community ('social entrenchment') (Lipka, 2010: 96). Individual changes can indeed contribute to collective change as the example of the word *jeans* shows. The word is the result of a collective upgrade to a basic-level term due to the high cultural importance of this type of trousers (Geeraerts et al., 1994).

When it comes to lexical items, entrenchment is sometimes also seen as a superordinate metalinguistic term for the hyponyms lexicalisation and institutionalisation, describing the cognitive side of those processes (Lipka, 2010: 96). As languages are regarded as being comprised of a structure inventory of linguistic units which are connected by relations, categorisation, composition, and symbolisation (Langacker, 2017: 39), entrenchment is seen as the automatization process through which a linguistic structure achieves unit status (Langacker, 1987: 57, 2008: 16). An increased entrenchment goes along with a raised automatization. This becomes clear when looking at processes such as reciting the alphabet: through repetition or rehearsal, a complex structure is thoroughly mastered to such an extent that it becomes automatic and requires little conscious monitoring (Langacker, 2008: 16). However, it is not a one-way process, and it is important to acknowledge that entrenchment is dynamic and can thus be defined as follows:

“Entrenchment is the continual reorganization of linguistic knowledge in the minds of speakers, which is driven by repeated usage activities in usage events and subject to the exigencies of the conventionalization processes taking place in speech communities.” (Schmid, 2020: 2)

Similar to conventionalisation, there are several different catalysators and forces that promote entrenchment.

3.2.1 Catalysators for entrenchment

While the EC-model links entrenchment to conventionalisation, they are nonetheless two processes with different feedback loops. Entrenchment is catalysed by routinisation and schematisation. Routinisation as a process entails a change in the strength of patterns of associations in the speakers' minds, leading to an individual, cognitive change rather than a collective one (Schmid, 2020: 317). This individual linguistic knowledge is represented in four different types of associations: *symbolic*, *paradigmatic*, *syntagmatic*, and *pragmatic*. Thus, according to Schmid speakers do not entrench utterance types but rather patterns of

associations, which become active while processing utterance types (Schmid, 2020: 206).

Symbolic associations process form, meaning and their connection and depict a “cognitive substate of multiplex semasiological and onomasiological connections within utterance types” (Schmid, 2020: 46). *Paradigmatic associations* establish a connection between the potential target of symbolic associations, also referred to as activation sets (Langacker, 2000: 105), which compete during productive as well as receptive processing. The winner is the one that will be used for production in the end (Levelt et al., 1999). *Syntagmatic associations* follow the principle of linearity (de Saussure, 1916). They connect the mental states with formal and semantic aspects of parts of the sequences of utterance types. They therefore operate on the level of communicational and contextual contingency. How much effort is needed to activate these syntagmatic associations depends on how often they have previously been activated. Highly activated syntagmatic associations enable the hearer to form expectations about what will come (Schmid, 2020: 47). *Pragmatic associations* affect the situational context and mediate between perception and processing of context-dependent and functional aspects of meanings such as reference, deixis, aspect etc. (Schmid, 2020: 48).

To exemplify this Schmid names the case of *fell*: the symbolic association connects the word-form *fell* and the meaning ‘suddenly went down to the ground’. The paradigmatic associations denote the competition with *fill*, *full*, *tell*, etc. The syntagmatic associations are triggered by the cotext (*the toddler stumbled and...*) and the pragmatic associations by the perception of the situational context which prime the target meaning, rather than the meanings ‘decrease’ or ‘be killed’ (Schmid, 2020: 49).

These associations are the ability of ‘one kind of experience [...]. to evoke another’ (Langacker, 2000: 94), by being part of an associative network in the brain. This network consists of countless connections between neurons, some stronger than others. When processing something, the network constantly moves from one constellation of

assemblies of firing neurons to the next. This entails that none of these constellations are fixed. This flexibility leads to *associative activation* (Kahneman, 2011: 51), implying that one cognitive process activates and triggers others. It should be noted that the term *associative activation* is misleading in so far that, although neighbouring areas in a network will get co-activated, patterns of associations are competing for activation. Frequent words are usually found in an entrenched pattern and thus, depict an attractor which is easily and effortlessly reached due to its frequent previous activation (Langacker, 2000: 7; 2017: 47). This entails that processing of unentrenched utterances, such as neologisms, is much more effortful and potentially slower.

While associations are not imperative to this thesis, it is important to be aware of the four types of associations as they depict an important part of the entrenchment feedback loop. The processing of language and its representation of linguistic knowledge is denoted by various degrees of routinised entrenched patterns of associations. These patterns of activation in the associative network consist of cooperation and competition between the four types of associations. When processing an utterance, these patterns become active in working memory (Schmid, 2020: 49). Therefore, rather than the pure retrieval, access, and storage models, this associationist model of linguistic knowledge does more justice to the flexibility, cotext and context adaptability of processing and the multi-dimensional contingency of representation (Schmid, 2020: 49). Speakers routinise patterns of associations since they provide the knowledge to de- and encode linguistic utterances in different situational usage events. In short, routinisation can be summarised as:

“Routinization (Haspelmath 1999: 1055, Langacker 1987: 100) strengthens patterns of associations representing the commonalities of highly similar recurrent utterance types, e.g., word forms and fixed expressions.” (Schmid, 2020: 6)

As mentioned, usualisation cooperates with routinisation, however, they are two distinct processes, since usualisation is located on the social

spectrum – indicating a change that concerns the regularities of behaviour shared in the community – in contrast to routinisation, which is about a change concerning the strength of associations in the individual’s mind.

Figure 9, taken from Schmid (2020: 298), shows how the collective level of conventions with the six dimensions of conformity, and the cognitive level, with the four types of associations, interact. While syntagmatic conformity and syntagmatic associations are a near overlap, since they both relate to the linear dimension of structure, the onomasiological (competition between forms and one communicative goal) and semasiological (competition between different meanings and one form) dimension of conformity bear resemblance to the paradigmatic and symbolic associations, as the cognitive substrate of form-meaning and meaning-form regularities. The cotext, context and social aspects of conformity are all within the range of paradigmatic associations. However, the context-related aspect of meaning and the communicative function are also linked to the onomasiological, semasiological and syntagmatic conformity, indicated by the dotted lines (Schmid, 2020: 298). The fact that the processes of usualisation and routinisation, and with them the conformity dimensions and association types, are closely linked makes the methodological problem to find which of these two processes has taken the lead, as this can only be found out if changes in the collective frequency (such as shown in corpora) are compared to frequency from individual speakers (Schmid, 2020: 317).

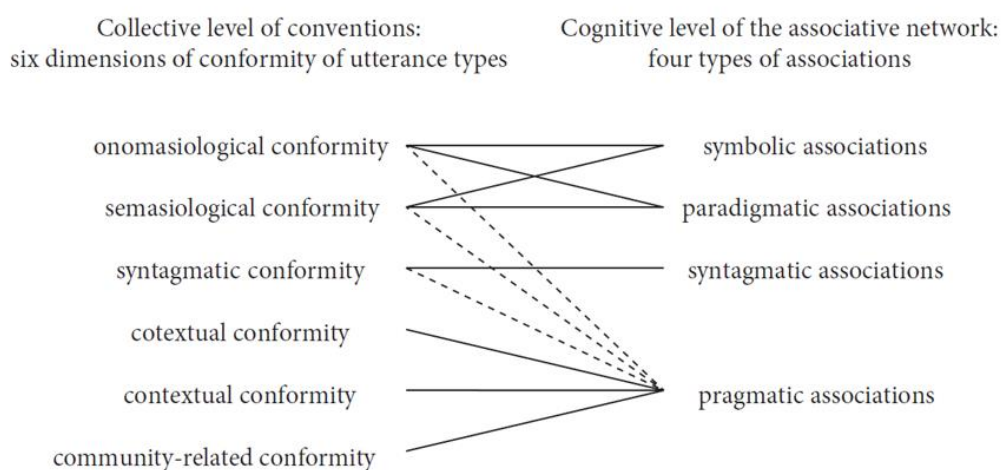


Figure 9 Relation between conformity profiles and types of associations (Schmid, 2020: 298)

While routinisation entails generalisation, schematisation can be seen as a side effect of routinisation and statistical learning (Frost et al, 2015: 118) (cf. section 3.2.2). Schematisation is an integral component of routinisation, but the degree to which it is effective is subject to what becomes routinised.

“*Schematization* (Abbot-Smith & Tomasello 2006, Langacker 2008: 17) strengthens patterns of associations representing the commonalities of variable recurrent utterance types, e.g. semi-filled lexico-grammatical patterns (e.g. that’s Adj, as manifested by *that’s right/great/lovely*, etc.)” (Schmid, 2020: 6)

Being an integral part, it can be said that there is no qualitative difference between routinisation (token entrenchment) and schematisation (type entrenchment) but a quantitative one, which correlates with the degree of variance of what becomes routinised (Schmid, 2020: 343). It should be noted that within the EC-model, conventionalised items do not depict an input to entrenchment but are a trigger for it. Patterns of associates are activated during co-semiosis, which are then eligible for routinisation and schematisation (Schmid, 2020: 205). In sum, routinisation and schematisation are two processes that are closely linked and work together, but at the same time affect different aspects of entrenchment.

3.2.2 Psychological foundations for entrenchment

Several psychological processes serve as the fundament of entrenchment. One of them is statistical learning, the basis for entrenched patterns of associations and one of the major psychological learning mechanisms behind entrenchment (Schmid, 2020: 207). In statistical learning, learners implicitly form associations between stimuli by tracking and storing the underlying statistical relationships between such elements. This allows the speakers to make predictions about what will happen next, based on whether the type of situation has been repeatedly experienced before (Schmid, 2020: 207).

In regard to content words, statistical learning of symbolic associations is fairly straight-forward, since the more often a speaker is confronted with a lexical item referring to a given entity, the stronger the associations between form and concept become (Schmid, 2020: 207). Statistical learning via syntagmatic associations conforms to the conception of entrenchment as unit-formation (Blumenthal-Dramé 2012; Langacker, 1987). Statistical learning of syntagmatic associations underlies the phenomenon of chunking and the learning of representations of generalised patterns (Christiansen & Chater, 2016a, 2016b). It is used by speakers to “constantly update their knowledge of lexical [...] co-occurrence tendencies that reflect the conventionality based on syntagmatic conformity among the members of the speech community” (Schmid, 2020: 208). Statistical learning of paradigmatic associations can only take place in a syntagmatic or pragmatic context. Through repeated co-activation between similar words (in the sense that they have a similar meaning, can be used in a similar pattern etc.), paradigmatic associations are established and routinised by means of statistical learning. Vice versa, speakers potentially also learn which words are not used in the respective pattern, which is denoted as ‘statistical pre-emption’ (Boyd & Goldberg 2011, Goldberg 2019: 75–84) or ‘negative entrenchment’ (Stefanowitsch, 2008). Pragmatic associations are also subject to statistical learning. From previous

experiences and from what speakers have heard before in similar situations, they can learn what to say or expect in similar situations – which is, for instance, important for style and register (Schmid, 2020: 209).

While learning is often of statistical nature and hence based on repetition, it is also possible to learn by surprise. If a speaker's routine is broken by something unexpected, they can learn from it (Barto et al. 2013, Rescorla & Wagner, 1972). However, in contrast to an increasing routinisation of already entrenched patterns of associations through repetition, individual surprise-based learning is more likely to trigger and support collective language change. Due to the surprise element in conversation and thus a difficulty in co-semiosis, the associative network is more likely to register something uncommon (Schmid, 2020: 210).

Apart from statistical learning, memory consolidation is a key psychological mechanism for entrenchment. Linguistic knowledge and skills to use a language must be stored. Therefore, speakers possess a mental lexicon, a neural network, that contains information regarding a word's meaning, pronunciation, and syntactic characteristics (Aitchison, 1994: 228). The “lexicon is conceptually necessary as the long-term memory repository of available pieces of language from which the combinatorial system can build up larger utterances” (Jackendoff, 1997: 109). However, for knowledge to be stored in long-term memory, it needs to go through the working memory. While this seems straight-forward, it is still debatable where the working memory is located (Fiez, 2016). What is clear though, is that the working memory is the active part of the limited capacity of short-term memory (Baddeley & Hitch, 1974; Buchsbaum 2016) and must therefore be separated from the long-term memory.

The prominently used model suggests subdividing working memory into four parts: the domain general executive control centre, two modality-specific parts - i.e. verbal working memory (*phonological loop*) and visuospatial working memory (*visuospatial sketchpad*) - and an *episodic buffer* which is a temporary storage for episodic information and

interacts with long-term episodic memory (Baddeley: 2000, 2010). The space within the buffer is limited though, so that the working memory has strategies to group information in chunks (Schmid, 2020: 211). It is generally agreed that four items or chunks are the maximum capacity of what the buffer can 'hold' (Baddeley, 2010; Cowan, 2001). In this context, chunking is defined as grouping primitive stimuli into larger perceptual and conceptual units, making "a chunk [a] collection of elements having strong associations with one another, but weak associations with elements within other chunks" (Gobet et al., 2001: 236). For information to go from working memory to long-term memory, memory consolidation needs to take place (Takashima & Bakker, 2017). Thus, a lexical item must firstly be processed in working memory and can then be consolidated from episodic to semantic memory. While the former stores autobiographic information, the latter is decontextualised. Entrenchment depicts the shift from episodic to semantic memory, going hand in hand with a reduction of the strength of pragmatic associations and an increase of symbolic associations. This shift goes along with the integration of a new element into existing lexical networks.

Apart from that, the item also needs to be routinised in the procedural memory, which is linked to routinisation and automatization of segmental aspects of fixed units and variable patterns and contributes to fast and automatic activation and production of implicit routines that are necessary for processing linguistic patterns. Furthermore, it was found that comprehension is more automatized than production (Schmid, 2020: 213). Especially when it comes to larger units, both fixed and variable, a collaboration between declarative and procedural memory is needed (Schmid, 2020: 216). Generally, there are some reservations as to how automatic complex and variable schemas are processed. Simple lexical items offer the best conditions for automatization (Schmid, 2020: 214).

A word can only be produced and understood if the currently processed working memory is brought together with the knowledge stored in long-term memory. Even though this seems clear enough, how this happens is a matter of debate. One approach are the productive

coding theories (for surveys, see Clark 2013; Huang & Rao, 2011; Rauss & Pourtois, 2013), which denote that experience-driven long-term representations contribute to predicting upcoming elements in processing and balance bottom-up and top-down processes (Schmid, 2020: 50). The models aim to integrate perception, cognitive processes, memory, representation, learning and action (Friston, 2010) as minds constantly generate context-sensitive hypotheses of what will happen next on the basis of stored representations that are formed from experience (Schmid, 2020: 50). Within these models, the ease of processing depends on the strength of symbolic associations of the target word so that the ‘resting activation’ for connecting form and meaning is seen as a result of long-term entrenchment by repeated exposure and use (Baayen, 2010). It also depends on the strengthening of pragmatic associations, which mediate the information about the context of the situation (Kroczek & Gunter 2017). Furthermore, ease of processing is subject to the strength of syntagmatic associations in regard to the preceding words and depends on their number and strength of paradigmatic associations to other words that are competing for an occurrence in the same target slot (Schmid, 2020: 51).

There are different ways of how processing can be measured, such as relative frequencies per million words, transition probabilities, log-likelihood statistics or dispersion (Schmid, 2020: 52).¹⁷ In this context, lexical items are regarded as points of access to a network (Langacker, 1987: 163). However, they are not readily made and do not correspond to “go-to-mental-lexicon-and-grab-it” kind of processes (Schmid, 2020: 53) but rather to dynamic and transient multidimensional activation

¹⁷ While these measurement tools will not be detailed, it should be noted that *relative frequencies per million words* are used as approximate indicators of the so-called resting activation level (Morton, 1969, Plag, 2003, Weber & Scharenborg, 2012), *transition probabilities* are used to approximate the strength of syntagmatic associations representing the likelihood of one element following another, *log-likelihood statistics* (Dunning, 1993) are used as a significance test, i.e. as an indicator of the confidence that we can have in the assumption that a prediction we make is not due to chance. The number of paradigmatic competitors is used as a measure of *dispersion* in the target slot.

patterns (Casasanto & Lupyan, 2015). How quickly they can be accessed, and how much effort is needed to do so depends on the availability of well-entrenched attractors. This idea agrees with the notion in neuroscience that the 'brain's' concepts are based on long-term representations, which derive from experience which makes them flexible and context-dependent (Kiefer & Pulvermüller 2012: 86–88; Ramscar & Port, 2015). Several effects have been found regarding dynamic lexical access and retrieval, such as the frequency effect, semantic priming effect, word length effect, neighbouring effect, recency or context effect, practice effect and word-superiority effect. In the framework of this thesis word frequency is an important factor. It was found that more frequently used words are recognised faster and more accurately (Rayner & Duffy, 1986). Due to the repeated processing of symbolic associations, there is a link between perceived forms and meanings.¹⁸

However, other effects should also be mentioned, as they could potentially be linked to the success of a neologism. Words that are preactivated by semantically related words are activated faster (Neely, 1977). Depending on whether such primes are related syntagmatically (*bread—butter*) or paradigmatically (*river—stream*), syntagmatic or paradigmatic associations are preactivated in the predictive model constructed during processing (Schmid, 2020: 54). If a word has been encountered or primed by context shortly beforehand, the word in question is also recognised faster, since syntagmatic and paradigmatic associations have been activated just a short while ago (Zwitserslood, 1989). Besides this, shorter words are recognised faster than longer words (Forster & Chambers, 1973). On the one hand, shorter words are generally more frequent than longer words (Zipf, 1949), and therefore more likely to become entrenched by strengthening the symbolic associations between form and meaning and the paradigmatic

¹⁸ It has to be said that the frequency effect as such is collinear with and possibly superseded by other effects (Baayen et al. 2016), such as local frequency effects determined by the current linguistic context (Baayen 2010, McDonald & Shillcock, 2001), and by effects of the dispersion across different contexts (Adelman et al. 2006, McDonald & Shillcock, 2001).

association with competing forms. On the other hand, short words are more amenable to holistic recognition and processing, which also favours the effortless activation of symbolic and paradigmatic associations (Schmid, 2020: 54).

In contrast to this, words with dense neighbourhoods, i.e. words with many very similar words, are recognised more slowly than those with few neighbours. Further, if neighbours with higher frequency are available, they will impede access to the less frequent target word (Grainger et al., 1989). In the EC-Model, these two effects can be explained by the cooperation and competition between symbolic, syntagmatic, and paradigmatic associations. Experience can also have an impact on how quickly a word is recognised. Experienced readers were found to recognise words faster, generally read faster, fixate fewer words, and process more deeply (Golinkoff, 1975). Further, words are recognised faster than non-words (Paap et al., 1982).

When bringing together associations and lexical-semantic processing, it can be said that associations, especially syntagmatic ones, create expectations about the length and potential components and their grammatical properties. Associations and lexical-semantic processing further limit the paradigmatic options and prepares the associative network for symbolic and pragmatic associations. Paradigmatic associations often depend on the strength of the bond between syntagmatic and paradigmatic associations and can have a permanent but changing effect on the other associations. Symbolic associations mainly process content words. This means that the processing model integrates context, long-term traces of competing syntagmatic associations (even to the extent of several words and beyond the sentence boundaries), semantic and conceptual layers such as symbolic paradigmatic and syntagmatic associations and conceptual projections (Schmid, 2020: 71). Thus, all four associations constantly work together.

Nevertheless, this is all about comprehension. Important for this thesis, however, is also the production side of things. When looking at speech errors, it becomes clear that the model also works for production. Such errors often imply problems with syntagmatic and paradigmatic

associations. For instance, word-initial anticipation (e.g., *bake my bike* instead of *take my bike*) and perseveration (e.g., *she can she it* instead of *she can see it*) seem to show errors that are mainly caused by syntagmatic associations, whereas word reversals (e.g., *a first fine half* instead of *a fine first half*) and substitution errors (e.g., *in our academic ivory league* instead of *in our academic ivory tower*) are caused by paradigmatic associations (Schmid, 2020: 72). However, it is difficult to make a clear cut, and to some extent there is always a cooperation of both dimensions of associations.

Important for both, comprehension as well as production, is the question of whether words are accessed and processed holistically or analytically. This also plays a role for neologisms, since it has an impact on how they are processed when encountered. While it is commonly agreed on nowadays that words are not necessarily ready-made tools (Schmid, 2020: 92), it was previously argued that the mental lexicon contains a full list of stored forms, maybe even combined forms (Brinton & Traugott, 2005: 9), and consists of unanalysable wholes, which are accessed holistically. This Full Listing Hypothesis implies that each word holds its own representation in the lexical memory, including inflected and derived forms (Butterworth, 1983). It was also argued that with monomorphic words it seems as if people like to memorise as many as possible (Jaeger, 1986: 76). How problematic this approach is, becomes clear when looking at agglutinating languages. If all forms were stored holistically, it would go beyond the human storage capacity (Bauer, 2001: 100).

Closer to the current research comes the idea of words being *semi-listed*, implying that some forms are stored while others are generated on-line (Aitchison, 1994: 228; Jackendoff, 1997: 231, note 11). The idea is that there is a differentiation between a *virtual lexicon*, which is a space for possible derived forms, and an *actual lexicon*, which is a list of occurring items (Jackendoff, 1997: 117). This goes in line with the position that lexical rules rather determine what is possible, not what is actual (Lees, 1960).

Nowadays, there are several competing models regarding the representation of simple and complex lexemes in the mental lexicon. The rule-based dual-route models (e.g., Pinker, 1998; Pinker & Prince, 1988; Prasada and Pinker, 1993) claim that regular complex forms are stored as stems and affixes and combined compositionally during processing, whereas irregular forms are stored and accessed holistically. Hence, regular forms are of analytical nature and irregular ones of holistic. This approach suggests a maximised rule-based online computation and a minimised demand on the memory and bears some similarities with the idea of items being *semi-listed*.

Opposed to this, exemplar-based and connectionist single-route models (e.g., Bybee, 1985; Bybee, 1995; Rumelhart & McClelland, 1986) claim that all forms are accessed by a single route to the associative network.¹⁹ Their weights are determined by the token and type frequency of the individual forms (Schmid, 2020: 242). It is assumed that an entrenched utterance “represents an automated, routinized chunk of language that is stored and activated by the language user as a whole, rather than ‘creatively’ assembled on the spot” (De Smet & Cuyckens, 2007: 188). Thus, chunk status refers to the idea that a string which can be analysed in smaller subcomponents is perceived as a unit, and thus can be retrieved in a single step rather than by accessing its component parts and their composition. Taking a common expression such as *the thing is that* as an example, it can either be stored as one unit or all the single parts can be saved in their own units, which means that non-unit and unit status are not dichotomous (Langacker, 1987: 59). Therefore, *chunking* indicates that there are sometimes two different lexical entries for the same word.

¹⁹ Various researchers from both groups have elaborated different models that deal with the processing of complex words. Since this thesis deals with the social aspects of how neologisms get accepted in society, rather than the cognitive aspects, the different models will not be elucidated in detail. A detailed description of the manifold models – from the Full-Listing-Hypothesis (Butterworth 1983) via the compromise Augmented-Addressed-Morphology-Model (Caramazza et al. 1985; 1988) all the way to the Single Direct-Access-Modell (cf. Marslen-Wilson et al. 1994; 1996) can be found in Seyboth, 2014.

Sheepish, for instance, will be stored in its literal meaning as well as with the meaning of ‘being embarrassed’. A way to prove chunking is by looking at the survival of obsolete elements in syntagmatic combinations. DeSmet points out that Old English *ræden*, meaning ‘condition’, is an extinct form, but it has survived as the suffix *-red* in present-day English words like *hatred* (De Smet, 2016: 86). It is argued that an innovation is more likely to succeed if it bears an analogy to an existing chunk or resembles an established form (De Smet, 2016: 87). However, it can also evoke the opposite effect. Similarities to well-entrenched constructions might hinder a certain change. Articles are, for instance, usually followed by nouns; a change in this kind of entrenched pattern seems very unlikely (De Smet, 2016: 87).

The assumption that people’s language representation is constantly updated from everyday life experiences, and that every encounter leaves a trace in the memory, are woven into usage-based models to determine the size and nature of mental units (Abbot-Smith & Tomasella, 2006; Bybee & McClelland, 2005). Previously, various ideas about the standard size for lexical units existed. From narrow syntax (cf. Aronoff, 1976: 94), to either morphemes or both morphemes and words (cf. Halle, 1973: 16), to the idea that lexical entries are rather big and include idioms or noun phrases and can thus come in all sorts of sizes (Jackendoff, 1997: 109) like *listemes*, which neither correspond to the morphological objects nor to syntactic atoms (Di Sciullo & Williams, 1987: 1).

Recently, two different types of models were found to depict the impact of string frequency (cf. more detailed section 3.2.3) on mental representation: a holistic and a syntagmatic model (cf. tbl. 3). Holistic models emphasise that every single usage event strengthens the memory trace for a complex and unanalysable string as a whole and thus it enhances the string’s relative prominence in the cognitive system, making it more easily accessible than its individual component parts (Blumenthal-Dramé, 2017: 130). In contrast to that, a “string that is only rarely encountered is assumed to be weakly represented as a whole and

therefore less readily activated than its component parts” (Blumenthal-Dramé, 2017: 130). The brackets in the holistic model in *table 3* depict mental units. The idea is that in cases of high frequency, a holistic storage overrides a syntagmatic storage, and in cases of low frequency vice versa. However, the opposite could also be possible. Children might already understand the sentence, *I don't know* but can only handle the separate parts later on (Blumenthal-Dramé, 2017: 132). Syntagmatic models highlight the sequential relations of strings, indicating that frequency in use increases the syntagmatic fusion between the morphemes of the string (Blumenthal-Dramé, 2017: 131).

	Holistic	Syntagmatic
High-frequency whole	[I don't know] [I] [don't] [know]	Idontknow
Low-frequency whole	[You won't swim] [You] [won't] [swim]	You won't swim

Table 3 Different ways to depict the impact of string frequency on mental representation (Blumenthal-Dramé, 2017: 131)

These two models are not too far apart from each other, they might even capture the same thing in a formally different way, as frequency is a major factor in both models. The main difference is that the holistic model allows an overlap in storage and thus, within the holistic memory, traces of different grain size and representation strength compete for priority in online processing (Blumenthal-Dramé, 2017: 130). In the syntagmatic system the relevant knowledge is available giving information about the relationship between memory traces and single representations (Blumenthal-Dramé 2017: 132).

Schmid argues that frequent opaque complex lexemes (e.g., *hotdog, butterfly*) have their own entry in the mental lexicon, whereas the constituents of frequent transparent lexemes (e.g., *birdhouse*), despite the lexemes also being entrenched and having their own entry in mental lexicon, have a better chance to reach a level of conscious processing (Schmid, 2008: 24). In contrast to this, rare transparent complex lexemes (e.g., *cartwheel*) are processed computationally, but once they get more frequent, they will start to be processed holistically. Rare opaque complex

lexemes (e.g., *boatswain*, *blackguard*) will have time-consuming race-horse type competition between holistic and computational processes (Schmid, 2008: 24 f.). While it can certainly be said that neologisms, in the beginning of their life, are low-frequency words, it is not clear whether they are stored holistically or syntagmatically once they start getting entrenched. However, it was found in several priming experiments, including visual lexical decision tasks and self-paced reading tasks, that neologisms are processed in a qualitatively different way after even a single exposure, implying that the neologism has started to build a mental representation and has been present in the participants' minds (de Vaan, Schreuder & Baayen, 2007).

The current consensus is that irregular forms mainly rely on symbolic, rather than syntagmatic associations. Thus, they link form and meaning like monomorphic words, while regular forms are syntagmatic and paradigmatic associations, with more frequent regular forms probably still being represented and processed holistically (Schmid, 2020: 242). Experiments have shown that high-frequency nouns, for instance, have independent representations for singular and plural forms (cf. Baayen, Van Casteren & Dijkstra, 1992). This finding is supported by eye movement tests that showed longer process and activation time spans for singular and plural forms, meaning that the forms are competing and thus each have their own lexical representation (cf. Baayen, Levelt & Haveman, 1993; Jackendoff, 1997: 123). This makes frequent word forms resist regularisation better than rare forms, as holistic syntagmatic associations are more often refreshed with frequent words. With compounds the pattern is similar, since it is assumed that frequent compounds are holistically accessed (cf. section 4.4). This means that the association machinery, as well as the way in which syntagmatic associations are strengthened, explain the process and representation of irregular and regular word forms.

Taking a step back, before words can be stored, they need to enter the speaker's mental lexicon, thus the question arises as to how this happens. Reiteration and thus frequency of exposure seems to play a

major role (Herbermann, 1981: 325ff.). Lexical decision task experiments (between words and non-words) have shown that participants usually react quicker to high-frequency words (Chialant & Caramazza 1995: 65–66). In reading tasks, readers spend more time on low-frequency words (Inhoff & Rayner 1986: 437) which even led to the word following a low-frequency word being read more slowly (Chaffin & Morris, 2001: 226). The question at hand is how frequently a word has to be encountered for it to acquire a representation in the mental lexicon?

Two opposing approaches exist. The so-called *fast mapping* suggests that a representation is already formed after the first encounter. However, it is not clear whether this also applies to morphologically complex words that are predictable from their constituents (de Vaan, Schreuder & Baayen, 2007: 3). It was shown empirically that, after having presented the nonsense word **cathedruke* again and again to participants, a competition to *cathedral* emerged already after a night's sleep; therefore 24 hours seem to be sufficient for a lexical representation to develop (Gaskell & Dumay, 2003: 108). Although it is unlikely that a full-fledged lexical entry already establishes after only one exposure, this approach assumes that even little exposure will lead to a word starting to have a representation in the speakers' mental lexicon.

In contrast to this, researchers who work with visual comprehension argue that no traces can be found instantly in the lexical memory, or only after a great amount of exposure (Alegre & Gordon, 1999: 41). This would imply that either only high-frequency words are stored, while inflection is processed by rules (Pinker & Ullman, 2002), or that most inflections are processed through memory (Eddington, 2004: 862). All in all, the mental lexicon is a very delicate construct that still needs to be researched further, as it is unclear what is stored, what sizes the stored items have, what role frequency plays in adding lexical representations to the mental lexicon, and whether different word types are represented distinctively.

Recapitulating, entrenchment is, amongst other things, based on statistical learning and memory consolidation. Lexical items are represented and processed in the speaker's mind through different

associations. Depending on the form of the lexical item, it can be accessed and processed either holistically or analytically. Frequency has a positive impact on this: the more often certain linguistic items are processed, the more the patterns of associations for processing are activated, the higher the entrenchment. Vice versa, lower frequency means lower entrenchment, so that, when a language is not used by speaker anymore, they forget words, expressions etc. (Steinkrauss, & Schmid, 2017).

3.2.3 Forces affecting entrenchment

Since it was shown previously that the more often a speaker has processed patterns of associations which control the production or understanding of an utterance, the more deeply they are entrenched in the speakers' associative network, it seems as if repetition and frequency are key factors for entrenchment (Bybee, 2006; 2010). Just like conventionalisation, entrenchment is a gradual process, implying that achieving item or unit status is a matter of degree, conditioned by constant repetition and thus frequency (De Smet, 2016: 75). Once frequent enough, a word loses its semantic specifications and pragmatic salience and therefore becomes less analysable and more entrenched (Haspelmath, 1999: 1055; Langacker, 2017: 42). A high degree of entrenchment thus goes along with chunk status (Croft & Cruse, 2004, 292; Langacker, 2008, 16, 21, 38). Due to their importance for entrenchment, frequency and chunking will be elaborated in the following paragraphs, nevertheless it should be kept in mind that neither the degrees of entrenchment, nor its existence can be measured with currently available means, so that only operational definitions that approximate the theoretical construct are available (Stefanowitsch & Flach, 2017: 121).

It is assumed that a structure exists, in the sense of neural processing, organised by recurring patterns, where some substrate features are hardwired, and others are not (Langacker, 2017: 40). These patterns must be restructured from previous activity, which is exactly

what entrenchment is about (Langacker, 2017: 40). The need for previous activity calls frequency into play. This aligns with the fact that “theorizing about language use tends to assume a tension among replication and creativity [with] entrenchment [...] [being located] on the replication side of usage” (De Smet, 2017: 95). Although it is argued that a lexical item or a unit might already leave detectable traces after one occurrence (cf. section 3.2.2) (cf. Tamminen & Gaskell, 2013; Van der Ven et al., 2015), repetition is assumed to deepen and accelerate its entrenchment (de Vaan, Schreuder & Baayen, 2007: 2; Schmid, 2008: 19). Frequency eases and speeds up lexical access and retrieval (Sandra, 1994: 30-31) until such a point that a unit is so entrenched that it constitutes ‘an event waiting to happen’ (Langacker, 2016: 41). Thus, entrenchment is being fostered by repetitions of cognitive events, i.e. by “cognitive occurrences of any degree of complexity, be it the firing of a single neuron or a massive happening of intricate structure and large-scale architecture” (Langacker, 1987: 100). It therefore depicts “the degree to which the formation and activation of a cognitive unit is routinized and automated” (Schmid, 2007: 119). Hence, its pattern is not only maintained, but its execution will become faster and easier, in other words automatic. Psychologically speaking, automatization indicates a thoroughly mastered routine that is executed without close monitoring (Hartsuiker & Moors, 2017: 201ff.). Nevertheless, it should not be forgotten that due to its dynamic nature, entrenchment can also be reversed and is not a stable condition (Schmid, 2020).

Over the last few years, a positive correlation between processing ease and frequency has also been increasingly supported by empirical neurolinguistic and psycholinguistic studies that examined effects of usage frequency on the processing of transparent multimorphemic strings (Blumenthal-Dramé, 2017: 133). Chaffin et al. (2001) conducted an eye movement experiment studying high-familiar, low-familiar and new words. Familiarity in such studies is usually operationalised by means of frequency of occurrence (Schmid, 2008: 12). The participants’ eye movements were tracked as they read a sentence that either contained a high-familiar, low-familiar, or novel word (Joe picked up the

(*guitar/zither/asdor*). The word was then put in context in the second part of the sentence (and *begun to strum a tune*) and a definitional association was made by using a superordinate term in a second sentence (He played the *instrument* to relax) (Chaffin et al., 2001: 226). Using fixation duration, gaze duration, first pass reading time and spill over in order to obtain information, the results showed that, with low-familiarity and new words, the initial processing time was much higher than with high-familiarity words (Chaffin et al., 2001: 229).

A possible interpretation of this is that low-familiarity and new words are less entrenched than high-frequency words. Only once readers received more information (in the 2. sentence), differences between low-familiar and new words became apparent. With new words, participants spend more total processing time in the informative context (in the 2. sentence) and referred more often back to the context region (in sentence one) than with low- and high-familiar words (Chaffin et al., 2001: 229). Thus, this shows that the processing of context with new words is more effortful (Chaffin et al., 2001: 229) but at the same time new words need more contextual information than existing words, entailing that the amount of contextual information can ease and speed up the comprehension of novel formations (Schmid, 2008: 13).

Furthermore, participants spend more total reading time on new words, probably due to the lack of familiarity. Similarly, low-familiar words were re-read more often than high-familiar words and were processed slower (Chaffin et al., 2001: 229). The speed of access as well as the retrieval from the mental lexicon show a certain routinisation which supports the idea that frequency and entrenchment co-vary (Schmid, 2010: 116). Therefore, it is common practice to regard frequency as a significant factor in entrenchment (Schmid, 2010: 101).

However, frequency is not that straightforward as it cannot be seen in isolation. It was shown empirically that the speed of activation for the word *nun* profits from the high-frequency homophone *none* (Caramazza et al., 2001). Thus, frequency might not be word specific but cumulative for all homophonic (however, not homographic) forms. Furthermore, *family size* can play a role. Bertram, Baayen and Schreuder

(2000) found that with well-established formations response latencies in visual lexical decision decrease with increasing family size. However, this effect could not be confirmed empirically for neologisms (de Vaan, Schreuder & Baayen, 2007: 24).

Furthermore, when investigating family size, a distinction between token frequency effect and type frequency effect was found. The token frequency effect affects individual words and, when high, ensures a word's establishment in memory, while a type frequency effect concerns simplexes and complex words in morphological relation to other words in the mental lexicon (Bertram, Baayen & Schreuder, 2000: 402). Thus, token frequency usually has a conserving effect since repetition strengthens the representation of linguistic forms in memory and makes them more accessible (cf. section 4.2).

Diachronically speaking there is a trend towards regularisation, such as giving verbs a regular *-ed* past tense form. High-frequency words resist this regularisation process due to their high accessibility and the fact that they are accessed independently and are not as interconnected in the network (Bybee, 2007: 10, 14).²⁰ In contrast to that, type frequency goes along with reduction, like in *god be with you* becoming *goodbye*. This change in form shows that expressions are stored as units and, over time, get more efficient through an increased overlap and reduction (Bybee, 2007: 11). While it seems that reduction and conservation are contradicting, as one favours change and the other one blocks it, they cause changes on different levels: the strengthening of memory makes complex units resist change by reformation or analogy, whereas greater fluency and reduction of repeated units is a phonetic and semantic change (Bybee, 2007: 13).

Hence, the family size of a word shows that frequency cannot be seen as an absolute, but rather a relative which does not only depend on

²⁰ This conservation effect can also be observed on a morphological level. When comparing *I know nothing about it* and *I don't know anything about it* (Bybee, 2007: 10), it was found that the first construction, which is the older and more conservative form, was preferred when the two constructions could be used interchangeably (Tottie, 1991).

the base form, but also on the frequency of related inflectional forms as well as the size of the associated word-family (i.e. the set of complex words with the same base) (Nagy et al., 1989: 267; Baayen et al. 1997: 865; Bauer, 2001: 102).

Nevertheless, frequency is often regarded as an absolute, as it seems obvious to assume that comparing two features with differing occurrences in texts of the same length, provides reliable information about frequency.²¹ However, if the two features have nothing in common (e.g., meaning, functional load, etc.), the retrieved frequency information gives little indication about the actual frequency (Hoffmann, 2004: 190). “[F]requency information for an individual linguistic item only becomes meaningful as a diagnostic tool if it is compared with the frequency of occurrence of related linguistic phenomena” (Hoffmann, 2004: 190). Thus, a frequency-based analysis, in the sense of lexical/textual frequency, does not only have to consider how often an item is found, but also in how many instances it could have occurred but did not, since the underlying concept was expressed by using a different item (Hoffmann, 2004: 190).

While counting the absolute frequency can serve as an indicator of the strength of symbolic associations of fixed and variable forms, counting the relative frequency can serve as an indicator of the strength of syntagmatic associations between parts of strings. Thus, although frequency cannot be accounted for in a vacuum, absolute frequency should not be ruled out completely. It is assumed that co-text free entrenchment does exist, as a high-frequency word like *time* is probably more entrenched than a low-frequency word, irrespective of its environment (Schmid, 2010: 120). At the same time, some words are entrenched in a co-text dependent way. The word *fact* for instance, is highly entrenched, but Schmid found that in 26,106 out of 68,472 investigated cases it occurs in combination with a *that*-clause and hence

²¹ Besides absolute and relative frequency, Hoffman also designates conceptual frequency. This type of frequency is quite hard to grasp, as it would require, that all paradigmatic competitors of a word are known in regard to their function and meaning. This might be possible for the lexicon but not for lexico-grammatical constructions (Hoffmann, 2004: 190)

shows high co-textual entrenchment (Schmid, 2010: 121). This demonstrates the tendency of a linguistic element to trigger or activate others (Schmid, 2010: 120). A further example is *kith*. On its own the item is hardly entrenched, however, in the combination of *kith and kin* it is (Schmid, 2010: 122).

Phrasal verbs also depict an example for co-text dependent entrenchment as they are usually entrenched as such, like *to get up*. This stands in contrast to absolute entrenched words which are not entrenched in one specific co-text, but in many different ones. In these cases, co-text dependent entrenchment is overridden by co-text free entrenchment like in the case of *way*, which can be entrenched in various constellations (*the only way, in such a way*) (Schmid, 2010: 121). This shows that the relation between frequency and entrenchment is still hard to grasp, also partially because the relation between absolute and relative frequency and co-text free as well as co-textual entrenchment is very complex. This problem is aggravated by the fact that it also depends on what researchers count as valid tokens and how frequency is measured, even when absolute and relative frequency are distinguished (Schmid, 2010: 125-126).

The fact that mainly corpus data, and thus print media, is used to assess frequency is also a problem, as it does not necessarily correspond to real life language usage. Frequently occurring words in corpora are usually in more prominent places than in the actual discourse and in the language system (Schmid, 2010: 102). Corpus data tends to work really well in terms of determining overall usage frequency but it fails to “capture more subtle determinants of usage intensity” (Stefanowitsch & Flach, 2017: 120) which might weigh disproportionately more than the actual frequency (Stefanowitsch & Flach, 2017: 121).²² This entails that frequency in corpora usually gives an insight into conventionalisation rather than entrenchment, and consequently into the

²² More detailed explanation on how one could correlate frequency of occurrence with salience of entrenchment in the cognitive system can be found in Stefanowitsch & Flach, 2017: 102 ff.

social rather than the cognitive side (Stefanowitsch & Flach, 2017: 117).²³ In order to translate the degree of conventionalisation directly into the degree of entrenchment, it needs to be assumed that the frequency of use and exposure reflects the degree of conventionalisation in a speech community and that this enhances entrenchment in the minds' of the individuals (Stefanowitsch & Flach, 2017: 117).

Furthermore, research often makes use of corpora in the sense of 'corpus as output view' (Stefanowitsch & Flach, 2017: 102-103) rather than 'corpus as input view' (Stefanowitsch & Flach, 2017: 103-104) which focuses on usage frequency as a force affecting entrenchment. Therefore, corpus as output only describes a potential state of entrenchment, but since entrenchment is individual, it would be necessary to get the individual speaker's data analysed and compare it to the corpus data (Schmid, 2020: 217). As entrenchment is relative, and individual entrenchment depends on social processes and effects, "the entrenchment processes are initially triggered by more or less conventionalized utterance types that a speaker produces and is confronted with in co-semiotic events in usage" (Schmid, 2020: 205). Moreover, low frequency can be overruled by forces, such as high salience or strong pragmatic associations (Schmid, 2020: 217). This implies that there is a tension between the psychological basis of entrenchment and the social nature of language change (De Smet, 2017: 77).

In sum, the fact that corpus data is not sufficient, that there are different types of frequencies as well as that frequency is not a primary

²³ Schmid considers it vital to keep conventionalisation, being a social factor and entrenchment, being a cognitive factor, separate from each other (Schmid, 2008: 21). While they do intertwine (cf. fig. 4), they are governed by different kinds of structures and processes. Social aspects are governed by motivation, accommodation, diffusion and normation in the social system, while cognitive factors are governed by associations, chunking, automatization, generalisation and categorisation (Stefanowitsch & Flach, 2017: 127). Less clear cut is Langacker's distinction, who regards entrenchment as a general, not language specific phenomenon taking place in any kind of learnt human activity, which therefore depicts the individual counterpart to conventionalisation, which is a social process (Langacker, 2017: 39).

force for entrenchment, but rather driven by others (such as the communicative goals, emotive, pragmatic, and social factors) prevents a satisfying conclusion being reached on the relationship between frequency and entrenchment. Therefore, despite claiming that frequency in processing and occurrence in discourse correlate with the strength of entrenchment, there is still an overall lack of knowledgeability, especially about co-text and co-text free entrenchment (Schmid, 2010: 126).

Besides frequency of repetition, other forces that affect entrenchment are self-priming, similarity or analogy and salience.²⁴ Self-priming can be regarded as a factor that co-controls frequency. Similar to what happens in co-adaptation between two speakers, speakers also sometimes repeat themselves, by means of *self-priming*, *intra-speaker priming*, or *self-alignment* (Barlow, 2013; Gries, 2005; Günther, 2016; Szmrecsanyi, 2005; 2006). However, as with co-adaptation, it is not quite clear whether self-priming leads to entrenchment (cf. section 3.1.1). The EC-model suggests that repeating oneself will routinise linguistic habits and it might be assumed that repetition is supported by and is a symptom of routinised patterns of associations in the individual speaker's associative network (Schmid, 2020: 219).

Repetition is linked to the idea that the speaker will recognise certain similarities. These can range from *perceptual similarity*, *partial perceptual similarity* to *analogy* (Schmid, 2020: 219 ff.). The first implies that a *type* can share the same physical form such as word forms like *runs* and *drinking* and are treated as *token repetition* or *string repetition* and counted as *token frequency* or *string frequency*.

Partial perceptual similarity assumes that, for instance, the forms *laugh*, *laughs*, *laughing*, and *laughed* are recognised as word forms of the lexeme type *laugh*, based on the perceptual similarity of the grapheme sequence <laugh> and the phonetic signal [lɑ:f] common to all four examples. Thus, this is treated as *type repetition*. Partial perceptual

²⁴ Schmid also names embodiment and iconicity as forces for entrenchment. Due to their lack of relevance for the empirical part they are not discussed here, for more information see Schmid (2020: 221-223; 225-226)

similarity increases the recognition of *relational similarity* or *analogy* by means of structural alignment (Ambridge & Lieven, 2015; Behrens, 2017; Paul, 1920; Tomasello, 2003), so that different suffixes, such as *-ing*, *-ed*, *-s* are all considered analogous due to their partial perceptual similarity of the stem *laugh*.

The last type of repetition is based on structural alignment and analogy alone. Therefore, the sentences *Paul kissed Mary*, *Sam is hugging Simon*, and *Jane nudges Peter* can all be treated as tokens of one pattern, e.g., 'X V Y' or 'N V N'. In these examples perceptual similarity does not play a role anymore, they are based on relational similarity only. For this recognition to happen, semantic aspects need to be considered, since not every three-word sentence does evoke analogical reasoning based on relational similarity (Schmid, 2020: 220).

Perceptual similarities are usually considered to have a stabilising and conserving effect on token repetition, indicating that the more a speaker repeats the same routine, the stronger it will become routinised in the form of *token entrenchment* (Ziem & Lasch 2013: 104). By way of contrast, type repetition and analogy contribute to the extension and productivity of patterns and are regarded as being conducive to *generalisation* (Goldberg, 2006), *abstraction* (Langacker 1987: 132–7), *schematization* (Abbot-Smith & Tomasello, 2006; Ambridge & Lieven, 2015; Langacker 2008: 17), or *type entrenchment* (Ziem & Lasch 2013: 104).

Salience, similar to frequency, indicates the connection between usage and entrenchment within the entrenchment feedback-loop. An utterance that is highly salient due to its high entrenchment, which is a consequence of high frequency, is more likely to be reused by a speaker (Schmid, 2020: 221). However, salience is context dependent. It depends on the task and goals of the speakers, on what is expected in the context as well as on the individual speaker. Therefore, salience as such does not exist. There are only tendencies in regard to the general potential for different types of extralinguistic and linguistic experiences to attract attention, depending on context and speaker (Schmid, 2020: 224). A higher perceptive salience, however, means that speakers pay more

attention which is better for a successful co-semiosis (Clark, 1996: 81) and will increase entrenchment (Schmid, 2020: 223).

In sum, there are several different forces that affect entrenchment. One of the most debated ones is frequency of repetition. The more frequently similar utterances types are processed, the more likely it is that their representative pattern of associations will become routinised. However, other factors also impact routinisation and schematisation and thus entrenchment. While they are different forces, they eventually are all intertwined with frequency. A summarising statement, capturing entrenchment is that

“[e]very use of a structure has a positive impact on its degree of entrenchment, whereas extended periods of disuse have a negative impact. With repeated use, a novel structure becomes progressively entrenched, to the point of becoming a unit; moreover, units are variably entrenched depending on the frequency of their occurrence” (Langacker, 1987: 59).

3.2.4 Entrenchment of lexical items

While institutionalisation and lexicalisation were mentioned as lexis and neologism specific subprocesses of conventionalisation, this section will give a brief overview of these two processes and their connection to entrenchment. Schmid's (2008) 'three perspectives and three stages of the establishment of new words' and Kerremans' (2015) adapted version of it, summarise and structure the multifarious processes leading towards the establishment of a new word (cf. tbl. 4). While some aspects of it do not fit the dynamics of the EC-model, it nonetheless offers a good overview of what happens in these subprocesses of conventionalisation and entrenchment. The continuum from the first use to complete integration of a term is described by the stages of *creation*, *consolidation*, and *establishment*. The three perspectives of *lexicalisation* (structural perspective), *institutionalisation* (socio-pragmatic perspective) and *concept-formation* (cognitive perspective) highlight different aspects of these three phases (Schmid, 2008: 3). Lexicalisation, in the EC-model denoted as the usualisation subprocess of symbolisation, enables nonce-

formations to stabilise and finally become lexicalised items, possibly accompanied by the loss of motivation and an idiomatisation process (cf. section 3.1.3). *Table 4* indicates exactly this process, in which a new word starts off being highly ambiguous and context-dependant and only through consolidation a more stabilised form and meaning develops which will eventually become lexicalised.

Institutionalisation allows a new item to spread amongst the speakers of a society and thus affects the diffusion of a new word. Thus, while new words in their creation phase are often characterised by individual occurrences by speakers/writers, their consolidation implies that they start to diffuse into the speech community. This leads to the establishment of the word which means that the majority of speakers are familiar with the item (cf. tbl. 4).

Lastly, entrenchment anchors a new word in the individual's mind through hypostatisation, that gives a word a specific, hypostatised meaning. While there is no individual entry for the new word in the mental lexicon to start with, through consolidation a hypostatised concept gets developed and the entry that, at first, is only tentative and loosely connected to other entries will eventually become a distinct and firmly connected entity (cf. tbl. 4).

A shortcoming of this depiction is the fact that institutionalisation and conventionalisation are regarded as one process, rather than institutionalisation and lexicalisation being subprocesses of conventionalisation. Further, the fact that different factors are arranged horizontally evokes the idea that all of them happen at the same time. This, however, is not true as it is logically necessary for a word to diffuse first before it can develop a hypostatised meaning (Schmid, 2008: 3). The table also does not indicate the dynamics and does not show the instability all these processes are subject to. Nevertheless, it offers a concise overview of the lexis and neologism specific processes that become effective in conventionalisation and entrenchment of new words.

Perspectives Stages	Lexicalization	Institutionalization and conventionalization	Hypostatization and entrenchment
Creation	nonce-formation: semantic and formal ambiguity and context-dependence	nonce-formation: • individual occurrence by speaker/writer • type-familiar to hearer/reader	• pseudo-concept • no individual entry in the mental lexicon
Consolidation	stabilization of form and meaning	diffusion into the speech community	development of hypostatized concept with a tentative entry in the mental lexicon, loosely connected to other entries
Establishment	lexicalized lexeme (possible demotivation and idiomatization)	• institutionalized and conventionalized lexeme for the majority of the speech community • item-familiar	hypostatized holistic concept with distinct entry in the mental lexicon, firmly connected to other entries

Table 4 The three perspectives and three stages of the establishment of new words by Schmid (2008), adapted by Kerremans, 2015: 40

3.3 What's it all about: usage

Within the last few sections, it became apparent that entrenchment and conventionalisation are the two dominant processes which are needed in order to establish and maintain (lexical) items in general, but also (lexical) innovations. However, the central component that puts the conventionalisation and entrenchment flywheels into motion is usage (cf. fig. 4). By the speakers' use of language, and by the repetition of recurrent linguistic forms for communicative goals, and a repeated correlation between linguistic form and communicative goals, it sets and keeps the social process of conventionalisation (establishing and sustaining a convention) and the cognitive one of entrenchment (establishing and sustaining the linguistic knowledge) afoot (Schmid, 2020: 2). Usage, social and cognitive processes therefore constantly reinforce each other. Continuous language usage allows "shared conventions and individual knowledge to emerge, persist, change, and

embrace more or less creative innovations, both in language and in the external world” (Schmid, 2020: 2).

Usage takes place in concrete usage events, which are “actual utterances in their full phonetic detail and contextual understanding” (Langacker, 2000: 2). It comprises four aspects: the physically observable utterance itself, the participants’ communicative goals involved, the cognitive and interpersonal activity that is needed to produce an utterance and to establish co-semiosis between the speakers, as well as the linguistic, situational, and social context (Schmid, 2020: 15). According to Schmid, all four usage aspects can become conventionalised and entrenched. This implies that interpersonal activities such as co-semiosis, co-adaption, and accommodation, as well as cognitive associations influence use.

The major forces that affect usage are efficiency, extravagance, solidarity, and power. They partially resonate in Haspelmath’s adaption of Keller’s five maxims of action (Keller, 1994: 95-107, Haspelmath, 1999: 1055):

1. **Hypermaxim**: talk in such a way that you are socially successful, at the lowest possible cost.
2. **Clarity**: talk in such a way that you are understood.
3. **Economy**: talk in such a way that you do not expend superfluous energy.
4. **Conformity**: talk like the others talk.
5. **Extravagance**: talk in such a way that you are noticed.

(Haspelmath, 1999: 1055)

Efficiency follows Keller’s economy maxims of action, which states that you should not make an unnecessary effort (Keller, 2014: 140, 142). For entrenchment this implies that you should say what you always say in comparable circumstances and if the circumstances are too different from familiar circumstances, stay as closely as possible to what you would say under comparable circumstances (Schmid, 2020: 77). Thus, if speakers follow this maxim, they stick to entrenched routines.

Cognitive and articulatory economy can be comprised by the wish to communicate efficiently (cf. Zipf, 1949; Haspelmath, 1999; Labov,

1972; Goldberg, 2019). In some situations, economy might not be an efficient solution for the speaker. For instance, if a speaker is not precise enough, the decoding burden on the listener is increased, or if an utterance is entrenched for one speaker but not for another (Schmid, 2020: 77). Hence, in order to communicate successfully and enable an effortless co-semiosis, the speaker usually considers the hearer's needs and therefore follows the second maxim to increase their chances to reach the communicative aims (Keller, 2014: 135). Hence, while economy is important, co-semiosis should still be effortlessly and quickly accomplished (Schmid, 2020: 77) which makes efficiency a combination of the second and third maxim.

There are various aspects regarding new words that can facilitate co-semiosis, and thus make communication more efficient. As transparency will be of importance for the empirical part (cf. H1, section 1.2), the following paragraphs show how phonological, morphological, and semantic transparency can increase efficiency. The fact that speakers often prefer transparent forms can either have cognitive reasons – as the used patterns are already firmly entrenched schemas abstracted from language use (cf. Bybee, 2006) - or pragmatic ones, since the ultimate goal of a coiner is to be understood. Thus, both listener and coiners usually prefer transparent forms (Schmid, 2008: 15). Although forms outside of these patterns are less likely to succeed, they are not unheard of (Bauer, 2001: 62–71), as the case of *bouncebackability* showed (cf. section 4.2). However, if a new lexical item is more phonologically, semantically, and morphologically transparent, it stands better chances to be used. Vice versa this implies that language can impose certain restrictions to whether the speakers' creativity and productivity will lead to usage (Kjellmer, 2000: 206).

To achieve phonological transparency as well as increasing the words' success chances, parallels and analogies are often exploited. A word like **thimp* (cf. thump) stands a better chance to survive than **thmip* (Kjellmer, 2000: 209). Parallel constructions also bear the advantage that they are usually easier to pronounce in contrast to rare sound

combinations (Kjellmer, 2000: 209). Along with this goes that pronunciation and spelling of words should agree according to established conventions (Kjellmer, 2000: 219). This goes hand in hand with the fact that established conventions are more entrenched. Adding *-ly* as an ending to adjectives that already end in *-ly* (e.g. **friendliely*) is a rather poor candidate for success, since it is not a conventionalised and entrenched sequence of sounds, despite the fact that morphologically the use of *-ly* makes sense. The same applies for cacophony such as in **wordlily* (Aronoff & Fudeman, 2005: 216).

When speakers encounter a new word and cannot find it upon checking their mental lexicon, they start looking at the root, stem, affixes etc. (Schmid, 2008: 14). Thus, morphological or, in other words, formal transparency can also be reached by making use of existing parallels. In the case of compounds or blends, it was found that people tend to decode a word more easily if they are either familiar with the parts of the word or a certain pattern (*elephant food* is formed in analogy to *animal + food*) (Kerremans, 2015: 56). Thus, the frequency of a word formation pattern can have a positive impact on how well a new word is accepted. This has been confirmed in various psycholinguistic experiments (Demske, 2006: 75). In a questionnaire-based experiment, Lehrer hypothesised that formally more transparent blend words will get better rankings (Lehrer, 1996: 366). In order to investigate her hypothesis, she gave participants fifteen real blends and one made up word. The participants were given two tasks: they were supposed to identify the target words (the constituents of the blend) and to rate the words on a scale between one (good word) to five (bad word). The results showed that words were rated higher when participants were able to recognise more constituents (Lehrer, 1996: 382).

Analogies can also be exploited in word formation processes involving affixes, so that **pensivity* as a form of *pensive*, following the pattern of *active – activity*, *creative – creativity*, *native – nativity*, would be easily understood by speakers (Kjellmer, 2000: 210). Affixes in general depict a common way to create new words:

"reliable comparative statistics are not yet available, but there does seem to have been a trend towards the increased use of affixes as a means of word-formation in English in the last decade or so. The trend looks set to continue."
(Crystal, 1995: 133).

However, not all affixes have a strong creative potential. The Old English *-th* ending (e.g., *warmth*, *length*, *depth*), for instance, is hardly ever used these days to create new words. *-Ness* on the contrary is a highly productive contemporary affix that is used with thousands of existing lexemes and will probably be used to form many more in the future (Crystal, 1995: 128). Therefore, amongst two potential words that only differ regarding their affix, the one with the most productive affix has a higher chance to get conventionalised, entrenched and ultimately survive (Kjellmer, 2000: 212).

Another promoting factor that can increase a word's usage is the etymological compatibility between stem and derivational affix like *writer/kingly* vs *actor/royal* (Kjellmer, 2000: 210-212). Mixing different etymological stems and affixes is impossible in some languages, like German: **Sterbation* [starvation] (Bauer, 1998: 410). New words can also be prevented from coming into existence by syntagmatic or paradigmatic blocking. Syntagmatic blocking refers to an existing form blocking the development of others. The noun *glory* for the adjective *glorious* blocks **gloriosity*, while in the case of *curious* nothing blocks *curiosity*, as no other competing noun exists (Scalise & Guevara, 2005: 164). Paradigmatic blocking prevents the attachment of rival affixes to the same base, **occurationor* and **occurrence* do not exist because of *occurrence* (Scalise & Guevara, 2005: 164).

Semantically transparent forms are often shaped by parallels to existing forms. The easier a word can be comprehended, the more successful it might be when competing for lexicalisation (Kjellmer, 2000: 209). This includes a familiarity with the constituent morphemes (especially the first one), family size of the constituent morpheme as well as familiarity with the semantic relation between the constituents

(Schmid, 2008: 13). Once more analogy proves to have a positive impact on transparency so that **unstaple* is clear enough because it is based on words like *unbutton* or *unlock* (Kjellmer, 2000: 209).

Words can also be blocked semantically. In Italian, for instance, *grande* only works for things that are smaller than humans or manmade: **lago grande, peitra grande, canale grande* (Ettinger, 1974: 389-39). ‘Enemies’ of semantic transparency are the involvement of specialist knowledge in a relevant field to understand a word as well as ambiguity (Kjellmer, 2000: 209). Nevertheless, there are always exceptions to the rules so that sometimes even highly ambiguous words succeed. Hearing the word *detweet* without context offers several possible interpretations. Nevertheless, the word succeeded and by gradually diffusing eventually lost its ambiguity (Kerremans, 2015: 178 ff.).

While all these factors are important for promoting the usage of a new word, some are less essential than others (Kjellmer, 2000: 219-220). Phonological parallels are vital, since if the phonological pattern is unknown to most speakers, the word would struggle to find its spot in a language. In contrast to this, easy pronunciation is not as essential, as loanwords (e.g., *bibliophagic*) show (Kjellmer, 2000: 217). Words that do not follow morphological principles do not really stand a chance, while productive affixes are helpful but not required (Kjellmer, 2000: 218). None of these semantic factors are really needed, however, generally the more of the above criteria a new word fulfils, the greater the positive effect on its success and vice versa (Kjellmer, 2000: 220). The reason being that a word’s potential is relative, just like its realisation and existence are relative (Kjellmer, 2000: 206).

In sum, transparency has a positive impact on how efficient a word is perceived as. With the previously mentioned types of transparency in mind, it seems that the more efficient, precise, and transparent a word, the more likely it is to be conventionalised and entrenched and thus, with these two feedback-loops in motion, also usage will increase.

In contrast to efficiency, the force of extravagance causes the opposite effect as it implies that a speaker usually does not make use of

conventionalised and entrenched utterances (Haspelmath, 1999). Extravagance depicts the more or less conscious attitude of a speaker focusing on expressivity, talking in such a way that people notice you and in an amusing and funny way (Keller, 2014: 139). Humour or puns can also help a word to get noticed and be remembered (Kjellmer, 2000: 215). Indeed, many neologisms, especially when used in advertisements and newspapers, try to be humorous. However, the positive impact of humour is often debated. Algeo argues that joke terms, such as *Alaskaphobia* ('a fear by a Texan of something bigger than Texas') are rather unsuccessful (Algeo, 1993: 289). Metcalf agrees that humour can hinder new words to be permanently added to the vocabulary, as they might be too funny to survive (Metcalf, 2002: 129, 144).

In the FUDGE scale (cf. section 1.2), U stands for unobtrusiveness, indicating that the lower the obtrusiveness, the higher the word's chances for survival (Metcalf, 2002: 155-157). Metcalf recommends for the creation of a new word to camouflage it, smuggle it into the language and talk it up (Metcalf, 2002: 185). However, as so often, there are exceptions to the rule, such as in the case of *couch potato*. Although the expression started as a joke, it soon became established (Metcalf, 2002: 130). Other counter examples are Carroll's funny sounding nonsense words *chortle*, *galumph* as well as the umbrella term for such words themselves: *portmanteau*. Although these nonsense words did survive, most of Carroll's innovations had a relatively low success rate (Metcalf, 2002: 33). Despite giving these examples, Metcalf is convinced that humour is not a promising factor, and words have a better chance if they do not stick out too much (Metcalf, 2002: 27).

There is also a link between extravagance and the social status of the speaker, which means that extravagance can be associated with Keller's hypermaxim (Keller, 2014: 143). This links extravagance to the notion of foregrounding and salience (Günther et al., 2017), implying that an intended violation of expectations is present, either by using non entrenched and conventionalised utterances such as neologisms – which are salient due to their novelty (Schmid & Günther, 2016) – or by using

utterances that are already entrenched and conventionalised but are used in unexpected contexts, leading to 'salience by surprisal' (Schmid, 2020: 78). It is also possible that an utterance type is generally catchy in the sense of swear words, interjections etc. However, how salient an utterance is, always depends on the linguistic experience and social background of the speaker and hearer alike (Schmid, 2020: 79).

The third force is solidarity, which aligns with co-adaptation and the accommodation theory (Giles et al., 1991). Solidarity works in two ways, either by creating solidarity to show the speaker's affiliation to a group, or by creating distance, if the speaker is outside of the group. Closeness is reached by talking in such a way that you show that you are a group member (Keller, 2014: 137) and talking in such a way the others who are present do (Keller, 2014: 138). Distance, in contrast, is created by talking in such a way that you show you are not a group member (Keller, 2014: 139) and by not talking like the others who are present do (Keller, 2014: 139).

Power can also be a force affecting usage. Based on Fairclough (2001) and Bourdieu (1991) it can be said that you should talk in such a way that you are able to affect the state of the social world according to your goals, to manipulate people, to assert your authority, and to keep up your social status. When doing so, you should try to exploit available institutional (e.g. media) and ideological power (political, societal, etc.) to reach your aims. While the former can be referred to as *individual or local power*, the latter invokes the institutional aspects revolving around language and power. Closely linked to this are processes of standardisation and the routinisation of pragmatic associations such as prestigious pronunciation variants, grammatical aspects associated with the discourse of science, academia, etc. (Schmid, 2020: 81).

Although usage is subject to the forces mentioned, it is also the motor for conventionalisation and entrenchment.²⁵ Therefore, concrete usage events themselves can be the source for change. Repeated usage

²⁵ A summary of all the forces that affect usage, conventionalisation and entrenchment can be found in appendix 5.

events drive conventionalisation and entrenchment processes that control change, and specific innovative usage events can trigger change (Schmid, 2020: 310). Innovations depict partially licensed utterance types on a gradient scale from complete novelty (amongst which borrowings are the most frequent ones, which can be innovative on all six dimensions of conformity), salient changes (using native sources that deviate from conventionalised utterance types such as blends) and non-salient innovations (e.g. new words derived from productive word formation patterns). However, due to co-semiosis, innovations are not only acts performed by speakers but might be seen as an act of the recipient who regards a conventionalised utterance type as unconventional. A mismatch between the speaker's intention and the hearer's interpretation is another source for innovation.

It is also possible that the repetition of fully licensed conventional utterances can trigger change, for instance by changes in the frequencies of repetition on the collective or the individual level (Schmid, 2020: 311). Thus, language change can be triggered by different phenomena. The most important one for this thesis is newly coined words. Other phenomena include utterances that are salient modifications of conventionalised utterance types, utterances that entail a small, unobtrusive modification of conventionalised utterance types as well as changes in collective or individual relative frequencies of repetitions of conventionalised utterance types (Schmid, 2020: 313).

Thus, as indicated in figure 4, while usage puts the flywheels in motion, co-semiosis contributes to the evolution but also persistence of conventionalised lexical items that form the basis for a mutual understanding. This understanding is subject to the individual being able to identify the form that has been used and therefore directly links the entrenchment feedback loop to usage and conventionalisation. The entrenchment feedback-loop is set into motion by repeatedly using and activating the four patterns of associations. The more these patterns are activated, the more entrenched they become. Thus, usage, entrenchment and conventionalisation are all linked through the self-referential feedback-loops that have the potential for licensing on the

conventionalisation side and an ease of activation on the entrenchment side (Schmid, 2020: 5). Consequently, the more frequently an utterance type occurs, the more it is subject to co-semiosis in a concrete usage event, which implies it will be more conventionalised, which increases its licensing potential. This again increases the likeliness of this utterance type being used, leading to repetition. At the same time, the pattern activation in the entrenchment feedback loop gets strengthened through repetition, the utterance type becomes more entrenched by routinisation and schematisation and accordingly a more entrenched utterance type is more likely to be activated in a concrete usage event and thus will 'win' the competition for activation with other patterns of associations (Schmid, 2020: 74). Hence, frequency is the cause and effect for conventionalisation and entrenchment (Schmid, 2020: 75).

Usage, as the centre of the model, depicts the main source for the development of linguistic conventionalisation and knowledge. While the conventionalisation and entrenchment feedback loops are self-referential, as well as connected with each other, there is also another level: the connection between the two feedback loops and usage. When taking obsolete words as an example, it becomes clear that they are regarded as less and less conventionalised for achieving a communicative goal, which goes hand in hand with a loss regarding their cognitive strength in the speakers' minds. Therefore, for utterance types to survive, the mechanism always has to be up and running. Nevertheless, every speaker uses this machine and mechanism in a different way which leads then to a source for variation (Schmid, 2020:7).

In sum, usage events are the only place where the collective and individual processes meet and affect each other. Recurring usage events therefore lead to an increased conventionalisation in society as a higher entrenchment in the speakers' minds (Schmid, 2020: 3). Several forces impact use and usage events. Despite the fact that the EC-model is highly dynamic, these forces are relatively stable, socio-pragmatic and emotive principles. Concrete manifestations that are carved by the different forces are subject to change over time, so that extravagant words might become conventionalised (Schmid, 2020: 83). Usage

events, however, are not only influenced by the two feedback-loops as well as external forces, but they also serve as a place and engine for change. The usage-driven activity of the machine can therefore be interpreted as corresponding to the observation that all living languages are subject to continual change (Paul 1920: 32; Schmid, 2020: 7).

After this section has introduced the most important processes that can lead to a new word's usage, as well as usage itself, the next section will focus on the two word-formation processes that will be relevant for the empirical part – namely blending and compounding.

4. Neologisms and the word-formation processes of blending and compounding

4.1 Blending

Neologisms can have all sorts of different linguistic structures, from compounds such as *blue space* ('any body of water or the area around it'), via affixations like *anti-fit* ('Anti-fit clothes are deliberately designed to fit the wearer's body very loosely') to blends like *smishing* ('an attempt to trick someone into giving personal information by text message').²⁶ The empirical part of this thesis, which investigates English neologisms, deals with N+N blends and although the underlying theory will not be a prominent factor, a brief overview of definition, structure and their mental processing will be given in order to ensure a general understanding of blends.

4.1.1 Defining blends

Despite the fact that blending is accounted to be amongst one of the fifteen methods of forming a new word (Simonini, 1966), it has been considered marginal, especially in comparison to the well-researched process of compounding (Lehrer, 2007: 115). Blends and compounds are often regarded as bearing similarities (Bauer, 1998; Lehrer, 1998; López Rúa, 2004), as both processes involve a subtype of fusion (a new word is made from two or more autonomous words), the so-called 'univerbation', followed by demotivation (Brinton & Traugott, 2005: 68). Hence, a blend is "a new lexeme formed by parts of two or more other lexemes" (Bauer, 1988: 238). Besides sharing properties with compounds, blends are also said to be close to derivation, as material is

²⁶ All three examples were taken from the Cambridge Dictionary Blog "About Words", (<https://dictionaryblog.cambridge.org/tag/neologisms/>) from the entries from the 30.03.2020, 30.07.2020 and 17.02.2020

deleted from one or both constituents (Plag, 2003: 13, 121 ff.). Therefore, blends are sometimes denoted as hybrid words (Katamba, 1994: 186). By deleting material, the morphological structure is sometimes neglected, as morphological boundaries are ignored (Kemmer, 2003: 75), which led to the assumption that blends are morphologically unanalysable and depict an unpredictable word formation process, that often does not allow a transparent analysis into morphs (Bauer, 1983: 234).²⁷ Instead of their morphology, their phonology seems to be more informative regarding their analysis (cf. section 4.2) (Kemmer, 2003:75).

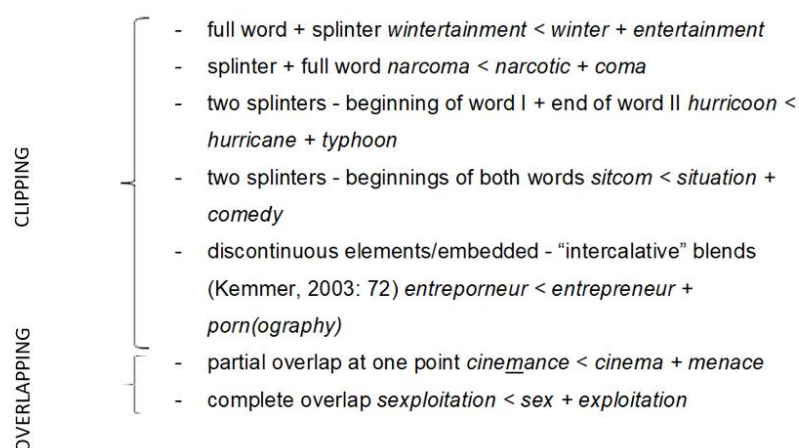
Although blends have long been marginalised and been regarded as random and unpredictable, “the process [of blending] is, of course, not at all new, and its very lack of subtlety probably accounts for its popularity; moreover, it is very easy to perform” (Pyles, 1952: 181). In fact, the first blends in English were detected as early as the 15th century, but most of them are nowadays obsolete. Some of the oldest current blends go back to the 19th century like *brunch* (1895) or *slanguage* (1870) (Merriam-Webster Online Dictionary). Despite being a long-established process, in recent years blends have become an increasingly important source for neologisms (Kemmer, 2003: 69). Along with this goes an increased linguistic interest (Lehrer, 2007: 115). One of the main reasons for their recent victory march is that, in the sheer number of stimuli who want to attract our attention (like media, adverts etc.), easily pronounceable and understandable as well as clever and creative novel words (Algeo, 1977: 48) will succeed (Lehrer, 2007: 116). Therefore, the main domains for blends are advertisements, newspapers, magazines,

²⁷ It should be noted that in morphological theory blends have been, so far, regarded as a mere footnote and therefore have hardly been researched. The biggest issue morphology has to face with blends is that the normal building block theory is based on morphemes, words and syntactic phrases. In compounds, inflection and derivation, the source elements are usually recognised as morphemes with clear boundaries. This, however, does not apply to blends as phonology seems to overrule morphology. Due to this they are often regarded as unanalysable: “It is [...] extremely doubtful whether such words can be analysed into morphs, and thus whether they form a real part of morphology” (Bauer, 1988: 38). Thus, blends do create issues for the field of morphology and despite more recent attempts to solve them (e.g. Kemmer, 2003), a lot of research still has to be done in this area.

and social media, as these types of media seek attention and want a favourable response from the audience (Lehrer, 2007: 129 ff.). Thus, blending has already some time ago moved from being a sporadic word formation process to a process “that has apparently led to the coining of many common words” (Marchand, 1969: 367). Nowadays it is a "creative technique" (Ronneberger-Sibold, 2008) and one of the most productive mechanisms for neologisms (Algeo, 1977: 74, Bauer, 1994, 37-39, Crystal, 1995: 130, Lehrer, 2006: 590). Blends became a productive morphological device on their own right, in spoken and written language, and skyrocketed over the last decades (Kemmer, 2003: 70).

4.1.2 The structure of blends

Blends can take on different forms, depending on how many words are combined and how much material is deleted. Although blends have been regarded as unanalysable in the past (cf. section 4.1.1), several scholars suggest different approaches on how to categorise them. One way is to look at how they are composed, distinguishing blends with overlapping and clipping:



(Lehrer, 2007: 117-118) ²⁸

²⁸ Although this will not be detailed in this thesis, it should be mentioned that the categorisation can become even harder when taking into account that blends are not just limited to words but also phrases. *A new time low*, for instance, depicts a combination of *new low* and *all-time low*. Another type of these rather extensive blends are mixed metaphors such as *keep your nose to*

The first five types show how clipping is involved in blending (Algeo, 1977: 48). A lot of research has been conducted about what and how much is deleted.²⁹ To some extent this is limited by the need of the hearer to still recover what is missing in order to understand the speaker. The so-called 'recognition point' refers to a point in a word, up to which the majority of speakers recognise it with an 80% probability (Gries, 2006: 539-45). This problem can be avoided by adding metalinguistic explanations if the source words are too hard to detect (Lehrer, 2007: 116-117). Sometimes words can get clipped at their morpheme boundaries as in *dumbfound* < *dumb* + (*con*)*found* (Algeo, 1977: 51). Through this process, blends can give new meaning to morphemes and thus can generate new words. *Parachute* for instance serves as the basis for creations like *parakite* or *paraglide* (Algeo, 1977: 52). Moreover, they can even create new morphemes like in the case of *Brexit* (cf. section 1.2). Other examples are *marathon* forming the base for *marrython*, *talkathlon* (Algeo, 1977: 52) or *cappuccino* resulting in *mochaccino*, *frappuccino* (Lehrer, 2007: 123).³⁰

the wheel (Algeo, 1977: 48). Located at the opposite end of the spectrum, also acronyms are sometimes seen as a subclass of blends (Algeo, 1977: 50).

²⁹ With clipped blends comes a great debate about the predictability of how much material is deleted. Despite this being a very interesting discussion, it would go beyond the scope of this thesis and therefore only some further literature will be presented here. While in the past predictability according to the prosodic structure was regarded as non-existent (Bauer 1983: 225; Cannon, 1986: 744), the perception of this has changed today and based on studies (however mainly on lexicalised existing blends), it is now possible to formulate generalisations in regard to prosodic restrictions (Bat-El & Cohen, 2012; Bauer, 2012a). Vital in this discussion is the position of the so-called *switchpoint*. Many different ideas about the position of the switchpoint can be found in literature (cf. Bauer, 2012a; Gries, 2006, 2012; Kelly, 1998). A first systematic overview of non-lexicalised blends and their switchpoint as well as stress pattern was conducted by Arndt-Lappe & Plag (2013). Other scholars also have had a close look at the stress pattern of blends and the role it plays in facilitating recognition (cf. Cannon, 1986; Bat-El, 1996; Fischer, 1998; Bauer, 2012a).

³⁰ It was found that blends are mainly disyllabic and trisyllabic. Longer constructions are usually avoided and the longer the base word, the more syllables get lost. Usually, the maximal length of a newly coined blend is never longer than its base word (Plag & Arndt-Lappe, 2013: 545-546).

Clipping at the morpheme boundary can sometimes make the distinction between blend and compound quite fuzzy. *Scape* for instance, was originally taken from Dutch *landscape*, over time other words like *cityscape* and *townscape* evolved, and *scape* soon was used as a word on its own. Therefore, more recent inventions such as *moonscape* are not considered as a blend anymore but rather as a compound. A similar example is *burger*, which evolved from being part of a blend to now forming new compounds (Algeo, 1977: 51-52).

Besides clipping, the last two examples above show that overlapping is also regarded as a major process in blending, which has increased in frequency over the last decades (Lehrer, 2007: 127). Apart from the distinction between partial overlap – in a single segment like *dumbsizing* – and complete overlap - with strings larger than a syllable like *glitterati* – the ‘spot’ of where the overlap takes place can vary, too (Kemmer, 2003: 73). This includes overlapping of

- hind part of word I + fore part of word II *film**ania*** < *film* + *mania*
- a discontinuous segment of one form with the fore and hind of the other word *can**imal*** < *camel* + *animal*
- a complete word + fore or hind part of other word, only marked by spelling *sin**ema*** 'adult film' < *sin* + *cinema*
- sandwich words (Wentworth, 1939) *autob**by**dography* < *autobiography* + *by dog*

(Algeo, 1977: 48)

Another subcategory of overlapping, which structurally differs from other overlapping blends but belongs to the same class conceptually, are

Hence, the maximum length of a blend, and thus the number of syllables, is determined by the longest base word (Bat-El, 2006:67; Bauer, 2012a: 14).

substitution blends (Kemmer, 2003: 73).³¹ They are characterised by the substitution of a part of one word with the whole part of another word like in *carjacking* < *car* + *hijacking* (Kemmer, 2003: 74). The various subcategories of overlapping show that it needs to be seen as a rather relative than a relative matter (Algeo, 1977: 54). A mixture of overlapping and clipping is also possible like in *Californicate* < *Californ(ia)* + *fornicate* and *motel* < *mot(or)* + *(h)otel* (Algeo, 1977: 52).³²

Blends can also be categorised by following the Saussurian dichotomy of syntagmatic and associate/paradigmatic relations (de Saussure, 1916: 123-126). *Syntagmatic blends*, or in other terms *telescope words* (Algeo, 1977: 57), are a combination of two forms that occur sequentially in speech. The most common variety amongst them is *haplology* – which corresponds to the above-mentioned overlapping of the hind part of the first word and the fore part of the second word (*Chicagorilla*) (Algeo, 1977: 56). *Associative blends*, or *portmanteau words*³³, can share a common base morpheme or affix or carry similarity in sound or meaning. Most of them bear a semantic link between the two constituents such as *needcessity* < *need*, *n(e)cessity* and are therefore also denoted as *synonymic blends*, as grammatically both words can stand in the same position and might be used instead of one another (Algeo 1977: 57). Additionally, there can also be morphological and phonological ties to reinforce the semantic connection or a sound similarity might lead to a semantic connection like in *buxom* 'bosomy' < *buxom* 'pliant,' *b(os)om* (Algeo 1977: 59).

Another type of blend that combines words of the same paradigmatic class are *dvanda blends*. In contrast to synonymic blends,

³¹ Due to the fact that it would go beyond the scope of this thesis, Kemmer's suggestion of schemas to analyse blends conceptually (Kemmer, 2003: 80 ff.) will not be discussed here.

³² An elaborated list of examples can be found in Algeo (1977: 52)

³³ Even though portmanteau word is also used as a general term for blends, it particularly fits this category as an opposite to telescope blends (Algeo 1977: 61)

the components of dvanda blends are not synonymous like in *smog* < *sm(oke) + (f)og* (Algeo 1977: 59). As in both, synonymic and dvanda blends, the two source words are from the same paradigmatic class, they are often referred to as *paradigmatic blends*. There is, to a certain extent, also a mixture of syntagmatic and paradigmatic blends. These so-called *jumble blends* contain semantic associations, but the components cannot substitute each other paradigmatically. For instance, *foodaholic* cannot replace *alcoholic* paradigmatically, however, it can be explained as a combination of two words associated with one another, since they are collocatable: 'He overindulges in food as an alcoholic does in alcohol' (Algeo, 1977: 58).

In sum, blends are more analysable than assumed in the past and linguists have suggested different ways to categories blends. However, one of the major problems is that, despite the fact that structurally some blends are alike, the system of their making as well as the psychological processes of forming them differ (Algeo, 1977: 62).

4.1.3 Identifying and processing blends

The increased interest in blends is accompanied by recent experiments that try to get a better understanding of how blends are identified by speakers and subsequently processed. Lehrer (1996) found that the following factors ease the identification of novel blends: a higher frequency of the source words, less neighbours³⁴, semantic priming, context, the number of present letters and/or syllables - probably connected to the 'recognition point' (cf. section 4.1.2) – as well as the identification of one part of a blend facilitating identifying the other(s) (Lehrer, 1996: 368 ff.). For the last point it was found that speakers who identified *jacket* in *swacket* < *sweater + jacket*, due to strategies of semantic plausibility, were usually able to guess *sweater* correctly

³⁴ When it comes to blends, neighbours are words that have the same letters in the same positions as the target word, like *psynergy* < *psychic + energy*, in which the splinter *-ergy* could also come from the neighbour *clergy* rather than *energy* (Lehrer, 2003: 371).

despite the fact that plenty of English words start with *sw* (Lehrer, 1996: 370). In *applicious*, the semantic context allows the recipient to decide that the second splinter is likely to derive from *delicious* rather than *vicious* (Lehrer, 2007: 126 ff.). The ease in identification follows the subsequent order (Lehrer, 2003: 371):

word + splinter > splinter + word > two splinters > complete overlap > embedded splinter

oildraulic (*oil* + *hydraulic*) > *narcoma* (*narcotic* + *coma*) > *sitcom* (*situation* + *comedy*) > *cattitude* (*cat* + *attitude*) > *entreporneur* (*entrepreneur* + *pornography*)

While all the above-mentioned factors contribute to correct identification, none of them are necessary or sufficient. The way blends are identified is in line with research on lexical retrieval, fostered by frequency, neighbourhood effects and semantic priming (cf. section 3.2.2) (Lehrer, 2003: 371). From that, Lehrer deduced that the factors of frequency, neighbouring effects and semantic plausibility will also quicken the processing time of novel blends. However, follow up experiments (in collaboration with Csaba Veres) including a lexical decision task and a priming experiment did not confirm this hypothesis (Lehrer, 2003: 379).

Therefore, the question at hand is still how blends are processed and whether they are processed differently to non-blend words. Unfortunately, not a lot of research has been done about this so far, however, an interesting preliminary study has been conducted by Juhasz et al. (2017), albeit only using lexicalised blends. By tracking eye-movement, they found that in lexical decision tasks, even relatively familiar blends were processed slower in contrast to (in length and rated familiarity) matched non-blend words (Juhasz et al, 2017: 287). However, in reading, when embedded in a non-predictable, but supportive sentence context, blends showed the opposite effect. Thus, they were read quicker than non-blend words (Juhasz et al, 2017: 290). Both experiments taken together indicate that lexicalised blend words are stored, accessed, and processed in a different manner than non-blend words (Juhasz et al, 2017: 290).

A follow up experiment (Johnson et al., 2019), which involved participants saying blend and non-blend words (that have been matched in word length, syllable number and frequency) as accurately as possible into a microphone as soon as they deciphered them, showed that naming time as well as accuracy for blends was usually longer/lower than for non-blends (Johnson et al., 2019: 850). This is in line with what has been found previously (cf. Juhasz et al., 2017). In a further experiment, in which participants were confronted with single line sentences, either containing a blend or a non-blend word (“Felicia watched the **tween/careful** girl look both ways before crossing the street”), participants had to read silently and once done with reading, press a button to move on (Johnson et al., 2019: 851). The outcome showed that fixation duration, gaze duration, single fix and the spill over were all longer for blends. This stands in contradiction to Juhasz et al. (2017), which might be explained by the fact that more context was given in their study. However, in general all these experiments show that it is very likely that blends are processed differently from non-blend words. This is supported by the findings that a blend, if it is perceived as such, is easier to decipher (Johnson et al., 2019: 855).

Thus, blends cannot be regarded as being similarly processed to other words and also the assumption that they might share characteristics with compounds, as both base words are co-activated when a blend is used (Kemmer, 2003: 70), bears a problem. If a blend like *beermare* < *beer* + *nightmare* is approached like a compound, it would lead the speaker to the assumption that it is ‘a horse that drinks beer’ (Lehrer, 1996: 381).³⁵ In summary, although a lot points in the direction of blends being processed differently to other words, research, especially into novel blends, is still in its early stages and the initial research presented above hopefully paves the way for future investigation.

³⁵ Despite the example, it should be mentioned that there is still evidence supporting the *dual-route processing* – as it is the case with compounds (cf. section 4.2.3) – in contrast to the *single unit processing*. For further reading see Häikiö, Bertram, & Hyönä, 2010; Kuperman et al., 2009; Pollatsek, Hyönä, & Bertram, 2000.

4.2 Compounding

As shown, compounds are often brought into context with blends, due to their assumed similarity in structure. While the main part of this thesis deals with blends, the study conducted in German focused on compounds. The reason for this is that blends are not widely used in German while compounds are extremely productive (Seyboth, 2014:1) and enable speakers to condense content areas to one word, which otherwise would have to be paraphrased (Ahrens, 1977: 74). Thus, German can be denoted as a “*kompositionsfreudige Sprache*” [a language that is generating many compounds] (Schlücker, 2012: 2). However, this mainly applies to nominal and adjectival compounds as it is, for instance, still debated whether verbal compounds even exist in German (Schlücker, 2012: 2). Since compounds are so productive in German, many neologisms are formed by compounding (cf. Jesenšek, 1995). Having decided to investigate German compound neologisms, a very brief introduction into the definition of compounds, their structural characteristics as well as how they are accessed and processed is given in the next sections. It should be noted that this brief overview cannot do justice to the extensive research that has been done in regard to compounds. However, since the word class is not of major interest for this thesis, giving a detailed description of the research done on compounds is not intended.

4.2.1 Defining compounds

The task to define compounds in German has a long history, going back as far as Jacob Grimm, who stated that compounds denote “das aneinanderfügen zweier deutlicher Wörter” [the combining of two distinct words] (Grimm, 1826: 383). This was followed by various definitions from linguists over the last century. Henzen for instance states that “Zusammensetzungen (Komposita) entstehen, wenn Elemente der Rede, die für sich als Wörter dienen können, zu einer neuen Worteinheit verbunden werden” [compounds come into being, when speech

elements that could serve as words on their own, are combined to a new lexical unit] (Henzen, 1965: 36). Another definition says that “im Kompositum [...] sind zwei (evtl. auch mehrere) [...] Wörter, die auch selbstständig in der Rede auftreten können, zu einem Bedeutungsganzen neuer Art verknüpft“ [in a compound two (or maybe more) words, which can occur independently in speech, are connected to a new word with a specific meaning] (Hempel, 1980: 152). Thus, compounds are words that are formed by combining two or more lexical items, thus words that consist of at least two free morphemes (Schlücker, 2012: 9). This process can include nouns, verbs, prepositions, and articles (Sauer-Egner & Reker, 2007: 1).

While this kind of definition seems obvious and although compounds, in contrast to blends, have been vastly researched in English and German, it is still difficult to give a single summarising definition. Some scholars argue that compounds are the connection of a minimum of two morphemes which can be free and bound (such as word formation affixes) rather than just free morphemes (Fleischer & Barz, 1995: 45). Furthermore, definition problems result from the fact that compounds touch various aspects of linguistics, such as phonology, morphology, syntax, semantics, lexicology etc. Especially the dichotomy between morphology and syntax is highly discussed. Some scholars talk about *compound words*, a term that already denotes that they are regarded as being part of morphology. This, however, ignores the fact that compounds are sequences of words which usually belong to syntax (Bauer, 2017: 3). Hence, depending on which criteria are applied, definitions can differ drastically (Czerwenka, 2007: 19). Based on this, some linguists have tried to distinguish between compounds that are the result of morphological and syntactical processes. Some further argue that there is distinction between root and synthetic compounds that can be taken from the basis of morphology versus syntax (Bauer 2012b: 134).

There are arguments in favour of both approaches. The fact that compounding is not that different from derivation supports the argument that compounds are allied more closely to morphology. Like derivation, compounding creates a new lexical item. Thus, a fascinating aspect of

compounds is the fact that they can evolve a meaning that cannot be derived from the single components anymore (Ungerer & Schmid, 1998: 77). Hence, they get lexicalised and are learnt by speakers and not analysed in regard to their internal structure. *Hedgehog* is a classic example of an English compound which native speakers learn as a whole, without analysing it as a 'pig which lives in hedges' (Bauer, 2012b: 134). Further, compounds, similar to derivatives, name entities, properties and actions, while syntax provides description. Thus, both words – the derivative *judo-ist* and the compound *judo-man* - denote a name for a person in contrast to the description a syntactic phrase such as 'an expert in judo' gives (Bauer, 2012b: 134).

However, there are also arguments that show how closely compounds are linked to syntax. They constitute a sequence of lexical items and usually such sequences are dealt with in syntax not in morphology. Furthermore, while it is true that many compounds are lexicalised and thus acquired as a whole and not analysed by speakers, especially compound neologisms and nonce-formations are usually, in the first stages of their lives, rather unlikely to be regarded as an unanalysable whole (Bauer, 2012b: 135). Thus, especially for novel compounds (cf. section 4.2.2) it can be argued that they are positioned closer to syntax than to morphology. Another argument that supports this notion is that (in English) the meaning of many N+N compounds is synonymous to an ADJ + N phrase (*atom bomb* – *atomic bomb*, *gold ring* – *golden ring*). Similarly, sometimes N+N compounds are equivalent to POSSESSIVE + N sequences (*birdfoot* – *bird's foot*, *student evaluations* – *students' evaluations*) (Bauer, 2012b: 136). In both cases the latter examples are considered as part of syntax, whereas the former, traditionally, belongs to the field of morphology. Bauer summarises:

“[T]here are [...] a good many reasons for seeing compounding as being more closely allied with syntax than with derivational morphology. [...] In current theories, compounding is nearly always dealt with as part of morphology. In either case, the point is made that the dividing line between morphology and syntax is a very fine one and not necessarily easily drawn.” (Bauer, 2012b: 137)

This summary exemplifies how difficult it is to find a straightforward definition for compounds. Trying to categorise compounds structurally is another approach which will be elaborated upon in the next section.

4.2.2 The structure of compounds

There are various structural and semantic categories of compounds that are widely agreed upon in linguistics, however, the terminology can differ. In German, three main types of compounds are distinguished: *Determinativkomposita* [determinative compound], *Possessivkomposita* [possessive compound] and *Kopulativkomposita* [copulative compound]. The first one depicts the most common type of compounds. Their structure is binary, and the constituent on the right-hand side of the word is seen as the semantic and grammatical head that determines word class, gender etc. of the compound. The constituent on the left-hand side of the word modifies the meaning of the head. Thus, these compounds follow the so-called Right-hand Head Rule (RHR), which states that “in morphology, we define the head of a morphologically complex word to be the righthand member of that word” (Williams, 1981: 248). Thus, the German words *Laubbaum*, *Nadelbaum*, *Obstbaum* all denote a type of tree with the head *Baum* [tree] and the modifiers *Laub*, *Nadel*, *Obst* [leaf, needle, fruit] indicating what type of tree they are (Schlücker, 2012: 5).³⁶

Possessive compounds (also called *Bahuvrihi compounds*) depict a small group of compounds that often refer to humans or animals. In contrast to *determinative compounds* the head does not denote the referent itself but a possessive relationship. Thus, a *Freigeist* [free spirit] is *eine Person mit einem freien Geist* [a person with a free spirit]. Therefore, this type of compounds is defined by a metonymic semantic interpretation (Schlücker, 2012: 5-6). Apart from this, however, they work similarly to determinative compounds, as they have an asymmetric modifier head structure. Therefore, they are often regarded as a specific

³⁶ For further reading on subcategories of determinative compounds cf. Schlücker, 2012.

semantic subclass of determinative compounds (cf. Fleischer & Barz, 1995: 46; Motsch, 2004: 376; Donalies, 2005: 59).

The last type, the so-called *copulative compounds* (also: *Koordinativkomposita* [coordinate compounds] or *Dvandva compounds*) consist of two or more coordinative constituents that usually belong to the same word class. This implies that the order of the two constituents is usually interchangeable, however, they are often lexicalised in a certain order (cf. section 3.1.3), for instance *schwarz-weiß* vs. **weiß-schwarz* [black-and-white vs. *white-and-black] (Schlücker, 2012: 6).³⁷

Another terminological way to describe compounds, which also points out the close relationship between possessive and determinative compounds, is Bloomfield's terminology of endocentric and exocentric compounds (Bloomfield, 1933) which are often used synonymously for determinative and possessive compounds (Schlücker, 2012: 6). Endocentric constructions include copulative compounds, which consist of two or more copulative components (Sauer-Egner & Reker, 2007: 4). Vice versa, with exocentric compounds the morphosyntactic properties or the semantic category of the whole word does not correlate with one of its constituents which entails that possessive compounds are often classified as exocentric (Schlücker, 2012: 6).³⁸

Furthermore, an aspect that is greatly discussed when it comes to German compounds and their form is the function of the *Fugenelement* [linking element].³⁹ These linking elements are usually used when either

³⁷ It should be noted that it was repeatedly shown that mostly the distinction between determinative und copulative compounds is not that clear cut in many cases (c.f. Breindl & Thurmair 1992; Donalies 1996, 2005).

³⁸ Another classification that can be found in the literature is between *Nichtrektionskomposita* [root compound] and *Rektionskomposita* [synthetic compounds]. For further reading see Lieber, 2004.

³⁹ Due to a lack in relevance, it should be briefly noted that the discussion about linking elements is dominated by questions such as the articulatory, prosodic, morphological or semantic functionality of the elements. For further reading see Fuhrhop, 2000; Aronoff & Fuhrhop 2002; Nübling & Szczepaniak, 2011; Krott et al., 2007; Becker, 1992; Krott, 2009).

a noun, or less often, a verb is the first component. However, not using a linking element is still the norm (Schlücker, 2012: 8). Apart from the *Nullfuge* [zero linking element], the most common linking elements in German, that can partially also be combined with an *Umlaut* (cf. Augst 1975, Ortner & Müller-Bollhagen 1991, Fuhrhop 1996, 1998), are:

+ -(e)s-	<i>Antrittsrede, Liebesbrief</i>
+ -(e)n-	<i>Bauernhof, Nervenfaser</i>
+ -er-	<i>Bilderrahmen</i>
+ -e-	<i>Tagebuch</i>
+ -ens-	<i>Schmerzensgeld</i>
+ e-Tilgung	<i>Schulranzen</i>

(Elsner & Huber, 1995)⁴⁰

In sum, when it comes to the structural and semantic criteria of compounds, there are various ways to categorise them. However, all these categories are not that clear cut and other factors, such as linking elements must be considered.

Since this thesis will deal with neological compounds, it is relevant to explain the difference between established, novel, and deictic compounds. Established compounds are accepted by the wider speech community; thus, they are understood by most speakers, are therefore conventionalised and lexicalised and can exhibit idiosyncrasy (Sauer-Egner & Reker, 2007: 3) (cf. section 3.1.3). When it comes to new compounds, there is a semantic distinction between deictic and novel compounds (Ryder, 1994). Deictic compounds denote compositional neologisms which are dependent on a non-verbal context and thus are “created to satisfy a fleeting discourse need” (Downing, 1977: 8). Novel compounds, in contrast, can be interpreted without reference to a specific context (Ryder, 1994: 9).

⁴⁰ The frequency with which these linking elements occur varies drastically. Some studies have shown that a zero linking element is most common with N+N compounds, followed by the linking element -s and -(e)n. All others are less common (cf. Wellmann et al., 1974; Kürschner, 2005; Krott et al., 2007).

Ryder names the example of the deictic compound *bike girl*, which was formed when a girl parked her bike inconsiderably in the hallway of a graduate department. While this compound is used in a certain context and only by a small group of speakers, it is nonetheless established within this group to denote ‘the kind of person who would inconsiderably leave a bike in the vestibule where everyone will trip over it’ (Ryder 1994:9). Zimmer argues that these compounds are “*naming devices*, which [...] denote ‘relevant categories’ of the speaker’s experience” (Zimmer, 1971: 9). According to Ryder there are two underlying processes in forming these compounds: “First, they will choose a noun to be the head noun [...] that describes a general class to which the new item could reasonably be assigned. Second, the element noun chosen as the modifier will have a relationship to the head noun that is relevant in a speaker’s cognitive organization” (Ryder 1994: 9). However, the distinction between deictic and novel compounds is not clear cut, since each deictic compound can also be denoted as a novel compound (Sauer-Egner & Reker, 2007: 3).

4.2.3 Accessing and processing compounds

Apart from the discussion about the semantic relationship between the components of a compound (c.f. Günther 1981; Heringer 1984; Fleischer & Barz 1995), it is also highly debated whether compounds should be treated as words, as syntactic constructions or as a mixture between both (c.f. Angele 1992). As a result of this, there are differing opinions about whether compounds are stored as whole units or in the form of morphemes, or even both. It is further debated whether access and production of compounds make use of the same processes, or whether completely different mechanisms take place (Seyboth, 2014: 1). Section 3.2.2 already mentioned the opposing model for accessing and processing complex lexemes and section 4.1.3 gave more details on how blends are identified and processed. Due to the close proximity of blends and compounds, it is sometimes argued that in processing blends “strategies for interpreting compounds will no doubt be utilized” (Lehrer,

1996: 376). However, despite the fact that compounds have been investigated much more profoundly than blends, a lot of inconsistencies can be found in literature.

Some argue that, in contrast to monomorphemic words, compounds are processed slower (cf. Inhoff, Brihl, & Schwartz, 1996) while others claim the opposite (cf., Drieghe et al., 2010; Fiorentino & Poeppel, 2007). The opposing views might be the outcome of the fact that transparent and opaque compounds seem to behave differently. Priming experiments showed that the meaning of transparent compounds is determined from the constituents' meanings. A compound like *milk bottle* can be primed with semantically related words, like *cow* and *flask* (Sandra, 1990: 556; Bauer, 2001: 110). However, opaque compounds (e.g. *buttercup*) and pseudo compounds (e.g. *boycott*) cannot be primed like this, which means that those words might have their own lexical and semantic representation, separate from the elements involved and are not split up into their constituents like transparent compounds are (Sandra, 1990: 543; Zwitserlood, 1994: 364). This agrees with Andrews' hypothesis that compounds are optionally accessed through their elements (in visual word recognition) (Andrews, 1986: 737). This all assumes, that - especially transparent compounds - can be easily split into their components while processing. This seems to be a logical conclusion, considering that compounds usually are the most early acquired words in Germanic languages. While this implies that the automated process of deciphering compounds should be independent, for instance, from the speakers' education level (Clark, 1993: 46), some studies found the opposite (Gleitman & Gleitman, 1979: 109). Furthermore, it was detected that some people struggle to find the head of a compound and, for example, regarded *quilt horse* as left-headed (Ryder, 1994: 1999). Thus, these findings pose doubt about how automatic and easily compounds are accessed. Therefore, it depends on the word itself. Nevertheless, many researchers agree that compounds are processed by decomposing them into two parts (Taft & Forster, 1975; Drieghe et al., 2010; Fiorentino & Poeppel, 2007; Juhasz et al., 2003).

In sum, compounds are a well-researched word class that, in contrast to blends, have attracted a lot of attention. Nevertheless, various aspects are still hotly debated, such as how they are accessed and processed, how the components relate to each other semantically and, particularly in German, what function can be attributed to the linking element.

5. Keeping up with the pace: neologisms on the internet– changes in methodology

The development of the internet and its current dominant position in our daily lives has led to many changes, including changes in linguistic methodology⁴¹, especially regarding neologisms. Since the empirical part of this study made great use of the internet as a source for retrieving neologisms as well as investigating them, it is important to explain how methodology has changed over the decades and what is the ‘new black’ these days.

Before the era of the internet, retrieving, collecting, and documenting neologisms was a long and slow process. Algeo states in his neologism dictionary *Fifty Years Amongst the New Words* that, as a first step in the process of creating this dictionary, contributors reported words they considered new. After this initial step, these words were counterchecked against reference dictionaries and their source material, if it was not available otherwise, was collected either as a xerographic copy or as handwritten or typed quotations. All these sheets, slips or clippings were added to the New Word files of the American Dialect Society and – already back then - computer records were created (Algeo, 1993: 3). While the collection process is still mainly manual nowadays, the internet changed the speed in which words are collected and documented. While lexical changes in the past have been recorded in ‘chunks’ of decades, the ‘turnover’ of new words became much quicker (Kerremans, 2015: 28). In fact, the Global Language Monitor (2009) estimates that every 98 minutes a new word is created (Paradowski & Jonak, 2012: 134).

In order to stay on top of all these new lexical innovations that the internet yields, researchers often make use of corpora. However, the fact that the sources of many corpora are often restricted to certain genres

⁴¹ Several statistical methods and tests were used to analyse the questionnaire studies. Detailed information is given in the method section of the studies (cf. section 6.1.3, section 7.2.3).

suggests that they might not depict lexical changes accurately, as the internet allows multifarious channels of rather informal information, like blogs and social media (Paradowski & Jonak, 2012: 134). Therefore, albeit not unproblematic, using the Web as a corpus – probably the best source to detect new words these days due to its up-to-dateness - is attractive for lexicographers and linguists alike (Gleick, 2006: 12). There are various homepages, run by dictionary publishers or based on crowdsourced user-content, which contain entries including a definition, selected quotations and often the first known attestation of new words. Examples, of which some will be used later in the empirical part, are *New Words* by Merriam-Webster⁴², *About words* by Cambridge University Press⁴³, *Urban Dictionary*⁴⁴, and *WordSpy: Dictionary of New Words*.⁴⁵ Outside of the English-speaking world, *Wortwarte*⁴⁶ is an example of a German project that documents German neologisms based on newspaper data (Lemnitzer, 2011). Thus, using electronic mass communication offers a unique opportunity to examine the diffusion of neologisms from a very early stage in their establishment process (Würschinger, et al., 2016: 35).

It is not surprising that more and more researchers make use of the internet as a source (cf. Kerremans, 2015; Würschinger, et al., 2016; Hohenhaus, 2006; Maybaum, 2013). However, it also brings about issues: some areas of the internet, for example search engines like Google or Bing, are often considered as unreliable for qualitative and quantitative linguistic research due to their inconsistency (Lüdeling et al., 2007; Renouf et al., 2005; Kilgarriff, 2003). First, most search engines are restricted to the surface web, which means that only some pages appear, others do not. Google for instance only returns the first 2000 hits, which are ranked and influenced by commercial factors. Thus, rather than

⁴² <https://www.merriam-webster.com/words-at-play/new-words-in-the-dictionary>

⁴³ <https://dictionaryblog.cambridge.org/>

⁴⁴ <https://www.urbandictionary.com/>

⁴⁵ <https://wordspy.com/>

⁴⁶ <https://wortwarte.de/>

reflecting the linguistic reality, it reflects economic structures. Secondly, most search engines do not distinguish between orthographic or morphosyntactic differences. While orthographic flexibility is a defining criterion of neologisms, advanced search options are needed to ensure that only variations from the target language are displayed (Kerremans, 2015: 70-71). Lastly, using search engines also means that data cannot be validated nor replicated by other researchers (Lüdeling et al., 2007: 10-12).

Due to these issues, several researchers have started to develop and programme different web-crawlers, both downloadable and on-demand ones (Kerremans et al., 2011: 62). Webcorp (© RDEUS 1999) is an example of a downloadable on-demand crawler, but, since it is only updated once or twice a year it is not sufficient for observing language change (Kerremans, 2015: 73). Other downloadable tools would be *KWiC Finder* (Fletcher, 2001/2007) and *GlossaNet 2* (Fairon et al., 2008). Although these tools are useful, their utility depends on the computer used and the available Internet speed (Kerremans, 2015: 72). In contrast to downloadable ones, Renouf et al. have programmed a crawler in Perl to select and download articles from UK newspapers (namely the *Guardian* and *Independent*) and crawl those (Renouf et al., 2005: 47). Outside the English-speaking world, Falk et al. have designed a crawler for French neologisms in certain newspapers (Falk et al., 2014) and Cartier has created a web platform called *Neoville* in order to track neologisms (Cartier, 2017). To improve the existing crawlers, Daphne Kerremans, Susanne Stegmayr and Hans-Jörg Schmid developed the *NeoCrawler*, which crawls the internet for new words on a weekly basis and accesses a great number of different webpages (Kerremans et al., 2011). Although this crawler also relies on Google, it addresses some issues that have been discussed previously and which are present in other crawlers (Kerremans, 2015: 72).

Since the *NeoCrawler* is one of the main tools used for retrieving neologisms in the empirical part of this thesis, I will give a very brief

description of its setup.⁴⁷ It consists of two parts: The Discoverer and the Observer. The Discoverer checks Google Blogs (a selection of URLs and local files) for unknown graphemic sequences (Kerremans, 2015: 80 f.; Kerremans et al., 2019). The grapheme sequences which are not contained in the Discoverer's dictionary will be extracted and treated as potential neologism candidates. Subsequently, the words are downloaded and set into frequency lists, where they get filtered in order to exclude the least likely neologism candidates. This filtering processing includes sorting out highly frequent words, words with more than two digits as well as words with less than three letters and proper names. Afterwards, they are compared to a reference dictionary in order to check their 'knownness' (Kerremans, 2015: 81). After this, in another filtering step, non-words amongst the unknown graphemic sequences are deleted by performing a frequency-of-occurrence analysis of all graphemic strings listed. Eventually, each candidate is awarded a 'type quality' that indicates how likely it is to be a neologism. After having concluded all these steps, a potential neologism is marked as such by being added to the database, where its diffusion will be monitored by the Observer (Kerremans, 2015: 84).

On a weekly basis, the Observer undertakes its automated crawling rounds, similar to a manual Google search. It sorts out blacklisted and irrelevant webpages, as well as false positives, duplicates, and outdated versions of homepages to ensure the integrity and validity of the collected data (Kerremans, 2015: 84). Once all the irrelevant and unusable pages have been deleted, and the number of tokens of the investigated neologism has been extracted, the pages are ready for linguistic analysis. Besides providing the user with technical information, like the process ID, time restriction and search string and date, the Observer also shows the number of pages and tokens that were returned for a certain word. This gives the user quantitative metrics of the

⁴⁷ A detailed description of how the NeoCrawler was developed and how it works can be found in Kerremans et al. 2011 and Kerremans, 2015. Within this thesis, only the basic outline, which is needed to understand how the words for the questionnaire were selected, will be given.

neologism in question. The Observer can also give a more detailed analysis, providing a list of the original pages, on which the tokens were first found and allowing the user to visit them in order to see the neologism in context. It also provides the option for the user to leave feedback on the field of discourse, type of source and the authorship in each individual hit (Kerremans, 2015: 84-92). Due to the dichotomy of Discoverer and Observer, as well as by incorporating the identification, retrieval and analysis of novel linguistic material on the internet in one programme, the NeoCrawler depicts “a convenient, reliable and fast application for investigating the development of English neologisms on the web” (Kerremans, 2015: 92). Therefore, it is used as the main source for retrieving suitable neologisms within the framework of the conducted studies. Secondary means, which were used in the empirical part of this thesis to assess the words further, are Twitter, Google, and online dictionaries. More details on how these tools were used will be given in the method section of each study (cf. section 6.1; section 7.2; section 9.2).

One of the studies conducted investigated German neologisms (cf. section 8). Since the NeoCrawler only contains English neologisms, other means for retrieving the German candidates had to be found. For this purpose, the neologism dictionary of OWID (*Online Wortschatz Informationssystem Deutsch*)⁴⁸, which is offered by the IDS, was used. The dictionary contains more than 2100 new words and phraseologisms as well as new meanings of established words. All words are sorted according to the year and month they were added. Besides this, I also made use of the afore-mentioned homepage *Wortwarte* (Lemnitzer, 2011), a weekly updated collection of new words that appeared in German media. In contrast to the NeoCrawler, the OWID and *Wortwarte* do not crawl the internet on a weekly basis and therefore do not provide information about the neologisms’ developments. Thus, additional tools

⁴⁸ <https://www.owid.de/docs/neo/start.jsp>

were used in order to monitor the progress and diffusion of the chosen neologisms, which will be explained in more detail in section 8.2.

Retrieving suitable neologisms was the first step in the methodological process, whereas investigating and analysing them, is the second step. In order to accomplish this, I decided to use (online) questionnaires. Questionnaires can be defined as “the self-completed, written questionnaire that respondents fill in by themselves” and which are employed “as research instruments for measurement purposes to collect valid and reliable data” (Dörnyei & Taguchi, 2010: 1). The use of questionnaires has a long tradition, especially in dialectology and sociolinguistics (Chambers & Trudgill 1998: 21– 4). Questionnaires were pioneered as early as 1876 by Georg Wenker in Germany (Milroy & Gordon, 2003: 14), a method later adopted by McIntosh (1952) in Scotland and Le Page (1954) in the Caribbean. Since questionnaires allow to quickly collect a large amount of easily processible data, they started to be widely used over the last decades (Milroy & Gordon, 2003: 14) and are even stated to have “become one of the most popular research instruments in the social sciences” (Dörnyei & Taguchi, 2010: 1). However, they are not suitable to the same extent to investigate variables like pronunciation, which is difficult to be assessed in questionnaires. In contrast to this, they have proven to be a valid tool for investigating lexical changes (Milroy & Gordon, 2003: 53). Various aspects of neologisms in different languages have been investigated by using questionnaire studies (cf. Ketabi et al., 2010; Sólyom, 2014). Additionally, questionnaires have been utilised to research blends (cf. section 4.1.3). Similar to the retrieval of words, the internet also changed how questionnaires were conducted. With the increasing popularity of the internet in the 1990s, a shift from postal to online surveys took place and more and more researchers started to conduct online studies (cf. Murray et al., 1996; Murray & Simon 1999).

Online questionnaires bring about various advantages. Once they are up and running, they run themselves and no further input from a researcher is needed. When enough participants have taken part, the

data can be easily downloaded into spreadsheets (Dewaele, 2018: 271). Furthermore, using the internet for questionnaires enables researchers to access either larger and more diverse populations all over the world, or, on the other end of the spectrum “small, scattered, or specialised populations” (Dörnyei, 2007: 121). Besides this, in contrast to questionnaires conducted by fieldworkers or interviews, written and online questionnaires bear the advantage of anonymity (Dörnyei, 2007). By not having a direct face-to-face interaction between researcher and participants, the pressure on the participants might be reduced and the level of honesty increased (Dewaele, 2018: 271). Hence, questionnaires minimise the risk of participants exaggerating or distorting their responses to please the authors of the questionnaire, the so-called *social desirability bias* (Oppenheim, 1992: 183).⁴⁹

However, online questionnaires also have limitations. One of the major issues is the inevitable self-selection bias (Dewaele, 2018: 271) as with “Internet-based research [...] it is not possible to apply a systematic, purposive sampling strategy, as all participants are self-selected” (Dörnyei, 2007: 122). A consequence of this is a possible unequal distribution in demographics such as age, gender, education, and socio-economic imbalances (Dewaele, 2018: 273 ff.). This can be the result of factors such as that younger people use the internet more, that participants with a higher socio-economic status are more likely to be able to afford a computer and access the internet, or that tools like Likert scales could look intimidating or even threatening to some participants, bringing back unhappy school memories, or provoking a fear of looking stupid in the eyes of the researcher (Dewaele, 2018: 275). While the main limitations mentioned are demographic in nature and since demography does not play a predominant role in the study, I decided that online questionnaires are a suitable tool for investigating the chosen neologism.

⁴⁹ In the past, sociolinguistics made use of questionnaires conducted by fieldworkers that similarly bear the disadvantage of fieldworker bias (Chambers, 1998) and a potential bias due to the mere presence of an unfamiliar fieldworker – the so-called Observer’s Paradox (Milroy & Gordon, 2003: 52).

Using online questionnaires especially stood to reason when considering that the extracted words were taken from the internet and thus should also be tested in the same medium. Hence, while questionnaires are a long-established tradition in certain fields and sciences, online questionnaires offer further advantages in contrast to traditional paper questionnaires and seem suitable in this context.

Before designing the questionnaires, I had to consider several factors, such as which programme to use and how I want to distribute the questionnaires. A great variety of tools are used by researchers for these tasks, such as *Google Forms*, *Survey Monkey*, *Lime* or *Mechanical Turk* (Dewaele, 2018: 271). After considering the various options, I decided to use a different tool called *QuestionPro*.⁵⁰ This decision was based on the fact that the tool offered all features needed, such as an unlimited number of questions and participants, the options to disable a *back* button and to incorporate a progress bar. The latter is important as it has an encouraging effect on the participants to finish the study (Brace, 2004: 147-160). The former, the option to disable the *back* button, was of great importance for the study as participants were presented with a neologism, asked whether they know it and then had to give/guess a definition. Later they were presented with the definition. If they have had the option to go back, they could have adapted their definition after having been presented with the real definition and thus could have altered the data/outcome. The fact that participants were able to fill in the questionnaire on a computer, phone or tablet also made it an appealing choice. As a method of distribution, I decided on snowball sampling. Being a German who lives and works in an English-speaking environment helped to get the 'ball' rolling for both, the English and German surveys. Therefore, the questionnaires were distributed amongst my students, colleagues and friends as well as published via all sorts of different private and institutional social media platforms. As an

⁵⁰ <https://www.questionpro.com/>

incentive three raffle prizes were offered, ranging from 25 to 100 GBP. The questionnaires were available online for about two weeks.

All in all, the internet has not only changed the ways and means in which we examine and research neologisms but also the speed and number of neologisms invented. While extracting new words is still manual labour, the methods for documenting and analysing them have undergone great changes. As a result of this, for the empirical part, the materials were retrieved from internet corpora, crawlers, and websites. They were further assessed by using search engines, social media platforms and online dictionaries. To investigate them, online questionnaires were used.

II. Empirical part

6. Questionnaire study I

6.1 Method

6.1.1 Materials - retrieving neologisms

After having laid the theoretical groundwork, the empirical part will show how the research questions and hypotheses (cf. section 1.2) were examined. While the general methodology was outlined in section 5, some questionnaire specific factors still need to be explained. As a first step I looked through the *NeoCrawler* (Kerremans et al., 2011) and retrieved neologisms. During this process I focused only on new lexical N+N blends, all other neologisms were not taken into account. Once I had an extensive list of new words, I manually counterchecked them in the Oxford Dictionary to ensure that they have not yet been added.⁵¹ In a next step, I decided on three distinguishing variables in order to find a comparable pool of words amongst the extracted ones. These variables were based on some of the hypotheses that will be investigated in the questionnaire studies (cf. section 1.2). Thus, in accordance with some of the hypotheses I decided on the following three variables: *frequency* (cf. section 1.2 **H5**), *first occurrence* (**H6**) and *early user group* (**H7**). Applying those variables to all the extracted neologisms from the NeoCrawler led me to the following 24 neologisms:⁵²

⁵¹ See appendix 1.1 for the full preliminary list of all N+N blend neologisms extracted from the NeoCrawler. All of them were investigated according to their frequency, first occurrence and the early user group. The presented 24 candidates have proven to represent their categories the best.

⁵² See appendix 1.2 for an overview of the definitions of the respective neologisms.

	High frequency		Low frequency	
	First occurrence 2001-2009	First occurrence 2010-2017	First occurrence 2001-2009	First occurrence 2010-2017
Private	SNACCIDENT CATIO FITSPIRATION	PRESSTITUDE BLEISURE TRUMPANZEE	BREADATARIAN RUNGER BROGA	CRUFFIN HONEYTEER BEGPACKING
Professional	VEGANUARY GLAMPSITE HAMDOG	BAECATION SHARENTING CRONUT	MANTRUM OBLICATION DOGA	MARANOIA BRONGERIE BELFIE

Table 5 Chosen neologisms for study I

To assess the new lexemes according to the three mentioned variables, several tools were used: the NeoCrawler, Google, Twitter, and Online Dictionaries. While the last three served as a source of information regarding the first occurrence and the early user group of the neologisms, the NeoCrawler mainly provided information about their frequency. While frequency plays a major role in the establishment of words, it also presents difficulties (cf. section 3.2.3), which affected my research, too. The first problem I encountered was that the *NeoCrawler* provides information about page and token frequency. I decided to focus on the former. Even though entrenchment is likely to be promoted if a person reads the same word several times in one article, it does not necessarily contribute to the word's usualisation and diffusion in society (cf. section 3.1.1; section 3.2). Another issue was that, while the NeoCrawler is reliable and automated to a high extent, it was still necessary to go through the hits manually to sort out false hits. The most common source for false hits was when the neologism in question was used on a non-English homepage, was the result of typos or was used in a different meaning.

Once this was done, I was able to calculate an average frequency for each neologism. Thus, I summed up all the frequencies for each word and divided the result by the number of weeks, for which the neologism was observed. Following this, a threshold between high and low-frequency words had to be found. Based on the available data, I decided for a threshold of five, with low frequency being assigned to words with an overall average below it, and high frequency above it. While I am aware that this is a somehow arbitrary value, I tried to choose words on the respective ends of the frequency spectrum when possible, to ensure

considerable differences between the neologisms. However, as the chosen neologisms are also meant to differ in their first occurrence, they have been added to the NeoCrawler at different times. This leads to bigger samples sizes for some words than for others. Regarding this issue, I decided to look at both, the general average of the respective neologisms over the whole course of its time in the NeoCrawler as well as in a restricted time window (from week six to nine in 2018) from which I had data for all neologisms in question.

Table 6 shows that most words have similar averages in both data sets, with the first twelve words (in purple) being high-frequency and the last twelve (in red) low-frequency words. The only two words that exhibit a great difference between their general average and their average in the restricted time frame are *veganuary* and *hamdog*. *Veganuary* shows two extremely high frequency values, so that the difference did not matter in so far as it was counted as a high-frequency word either way. The scenario was slightly different in the case of *hamdog*, which had a high overall frequency of 7.35, but only 1.75 for the chosen weeks. In this case, I decided to rely on the overall average rather than on the shorter time period and thus considered *hamdog* as a high-frequency word. While I managed to choose mainly words located on the extreme ends, *catio*, *doga* and *belfie* depicted borderline cases in which, once more, the overall frequency average was taken for their classification.

NEOLOGISM	OVERALL FREQUENCY AVERAGE	FREQUENCY AVERAGE WEEK 6-9/2018
SNACIDENT	6.37	6
CATIO	5.17	4.25
FITSPIRATION	9.47	8.75
VEGANUARY	17.29	13
GLAMPSITE	9.39	10.5
HAMDOG	7.35	1.75
PRESSTITUDE	11.94	10.5
BLEISURE	9.5	9
TRUMPANZEE	11.75	13
BAECATION	13.94	14.5
SHARENTING	12.11	14.75
CRONUT	8.83	8.5
BREADATARIAN	0.12	0.25
RUNGER	0.21	0.25
BROGA	3.05	4
MANTRUM	2	2.25
OBLICATION	0.29	0.25
DOGA	3.21	5.5
CRUFFIN	2.25	2.25
HONEYTEER	0.75	0.5
BEGPACKING	2.8	3.5
MARANOIA	1.417	1
BRONGERIE	0	0
BELFIE	4.23	5.5

Table 6 Frequency averages of the chosen words with purple marking high-frequency words, and red marking low-frequency words

Once the frequency of the words was determined, I investigated their first appearance on the internet and their early user group. The appearance spectrum reaches from 2001-2017, with me choosing 2009/2010 as a cut-off between ‘old’ and ‘new’ neologisms. In accordance with Kerreman’s *page-level classification scheme*, I divided the category of early user group into *professional* and *private* (Kerremans, 2015: 90 f.).⁵³ Professionals denoting people who are paid for writing an article in contrast to private people, who write for their own pleasure. This implies that I classified blogs, social media, forums, and online dictionaries (here mainly Urban Dictionary) as private and newspapers, tabloids, magazines, and journals as professional. The fact that some of the latter

⁵³ Appendix 1.3 shows all 24 neologisms categorised according to their first occurrence and their separation into private and professional.

do have a social media presence is considered. Since there are big differences between what is discussed in newspapers and on social media (Paradowski & Jonak, 2012: 134), it was important to incorporate various types of each group. Following the idea of the early user group, it should be mentioned that - due to the sheer lack of reliable information on it – I did not necessarily take the very first instance of a neologism into account, but the first few hits to get a broader picture of the context in which it was first primarily used.

In order to retrieve information about the early user group and first appearance, Google, Twitter and the Urban Dictionary were used. Google offers an advanced search function to look for a word in a certain time frame and restrict it to English speaking homepages. Thus, using this search tool enabled me to extract information about the first occurrences of the neologisms. A problem I encountered was that some homepages/articles are without a date of publication, hence they could not be used for this study. Google search also provided me with a rough idea about the early user group. However, since Google is commercially governed it should not be used as the only source (cf. section 5). Thus, to back up my findings from Google, I went through a similar process on Twitter. The advanced search also provides the option to customise date and language. However, while Google has no time restriction, Twitter only goes back till 2006. Therefore, it was only useful for neologisms that appeared online after 2006. Nevertheless, it has proven to be a valuable addition to the results taken from Google. Besides first appearance, Twitter also allows to draw conclusions about the early user group, indicating whether private people, public figures, or newspaper companies with a Twitter page 'kicked-off' the neologism's Twitter presence.

To get a final confirmation of first occurrence and early user group, I checked online dictionaries for the respective words. In most cases Urban Dictionary has proven to be useful since individuals can write down timestamped definitions. After having gathered all this information and eliminating unsuitable neologisms step by step, the 24 neologisms

presented in *table 5* were classified as suitable candidates for investigation in questionnaire study I.

6.1.2 Questionnaire design and participants

Due to the large number of tested words, two equal questionnaires were designed.⁵⁴ Both questionnaires were split into several subsections concerning their content.⁵⁵ A short introduction served the purpose of giving the participants some background knowledge about the study and making sure that they are, in accordance with the ethical standards, informed about the nature of the study. This is followed by routine demographic questions, such as the participants' gender, age, level of education, location, and native language. Checking the native language enabled me to make sure that only English native speakers took part in the study. While most of these variables are straightforward, a small-scale pilot study showed that the parameter *education* caused a problem. As the study was meant to go out to various English-speaking countries, all of which have different education systems, the question had to be simplified. Hence, in consultation with the participants from the pilot-study, who came from different English-speaking countries, I decided to only distinguish three groups: school degree, university degree or none of them. While not being very accurate, this seemed the only way to get information about the participants' educational background across countries.

The demographics were followed by questions about the participants' media usage, which were supposed to reveal information about the hypothesis that the time spent on the internet and the number of social media platforms used impacts whether people would know and use a neologism.

⁵⁴ For details about which words were used in which questionnaire see appendix 1.5

⁵⁵ A sample questionnaire can be found in appendix 1.4

The next part assessed quality and influence of media type and early user group. Virtual communities are more flexible and have more members and therefore more means to spread a word (Kerremans 2015: 158 ff.). This gives them a disproportional influence on language and its speakers' habits (Kjellmer, 2000: 224) which consequently means that the medium and the way it is perceived might also have an impact on diffusion and use of neologisms (Kerremans, 2015: 158). To assess the quality and influence of media types and the early user group(s), the participants had to rate the quality of information provided by different sources – from which the neologisms were retrieved - such as broadsheets, tabloids, blogs etc. Initially giving examples of each category seemed helpful (e.g., *The Times* for broadsheet, *The Sun* for tabloid etc.), however, the pilot study showed that this led to confusion since it automatically makes the questions culturally loaded. Besides the quality ranking, I also asked the participants to rank the influence of the respective author group. From these ratings, information about the assumed connection between early user group, high-quality media types and usage of the neologisms was assessed (**H8**).

Moreover, I assume that the participants' interests will also resonate in factors such as what they read and which language they use. Hence, the fifth section of the study investigated the participants' interests. Kerremans' fields of discourse from her *Page-level classification scheme* formed the basis for the different interest categories and their subcategories, only minor adaptations were made (Kerremans, 2015: 90 f.).⁵⁶ The participants had to mark in which of the given areas they are interested. Each area was represented by one or several of the chosen neologisms in order to find out whether interest has an impact on the potential use of neologisms (**H9**).

⁵⁶ The adaptations mentioned include that the category *advertising* was not used, as it did not seem suitable. Further, using each of the subcategories of *lifestyle* as a category on its own seemed useful (cf. appendix 1.4)

In the last and major section of the questionnaire, participants were presented with one neologism at a time (without context). Apart from the neologisms, participants were also confronted with twelve distractors. They were meant to fulfil two functions: first, they served as control words. Secondly, they were thought to keep the participants keen and give them a ‘sense of achievement’, as being constantly confronted with unknown words might be demotivating and make people quit the study. The distractors were supposed to be formally as close as possible to the neologisms, thus N+N blends. However, they were meant to differ in the sense that they are established words that have been conventionalised, institutionalised, lexicalised, and entrenched by the majority of speakers. Their entry in the Oxford Dictionary was regarded as an indicator for their successful diffusion and usualisation.⁵⁷ As this study aimed at speakers from all over the world, ‘locally’ loaded terms (e.g. *Oxbridge*) had to be avoided. This left me with the following twelve distractors:

<i>Distractors</i>		
MOTEL	SITCOM	ROMCOM
WORKAHOLIC	BRUNCH	CAMCORDER
SMOG	EMOTICON	BOLLYWOOD
SHOPAHOLIC	CHOCAHOLIC	SPANGLISH

Table 7 Distractors for questionnaire study I

Upon being presented with the words, the participants were asked an array of different questions, aiming at gaining data for the different hypotheses. As a first step they were asked whether they have come across the presented word beforehand as well as whether they know the meaning of the word. This way familiarity and recognisability were assessed (**H2**). In order to acquire data on semantic transparency, the participants were asked to give the meaning of the word, and if they did

⁵⁷ While it is true that not all words in dictionaries are well established amongst the majority of speakers, using the Oxford Dictionary as a reference to distinguish neologism from non-neologism seemed valid (cf. section 2.2). The pilot study further showed that all distractors were well-known amongst the participants.

not know it, to try to guess it (**H1**). This also enabled me to make sure that participants who claimed that they knew the given neologism were not led on by a *social desirability bias* (Oppenheim, 1992: 138) (cf. section 5). For retrieving data about formal transparency, participants had to name the two constituents of the blend neologisms (**H1**). In a next step, the participants were given the meaning of the lexical item and were asked whether this is ‘a good way to say this’, aiming at formal appeal. For conceptual appeal they were asked whether they like the idea/concept of the word (**H3**). This was followed by the question of whether they personally might find themselves in a situation where they would use this word. This question explores the link between the words and the participants’ interest (**H9**).

In order to create a connection between the early user groups’ influence and media quality, the participants had to state in which media types the word might be most commonly found – using the same categories as in the previous ranking question about the quality of media types (**H8**). In the next step, the early user group of the neologism was revealed, and the participants were asked whether they would use the word. Subsequently it was inquired why participants would or would not use the new lexeme. This open question added a qualitative element to the otherwise quantitative study, which might allow the identification of significant trends within the quantitative data analysis (cf. Creswell & Plano Clark, 2010). The final question of the study aimed at investigating general, overall appeal by asking whether the participants think that the word deserves to become a thing (**H4**). The rather casual wording of the question is intentional in order to try to avoid being too suggestive. After completing the study, the participants were given a link that led them to a raffle.

With 72 and 74 participants in the respective questionnaires, an overall of 146 participants took part.⁵⁸ Amongst the participants a rather

⁵⁸ Some questionnaires had to be filtered out due to incompleteness or being filled in by non-English native speakers.

unequal distribution of groups could be found (cf. tbl. 8)⁵⁹, one of the major disadvantages of online questionnaires and the snowball distribution (cf. section 5). *Table 8* provides a summary of the demographic distribution and *table 9* visualises the big disproportion amongst different age groups⁶⁰, as well as location and, to a lesser degree, education. In sociolinguistics it is assumed that four to five subjects per cell are enough to be representative for the population tested (Labov, 1972: 38). Although this study does not fulfil this criterion, the participants' demographics play a minor role, with only one hypothesis investigating them. Thus, it was nonetheless decided to still attempt to identify trends regarding age, education and location within the available data.

	School		University		
	Female	Male	Female	Male	Age Groups
City	23	10	17	7	<25
	2	0	18	13	26-45
	1	0	4	3	>46
Countryside	12	2	9	3	<25
	0	0	3	0	26-45
	2	0	4	5	>46

Table 8 Distribution amongst participants in questionnaire study I

⁵⁹ See appendix 1.6 for a complete overview of all participants, including those who preferred not to provide information for all categories.

⁶⁰ The age groups in the questionnaire were initially divided into <25, 26-45, 45-65, >66 (cf. appendix 1.4). However, as I could not find any participants for >66, I was left with only three age groups for the analysis.

	Categories	Participants
Gender	Female	99
	Male	43
Education	School	54
	University	88
Location	City	104
	Countryside	40
Age	<25	90
	26-45	35
	>46	20

Table 9 Participants per category

6.1.3 Statistical analysis

Before the results are presented, I will briefly outline which variables and statistical tests were used when analysing the questionnaire study. Following the hypotheses stated in section 1.2, the demographic input variables are *age*, *education*, *location*, *internet usage*, and *social media platform use*. The analysis investigates their impact on the output variable *renown of neologism*, which denotes whether participants have come across a neologism before and whether they know its meaning, and the variable *using a neologism*. Further, the influence of the independent variables of *frequency*, *first occurrence*, *prestige/influence of the early user group/media type*, and the *participants' interest* on the dependent variable *use of a neologism* is analysed. If the subgroups of the respective independent variable consisted of more than two groups, a one-way ANOVA test was used to analyse the data. If the subgroups of the respective independent variable consisted only of two group, (one sample or independent) t-tests were applied. Furthermore, some input variables were tested for possible correlations with the respective output variable, such as *transparency*, *appeal*, and *recognisability*.

The outcome of the study led to a further investigation (cf. section 6.2.5) in which all three factors were combined in a regression which

ultimately led to the use of a mediation model. These models seek to sort out a process whereby a predictor influences a response, by considering a mediating variable (Hayes, 2018: 77ff). In other words, such models identify and explain the underlying processes and mechanisms of an observed relationship between an input variable and an output variable via including a hypothetical variable, the mediator variable. Instead of a direct link between the input and output variable, a mediation model suggests that the input variable influences the mediator, which then impacts the output variable. Hence, the mediator variable explains the character of the relationship between in- and output variable (MacKinnon, 2008: 5ff.).

6.2 Quantitative results

6.2.1 Demographic data: Age, Education, Location, Internet, and Social Media Usage

One hypothesis investigated states that “younger people, people in urban areas, highly educated people, and people who spend more time on the internet or use more social media platforms will know and use more words in contrast to older people, people in rural areas, less educated people, and people who spend little time on the internet and use a limited number of social media” (H10) (cf. section 2.1). Amongst the input variables *age*, *education*, *location* and *internet usage*, only *age* is a significance predictor for the output variable *renown* [$F(2,142)=5.122$, $p=.007$].⁶¹ Younger people (up to the age of 45) tend to be significantly more familiar with the presented novel lexical items than older people [$t(143)=2.986$, $p=.003$]. *Graph 1* visualises how with rising age, the familiarity with the 24 neologisms in question declines.

The variable *age* is also correlated to the number of different social media platforms used. *Graph 2* shows that younger people use

⁶¹ Although it was not part of the hypotheses, I also tested whether there is a difference in gender and found that women within the sample knew significantly more neologisms than men [$t(140)=3.037$, $p=.003$].

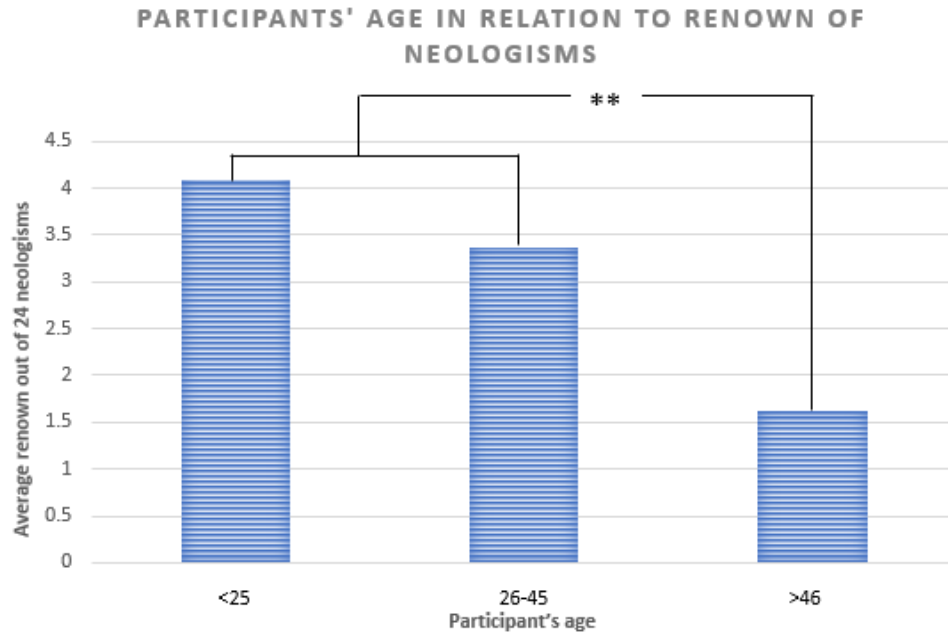
significantly more social media platforms than older people [$F(2,139)=19.748, p<.001$].⁶² While there is no significant correlation between using more social media platforms and the inclination to use more neologisms, there is a low but significant correlation between the independent variable *social media platform use* and the dependent variable of *knowing a neologism* [$r(145)=-.207, p=.013$].⁶³ However, due to the low correlation value and the high number N ⁶⁴, this significance has to be treated with caution and cannot be seen as a very reliable result. Similarly, in regard to the independent variable *time spent on the internet*, the statistical data shows a trend which hints towards the fact that people who spend more time on the internet on a daily basis are more likely to know the presented novel lexemes [$F(1, 145)=3.882, p=.051$].⁶⁵

⁶² For more details about the different social media platforms used per age group, see appendix 1.8

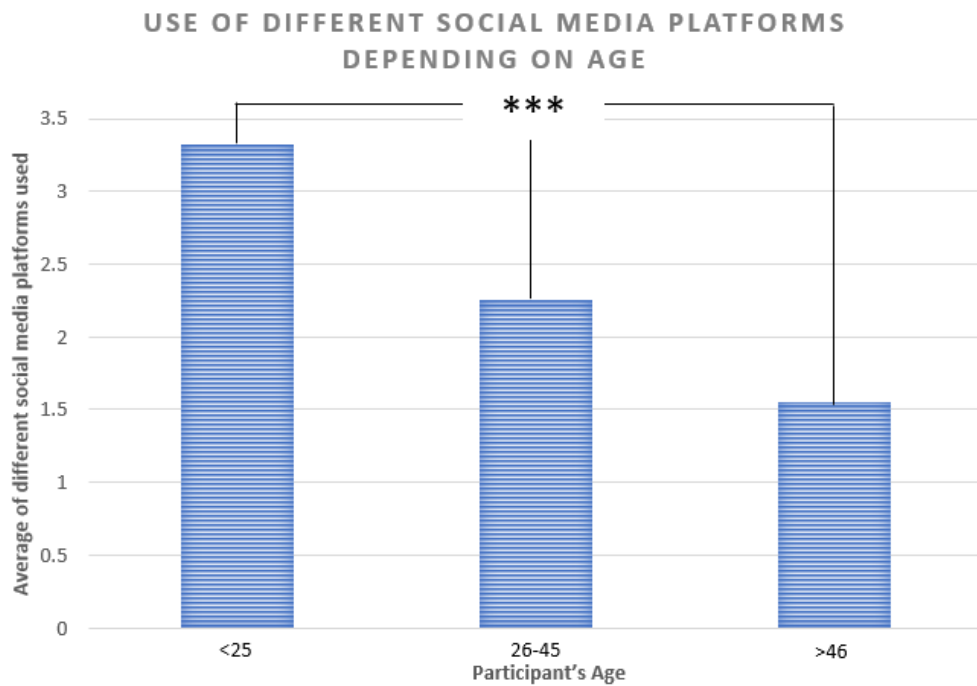
⁶³ No graph is provided since a graphic illustration with such a high number of N does not provide visual clarification.

⁶⁴ N indicates the number of participants. While 146 participants took part in the study, only 145 gave information about their age.

⁶⁵ The corresponding graph can be found in appendix 1.9



Graph 1 Influence of participants' age on renown of neologisms⁶⁶



Graph 2 Social media use graded by age

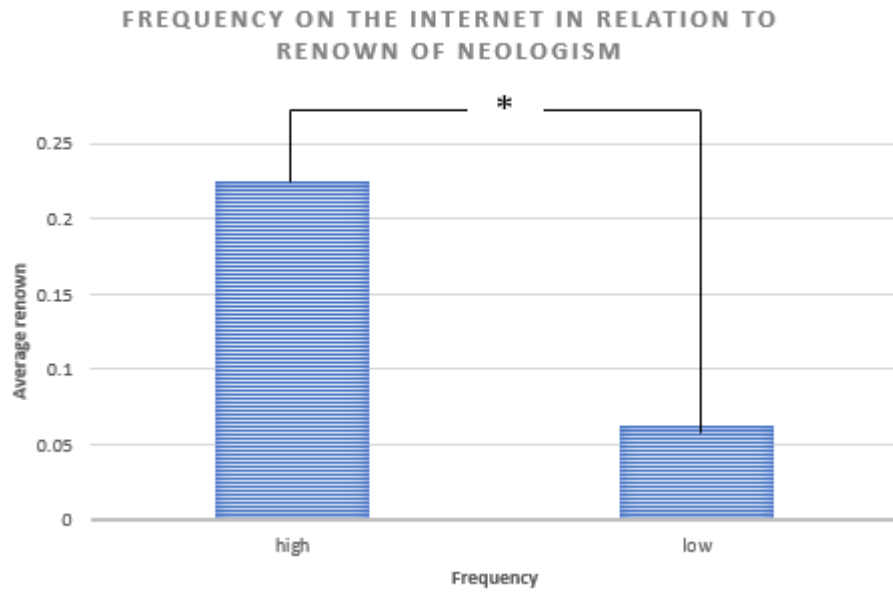
⁶⁶ For all statistical analysis, the following key will be used: * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$

6.2.2 Frequency, first occurrence, and quality of media type/influence of the early user group

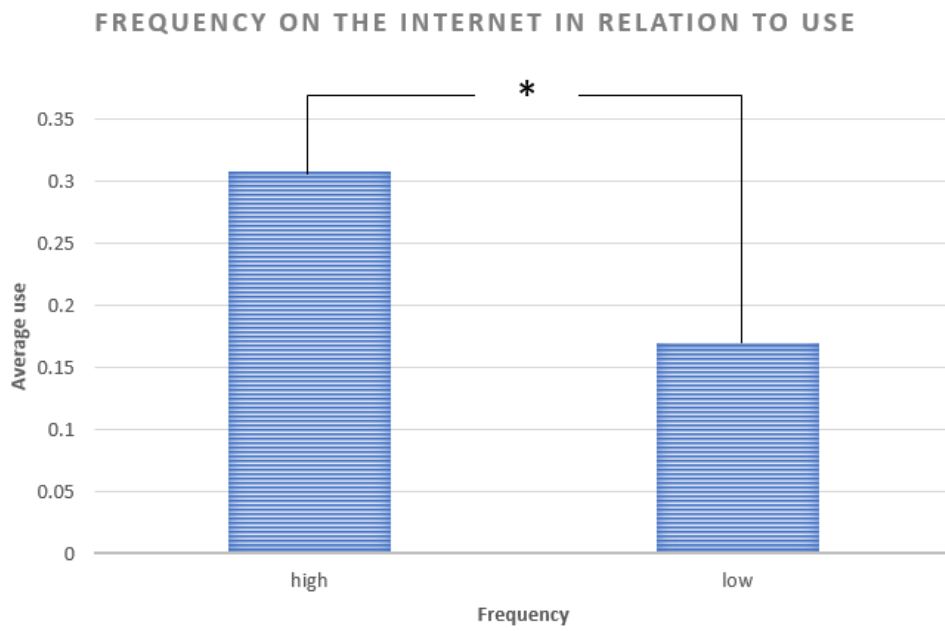
While the input variable *quality of media type/influence of early user group* is analysed on the participant level, *frequency* and *first occurrence* are analysed on the word level. Thus, instead of looking at the results for each participant, the overall diffusion for each word was assessed and correlated with the neologism's frequency and first occurrence, respectively. Frequency depicts a rather complex variable, as it serves as a dependent as well as an independent one. As a dependent variable, *frequency* compares the measured frequency online to how frequently a neologism is known 'offline'. To investigate *frequency* as a dependent variable, I divided the neologisms in two groups: high and low frequency. It was found that frequency within the internet is reflected by how widely diffused the neologisms are offline [$t(22)=2.115, p=.046$]. *Graph 3* shows that the twelve high-frequency neologisms used in this study have a significantly higher average renown outside the internet in contrast to the twelve low-frequency words.

Frequency is also used as an independent variable in this study in order to gain information about the influence of frequency on the use of neologisms. As *graph 4* shows, it was found that more frequent words have a significantly higher average usage in comparison to low-frequency words [$t(22)=2.261, p=.034$].

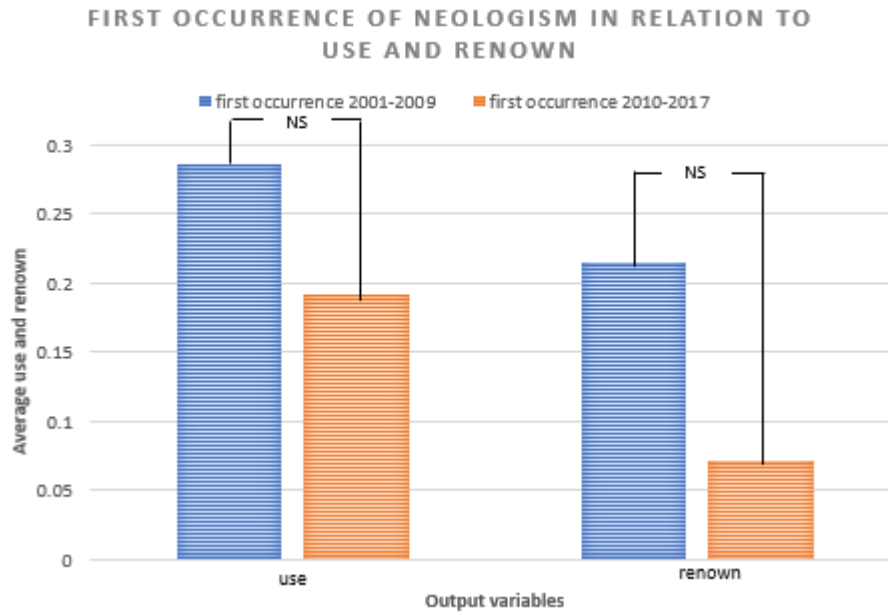
Unlike *frequency*, the predictor variable *first occurrence* does not have a significant influence on whether a new lexical item is known [$t(22)= 1.832, p=.080$] or used by speakers [$t(22)= 1.430, p=.167$] (cf. grf. 5). Only a descriptive trend towards older words being more used and known can be observed, however, no clear evidence for a difference can be found in the data.



Graph 3 Frequency of the neologisms on the internet in relation to renown of neologisms amongst participants



Graph 4 Frequency of the neologisms on the internet in relation to participants using them

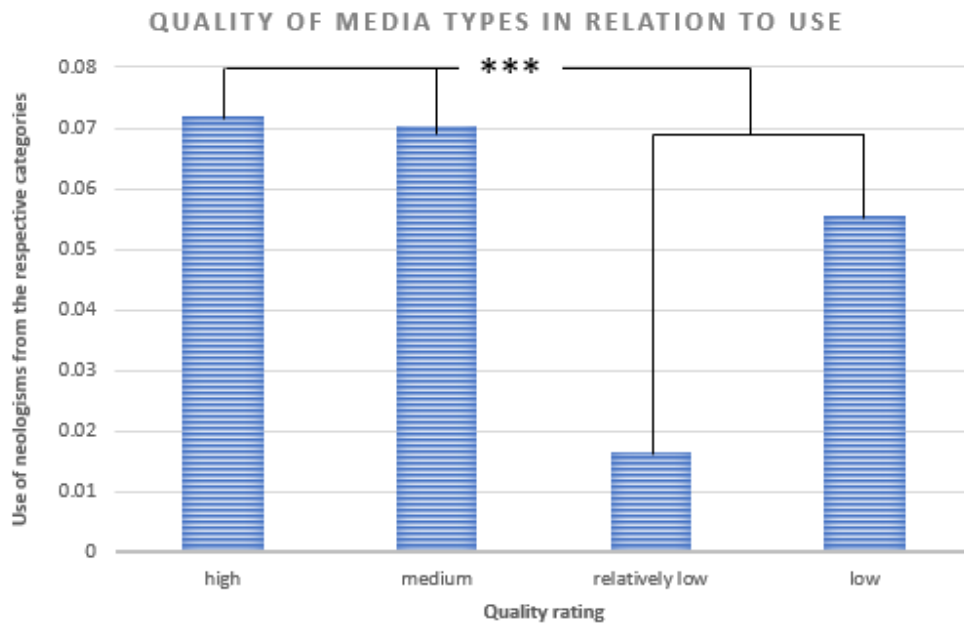


Graph 5 First occurrence in relation to use and renown of neologisms

To investigate the quality of media types and the influence of the early user group – back on the participant level - eight different media types with their respective authors were looked at: broadsheet newspaper, tabloids, magazines, journals, forums, blogs, online dictionary, and social media (cf. appendix 1.9). Each participant had to rank them according to their perceived quality and influence from one (very good/very influential) to five (poor/no influence) (cf. appendix 1.4 & 1.11). Looking at how the participants ranked the quality for the respective media types, a difference between the media types was found [$F(3,1164)=218.458$, $p<.001$]. Based on the outcomes, four significantly different categories were established: high quality (scientific papers and online dictionaries), relatively high quality (broadsheet newspaper), relatively low quality (blogs, social media, forums, magazines) and low-quality media (tabloids).⁶⁷ Graph 6 shows whether participants would use the neologism from the respective category. There is a constant, statistically significant decline in use from high quality media to low quality media [$F(3, 1748)=11.825$, $p<.001$]. Despite the fact that four different quality groups could be established, and usage differs amongst three of them,

⁶⁷ For full statistical details refer appendix 1.12

no positive correlation between the different quality groups and usage was found [$r(4)=-.490, p=.510$].⁶⁸⁶⁹



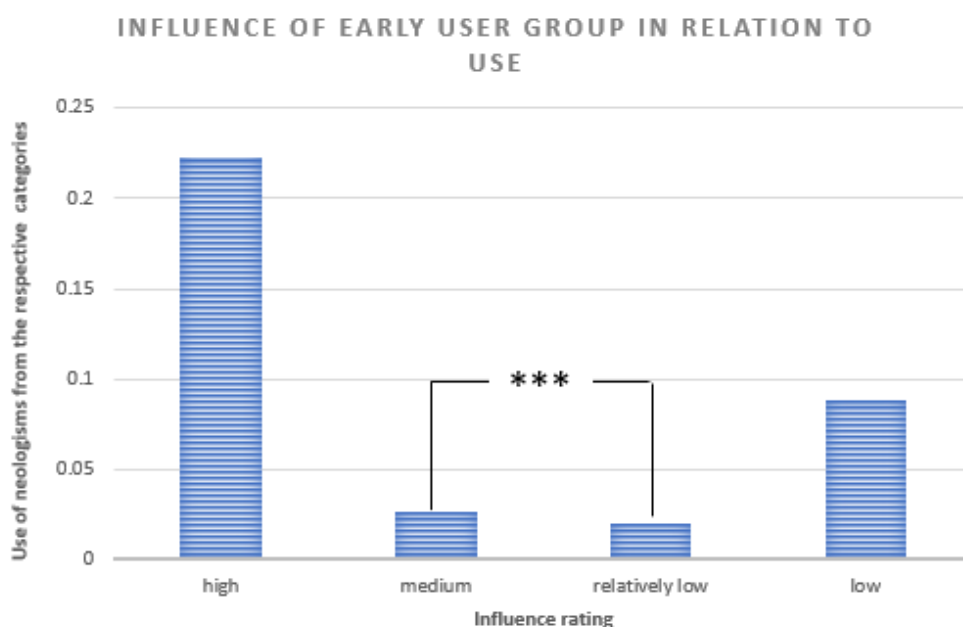
Graph 6 Rated quality of media types and use of the neologisms from the respective categories

A significant difference between groups was found in regard to ranking the influence of the early user group [$F(3, 1164)=90.187, p<.001$]. Four distinct groups were attested: high influence (scientists), relatively high influence (broadsheet newspaper, tabloid, magazine authors), relatively low influence (social media users, bloggers, encyclopaedia writers) and

⁶⁸ A graph showing the quality of the different media types established and their relation to usage can be found in appendix 1.13

⁶⁹ Participants were also asked to denote what media type they think are most likely to use the neologisms. In contrast to the connection between quality and use, this aimed at finding whether there is a positive correlation between perceived quality and use. Thus, while a word might derive from a high-quality media type, a person could perceive it as coming from a low-quality media type and therefore base their usage decision on their own perception. However, the data does not support drawing any conclusion, because only three of the used neologisms were ranked into the category of relatively low quality. Besides, for some words it was impossible to clearly put them into one category since no statistically significant differences between groups were found (cf. appendix 1.14).

low influence (forum users).⁷⁰ Although being similar, they are not congruent with the quality ratings. This becomes particularly clear when comparing the status of tabloids and tabloid authors. While tabloids are ranked as low quality, the authors/user groups are ranked as having relatively high influence (cf. grf. 6 & 7). In *graph 7* only the usage between the two middle groups of relatively high (magazines, tabloids, broadsheets) and relatively low influence (bloggers, encyclopaedia, social media) exhibits a statistically significant difference [$t(1530)=-4.201$, $p<.001$]. There is no significant correlation between use and influence of the early user group [$r(4)=-.598$, $p=.402$].



Graph 7 Rated influence of early user groups and use of the neologisms from the respective categories

6.2.3 Participants' interest

The participants' interest is assumed to be another predictor variable for an increased use of a neologism. To assess this, the participants were presented with several categories and were asked which ones describe their interests (cf. appendix 1.4). All neologisms were put into either one

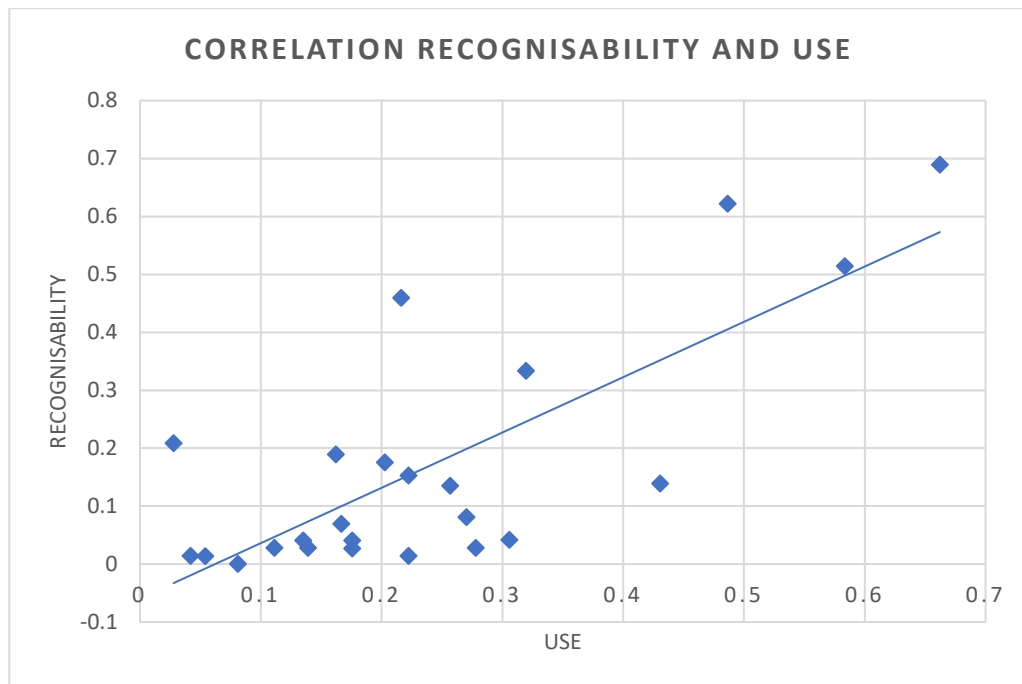
⁷⁰ For full statistical details refer appendix 1.12

or several of these categories (cf. appendix 1.15). Interest and usage were linked by looking at how many new lexemes participants could imagine using from areas they are interested in, in contrast to the ones they are not interested in. However, statistically there was no such difference detectable and hence *interest* did not prove to be a predictor for the output variable *use* [$t(290)=1.826, p=.069$].

6.2.4 Recognisability, transparency, and appeal

The predictor variables *recognisability*, *transparency* and *appeal* were investigated on the word level. In order to do so, all of the participants' answers for each word for the respective variable were combined and averaged (cf. appendix 1.16). Recognisability, tested by the question whether the participants have encountered the novel lexical item before, showed a significant correlation with the output variable *use* [$r(24)=.762, p<.001$]. Thus, participants who recognised a neologism tended to also give positive feedback about using it and vice versa. *Graph 8* shows the correlation between recognisability and use, with each dot denoting one of the 24 neologisms.⁷¹

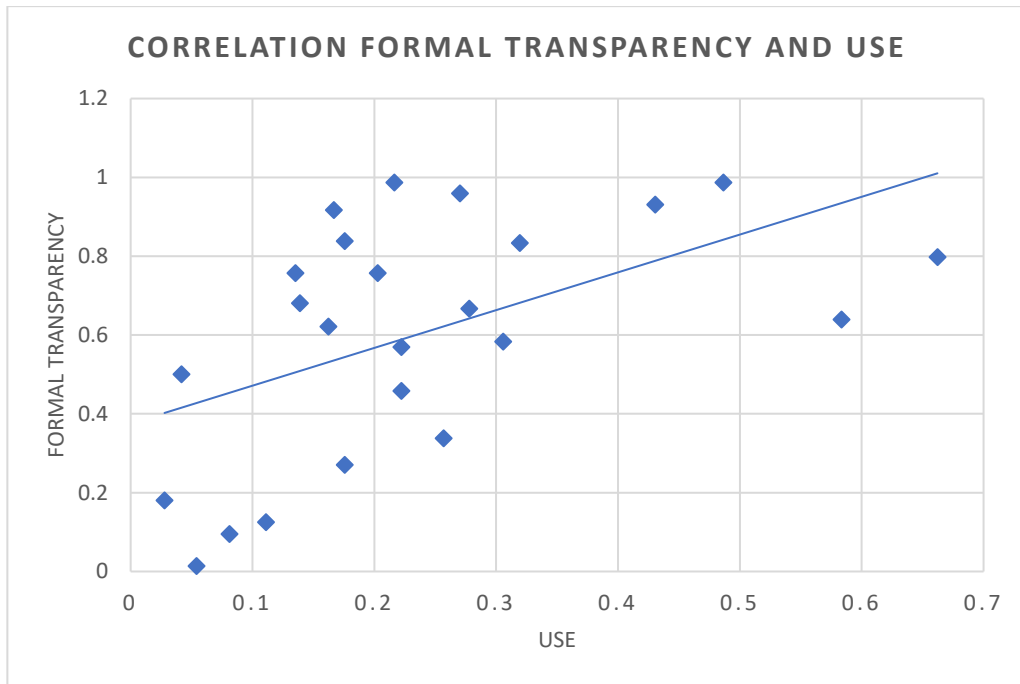
⁷¹ Unsurprisingly, it was also found that participants who not only recognised, but already knew a higher number of presented neologisms, were more likely to use more of the words in contrast to people who did hardly know any of them [$F(8, 137)=2.468, p=.016$].



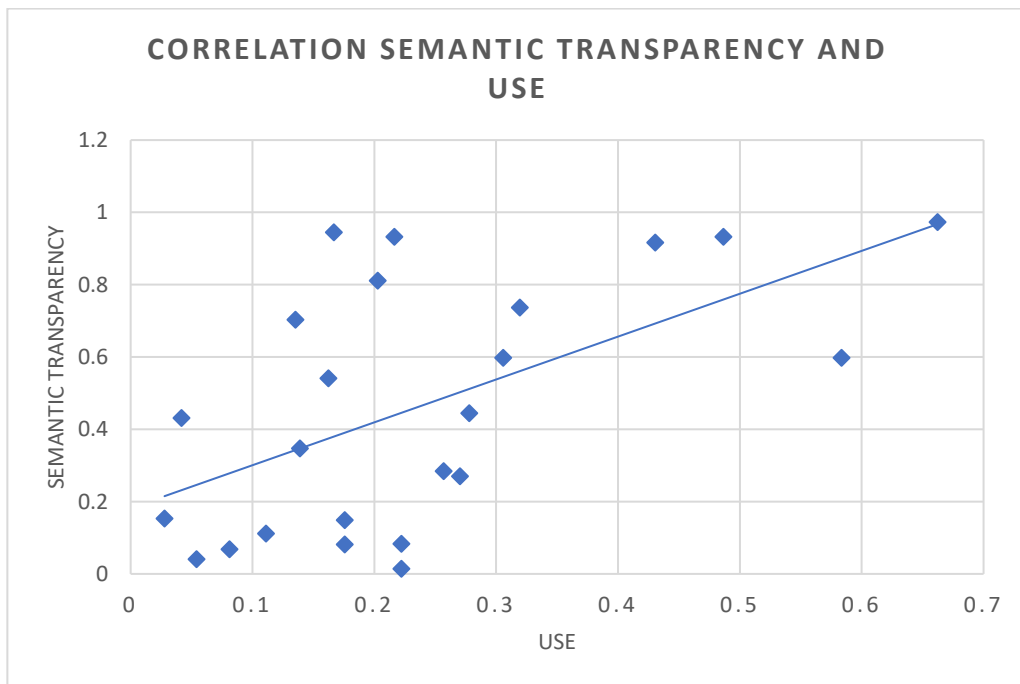
Graph 8 Correlation between recognisability and use

The predictor variable *formal transparency* was tested by asking participants to give/guess the two components of the presented blend. It was found that there is a correlation between the amount of correctly identified components of the blend neologism and the response variable *use* [$r(24)=.522$, $p=.009$] (cf. grf. 9). Hence, neologisms in which the two components were clearer, tend to be more likely to be used and vice versa.

For evaluating the variable of semantic transparency, the participants were asked to give/guess the definition of the respective words. *Graph 10* shows that, like formal transparency, also semantic transparency and use correlate significantly with each other [$r(24)=.568$, $p=.004$], implying that the clearer and more straight forward the meaning of a neologism, the more likely it is used by the participants.



Graph 9 Correlation between formal transparency and use

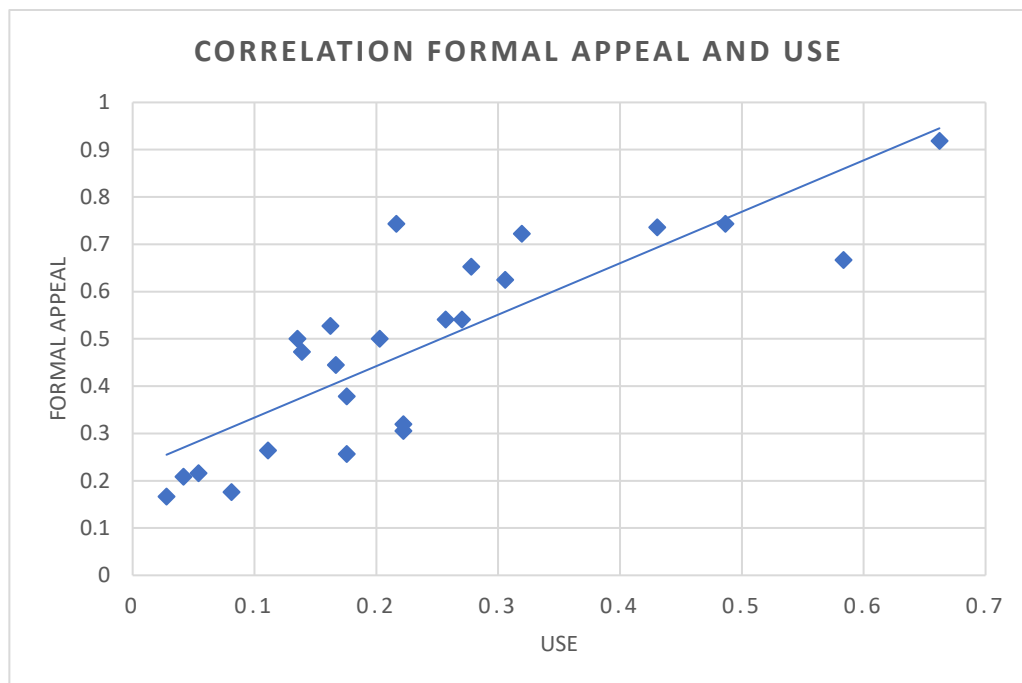


Graph 10 Correlation between semantic transparency and use

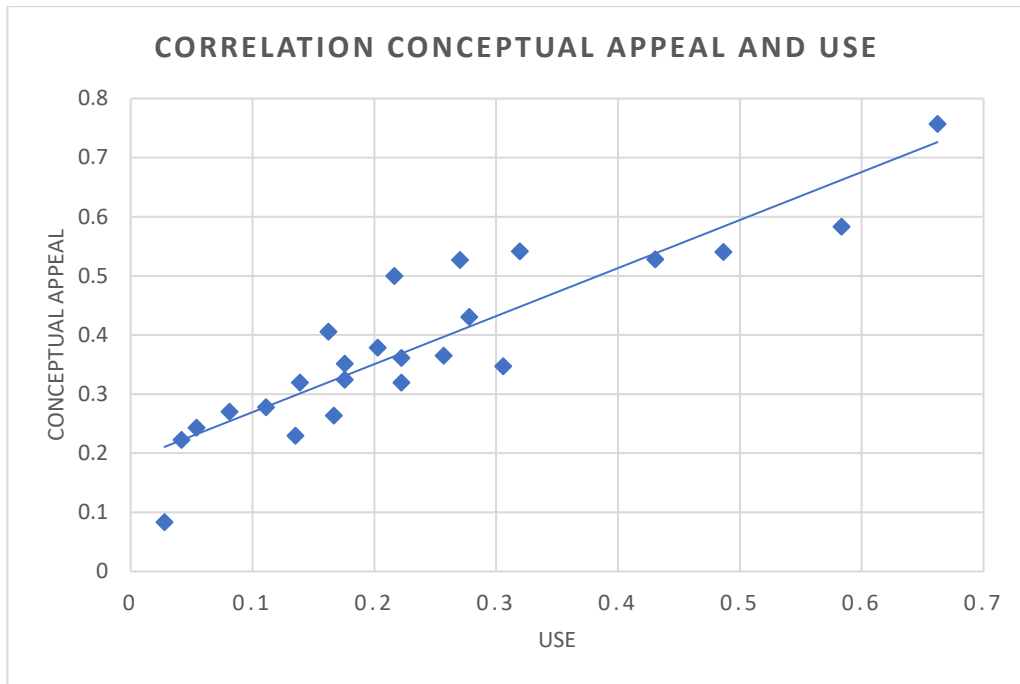
The final three variables that are hypothesised to alter use are *formal*, *conceptual*, and *general appeal*. To find out whether people perceive a word as formally appealing, they were given the definition and asked whether the presented lexical item is a good way to describe it. Graph 11 shows that the variable *formal appeal* correlates significantly with the outcome variable *use* [$r(24)=.830, p<.001$]. Moreover, a positive

correlation between whether participants like the idea/concept behind a word and thus find it conceptually appealing, and use was found in the data [$r(24)=.899$, $p<.001$] (cf. grf. 12).

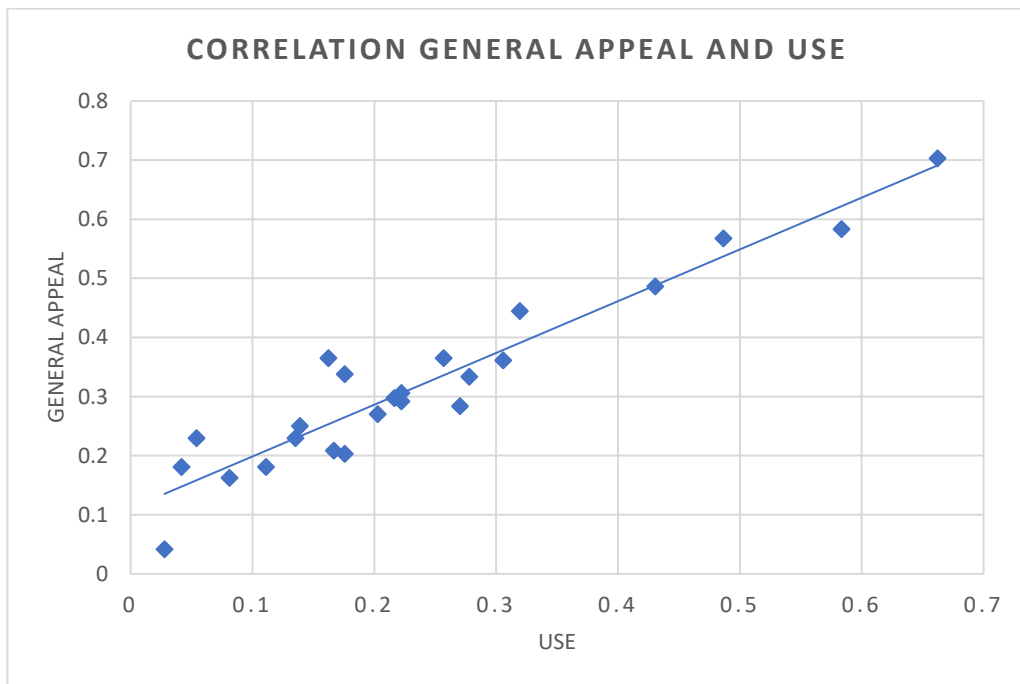
As it was hypothesised that the variables of formal and conceptual appeal impact use, also general appeal is regarded as a predictor variable. In order to assess this, the participants were asked whether the respective neologism should become a thing. The results show that the correlation between *use* and *general appeal* is so high, that attesting an actual difference between them is difficult [$r(24)=.949$, $p<.001$] (cf. grf. 13).



Graph 11 Correlation between formal appeal and use



Graph 12 Correlation between conceptual appeal and use



Graph 13 Correlation between general appeal and use

6.2.5 Further analysis – recognisability, transparency, and appeal

From the outcomes presented in the previous section, it became clear that all investigated independent variables display a significant correlation with the dependent variable *use*. Based on these results, an

analysis was conducted to find whether all input variables have an equal impact on use, or whether they differ amongst each other. Two out of the three predictor variables have subcategories. Therefore, as a first step I had a closer look into the relationship between formal and semantic transparency as well as formal and conceptual appeal, respectively. Up till now it was taken for granted that these variables differ from each other.

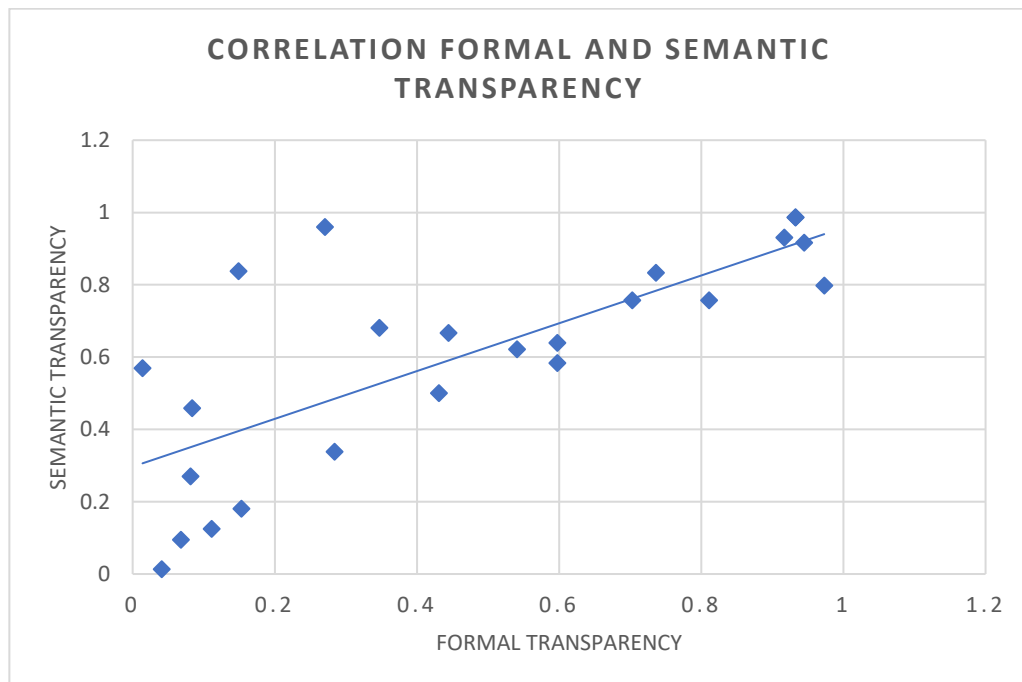
To start with transparency, it was shown that both types of transparency, when looked at independently, impact use. A regression model, for which these two variables must be different, showed that the two transparency types are undistinguishable in their impact on use (cf. fig. 10). This result is further underlined by a rather high correlation between formal and semantic transparency [$r(24)=.752, p<.001$] (cf. grf. 14).

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations		
		B	Std. Error	Beta			Zero-order	Partial	Part
1	(Constant)	.077	.065		1.180	.251			
	formal_transparency	.119	.146	.219	.816	.424	.522	.175	.144
	semantic_transparency	.193	.128	.403	1.504	.147	.568	.312	.266

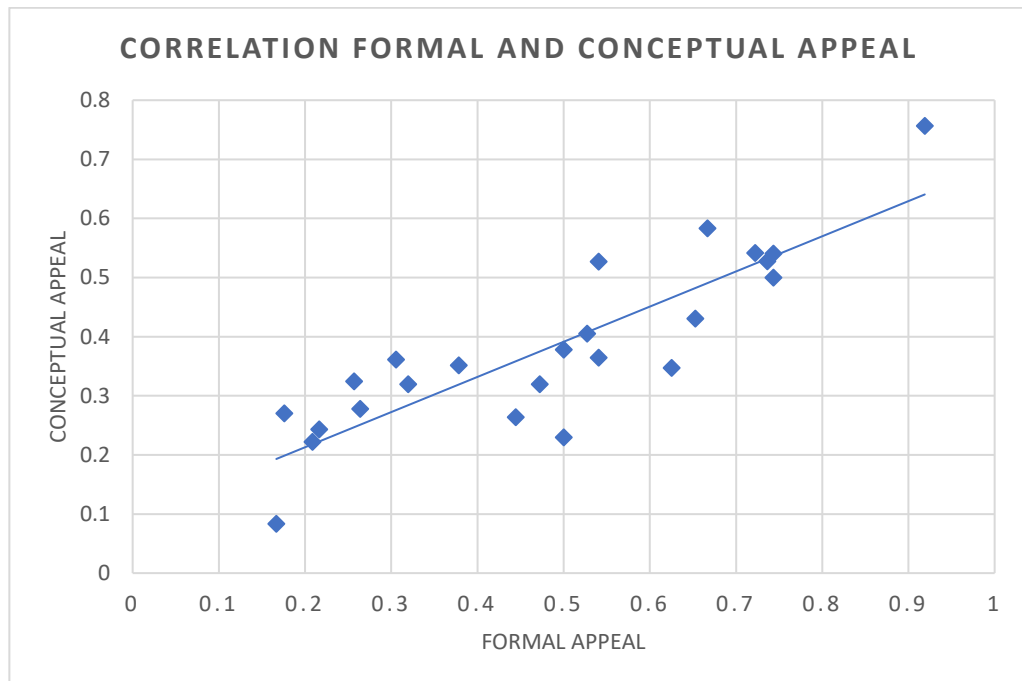
a. Dependent Variable: use

Figure 10 Regression result for use, formal, and semantic transparency



Graph 14 Correlation between formal and semantic transparency

On the basis of the results for transparency, I had a look at the correlation between formal and conceptual appeal and found an even higher correlation between the two of them [$r(24)=.863, p<.001$] (cf. grf. 15).



Graph 15 Correlation between formal and conceptual appeal

Thus, the assumed distinctions are not supported statistically, which means that the categories have to be summarised under their umbrella terms of transparency and appeal. This step left me with three rather than five input variables that have a statistically significant impact on the output variable *use*. When combining all three factors in one regression analysis, it was found that recognisability [$B=.247, p=.106$], in contrast to transparency [$B=-.564, p=.022$] and appeal [$B=1.208, p<.001$] does not have a significant impact on use anymore. The regression outcome further shows that appeal is by far the most significant independent variable.

This led me to investigate whether appeal functions as a mediator between the output variable *use* and the independent variables *transparency* or *recognisability*. Hence, the data was put into mediation models (cf. section 6.1.2). The model examines the process whereby a predictor (*transparency/recognisability*) influences a response (*use*), by considering a mediating variable (*appeal*) (Hayes, 2018: 77ff). Thus, the model depicts an attempt to identify and explain underlying mechanism

of an observed relationship between an input variable (*transparency/recognisability*) and an output variable (*use*) via including a hypothetical variable, the mediator variable (*appeal*). The model suggests that the input variable influences the mediator, which then impacts the output variable. *Figure 11* demonstrates the different steps and paths in the mediation model between transparency, appeal and use. Step 1 of the model, the path between transparency and use – also referred to as the direct effect - shows that the regression of transparency on use, ignoring the mediator *appeal*, is significant [$B=.733$, $t(23)=5.055$, $p<.001$]. The same applies for step 2, the path between transparency and appeal, with a significant regression of transparency on the mediator [$B=.917$, $t(23)=10.815$, $p<.001$]. Step 3, the path between the response variable and the mediator *appeal*, which is controlling for *transparency*, is also of statistical significance [$B=1.336$, $t(23)=5.727$, $p<.001$]. The final step of the analysis revealed that, controlling for the mediator, *transparency* was still significant [$B=-.492$, $t(23) =.651$, $p=.047$] but to a far lesser degree than appeal. Thus, the fact that step 2 and 3 – the so-called indirect effect – are significant shows that mediation has occurred. Therefore, a more transparent new lexeme is regarded as more appealing and thus is more likely to be used.

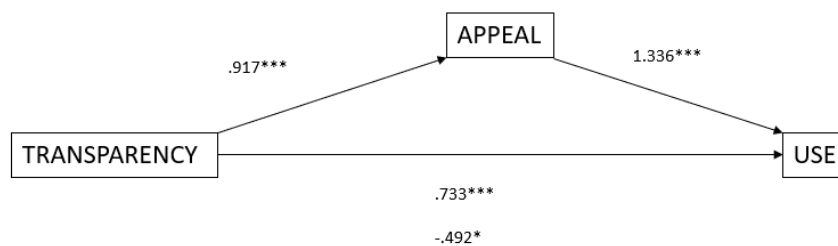


Figure 11 Mediation model transparency - appeal - use

Figure 12 shows the same process, but with the independent variable *recognisability*. The direct effect between *recognisability* and *use*, ignoring the mediator, is significant [$B=-.762$, $t(23) =5.527$, $p<.001$]. Step 2, the path between *recognisability* and *appeal*, shows that the regression of *recognisability* on the mediator is also significant [$B=.785$, $t(23) =5.939$,

$p < .001$]. The third step, the path between the output variable *use* and the mediator *appeal*, controlling for *recognisability*, is also significant [$B = .744$, $t(23) = 4.650$, $p < .001$]. The final step of the analysis reveals that, controlling for the mediator, *recognisability* was not a significant predictor of *use* anymore [$B = .179$, $t(23) = 1.119$, $p = .276$]. Thus, once more the indirect effect is significant, and mediation takes place. Therefore, if a word is recognised it is also appealing which leads to an increased use.

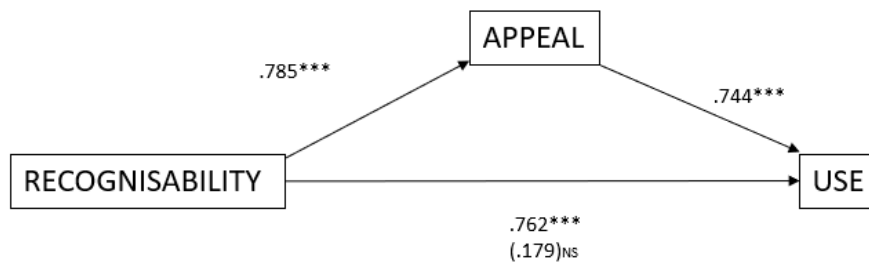


Figure 12 Mediation model recognisability - appeal – use

6.3 Qualitative results

In order to get a qualitative impression beyond the quantitative analysis, the participants were asked about their motives for deciding to use or not use a new lexical item. I looked through all the replies and based on them established the following groups:

- applicable/relatable/personally relevant
- funny/clever/creative
- precise/accurate/efficient/descriptive, useful, Zeitgeist/relevance
- clear meaning, familiar, good concept, good expression.⁷²

It seems that one of the main motives for using new words is personal, hence the applicability of the word to the participants' life as well as the

⁷² Since the whole data set is extremely big, the data from *glampsite* serves as an excerpt to provide an insight into the participants' replies (cf. appendix 1.17). Appendix 1.18 gives a visualisation of the absolute values for each of these groups.

reliability of the participants to the new lexeme. Therefore, the more a word has something to do with the speakers' personal life, the more likely it seems that it will be used. Besides, also humour and unusualness play a role. If words are perceived as witty, funny, clever, and creative, participants seem more inclined to use them.

As communication is about being understood, participants also repeatedly stated that if a word is precise, accurate and efficient they would use it. This goes along with the fact that words, which have a clear meaning, in the sense of being guessable or having easily detectable components, have a better chance to be used. Furthermore, a word that is perceived as useful and corresponds to the current zeitgeist, and thus bears a certain relevance, has a positive impact on whether participants would use it. Finally, if the participants are familiar with the word, if they like the concept and if they see it as a good way for describing the concept, they stated that they are more inclined to use it.

The reasons people named regarding why they would not use the presented neologisms were partially a mirror image of why they would use them. The participants stated that they would not use a word if it is not applicable to their daily life due to a lack of reliability (e.g. some participants mentioned that they would not use *runger* as they do not run and are not interested in sports). Furthermore, they would not use it, if it is not straight-forward and could lead to misunderstandings in communication. Other reasons were uselessness, if the concept was disliked or if the word was perceived as silly. Besides this, hard pronunciation, the term being either sexist or derogatory, not being familiar with the term as well as a general dislike towards new linguistic innovations could hinder people from using neologisms. Another minor factor was of regional nature, so that some participants would not use a neologism which contains a component which is not from their variety of English (e.g. some British stated that they would not use *baecation* due to *vacation* being an American term).

Thus, first, the personal situation of the participants as well as the extralinguistic reality seem to fuel the use of a new words. Hence, the more applicable, relevant and useful a word is seen as by a speaker, the more inclined they are to use it. Secondly, a new word is more likely to be used if it makes communication easier due to its clear meaning and accuracy. Lastly, the way a word is perceived, either positively in the sense of being funny or creative, or negatively, being discriminating or silly, seems to impact whether it will be used or not.

6.4 Discussion

The study has investigated a broad array of different aspects. Regarding the connection between the participants' personal data and diffusion and usage of new lexical items, the data do not support a definitive statement. However, it was found that younger people know significantly more of the presented neologisms in contrast to older people. A possible explanation for this can also be found in the data. Young people use considerably more social media platforms than older people. This goes in line with different studies from research centres investigating the demographic distribution in social media use, nationally as well as internationally (cf. Ortiz-Ospina, 2019; Chen, 2020; Johnson, 2020). Therefore, there might be a connection between the number of social media platforms used, age and familiarity with the presented novel lexical items. Although not all chosen neologisms were primarily used on social media, the platforms could potentially help to diffuse them (cf. section 3.1.3). The potential of social media for diffusing new words is mirrored in the increased interest in using social media as corpora for extracting and analysing neologisms (cf. section 5). Thus, since young people use more social media platforms, they might have encountered more neologisms by doing so, which would explain why young people tend to know more new lexemes than older people. None of the results from the study, however, show a significant impact of age and social media use on participants using new words.

While it seems somehow surprising that young people know more neologisms but are not more likely to use new words, this can be explained when looking at the nature of the diffusion and usualisation process (cf. section 3.1.1). First, it is generally possible that innovations are diffused but not usualised (Schmid, 2020: 95). Secondly, while mass media is a great tool for rapid diffusion, it does not necessarily lead to usualisation and entrenchment (at least not in the same speed) due to its one-directional character which makes co-adoption rather impossible (cf. section 3.1.1). Hence, even if a new lexical item goes *viral*, not all people who come across it will also partake in its usualisation and eventually push its diffusion and usage.

Another explanation for the discrepancy between knowing and using a neologism can be found in the qualitative data. The main motives for people to use a new word are the relevance of the word in their life and its usefulness. Independent of age, speakers might know a word but do not find it applicable to their lives and thus, would not use it. It should also be stated that the age groups used were rather unequal in size and distribution. The oldest age group used in this study was 46 and older, hence a relatively large group considering that there might be considerable differences between the behaviour of a 46-year-old person and a 76-year-old person. However, hardly any participants were found for this age group. In contrast to this, many young people below 25 participated. This left me with many participants between 18-25 and little participation in the age range above 46. A more equal distribution amongst subject could have yielded different results.

Linked to the consumption of social media is the daily amount of internet use in general. However, only a slight trend can be detected in the data which shows that people who spend more time on the internet are also familiar with more new lexical items. An explanation could be that even if participants spend a lot of time on the internet, they might only use the same, or a very limited number, of sources available, which implies that they are less likely to come across new words. To exemplify this, if a speaker reads an online newspaper for more than two hours a

day, they have potentially less chance of encountering a neologism in contrast to participants who use social media, forums, and different news channels in the same amount of time. It seems logical that, since more time on the internet only shows a slight tendency towards a higher familiarity with neologisms, media consumption has no impact on the use of neologisms.

The fact that no distinction in education level and its impact on diffusion and usage can be made, could simply be due to the very broad tripartite categorisation of either having a school degree, a university degree or none of them. While this simplification is not ideal, it was the only way to ensure that participants from English-speaking countries all over the world could participate. Nevertheless, a more detailed distinction could have revealed more information about the connection between education and the diffusion of neologisms.

A distinction between city and countryside could also not be found. Since all words have been taken from the internet, and as rural areas have internet access nowadays, it might make sense that no significant difference is present. While people in cities are surrounded by more people and do hear and see more things on a daily basis, the internet provides the same amount of information to anyone, independent of their location.

All in all, one of the main issues within this study concerning the demographics of the participants is the unequal distribution of subjects amongst the different groups as well as the partially very broad categorisation, for instance when it comes to education (cf. section 6.1.3). Since the study's main focus did not lie on demographic and social predictors, it was regarded as acceptable to not have an equal distribution amongst participants. However, this also means that the outcomes have to be treated with caution and that if there was a more even distribution amongst participants, the outcomes could have been different and more precise.

Apart from demographic factors, the investigation of the variables that were used in order to choose the neologisms for the study revealed that frequency is an important independent and dependent variable when it comes to the diffusion and use of neologisms. While measuring frequency is tainted with issues, such as the type and token distinction or absolute versus relative frequency (cf. section 3.2.3), I have tried to consider possible frequency issues. The outcome of the study shows that the frequency assessed on the internet (cf. section 6.1.1) mirrors the frequency with which the neologisms are known amongst the participants of the questionnaire study. This makes the measured frequency on the internet appear to be a reliable variable.

Furthermore, frequent neologisms were more likely to be used. This goes in line with the underlying theoretical background of the EC-model (cf. section 3 & 4). More frequent words are already more diffused by definition and possibly also more usualised as well as more entrenched and can therefore be activated more easily. A higher conventionalisation (which frequent words exhibit) also goes along with increased institutionalisation and lexicalisation and thus a loss of ambiguity and a gain in clarity (cf. section 3.1.3). Thus, the more frequent a word, the more conventionalised and entrenched it is, the less ambiguous and hence the more likely it will be used. The qualitative analysis supported this and showed that participants are more inclined to use a clear and straight-forward neologism.

While frequency is an important indicator for the diffusion of new lexemes, the point in time of when a word first occurred did not reveal any significant impact. Considering that “[m]any successful words [...] lurked in the language for a long time before attaining general notice” (Metcalf, 2002: 164), makes this seem less surprising. Furthermore, extralinguistic events might spur on the diffusion of a word, so that even a relatively new term could gain relevance and can spread and diffuse quicker than an older term. Two previously mentioned examples are *bouncebackability*, which got artificially institutionalised by means of media promotion (Hohenhaus, 2006: 22), and *Brexit*, which is by now

much more well-known, used and established than its older base word *Grexit*. However, time and use of a neologism always has to be seen as being relative. Thus, one of the main issues with assessing the connection between time and diffusion/usage of new lexemes is its relativity. Time is also problematic insofar, as the self-referential feedback loops for conventionalisation and entrenchment, and its motor usage, constantly have to spin in order to keep a lexical item conventionalised, entrenched and used. Hence, even if a word has been around for a long time, it does not mean that the wheels have been permanently in motion. At the same time, a relatively 'young word' can have a lot of momentum and can have all wheels spinning constantly. As shown, the wheels of obsolete words spin rather slowly and might eventually completely cease despite the fact that the word might have been around for a longer time (cf. section 3.3).

When it comes to media, its prestige, or better quality, and the influence of the early user group do not exhibit a significant impact on diffusion and use. Even though the participants graded the different media types and their authors in significantly different categories, and the use of new words between these categories partially differed, no direct link could be detected. This is surprising, as various studies have found the prestige of the *innovator* or *coiner* can be influential (cf. Labov, 1980; Milroy, 1992; Kerremans, 2015).

There are several possible explanations for this. One potential problem is the unequal distribution within the data. For instance, only one out of the 24 new lexemes investigated was initially used in a scientific journal. In contrast to this, five neologisms were taken from broadsheet newspapers. Thus, although participants did agree on scientific papers having the highest quality and scientists having the greatest influence, the word taken from this category might not be a good representative and participants decide against using it for other reasons. This inequality can probably impact the reliability of the result. The cause for this is simply that it was impossible to find the same number of neologisms from each media type that also matched the other, decisive variables.

Furthermore, there might be a terminological and precautional problem involved. While participants ranked online dictionaries as high quality, they could have had in mind *The Oxford English Dictionary*, while the actual dictionary that was primarily used for this thesis was the *Urban Dictionary*. The fact that I refrain from giving the participants examples of the respective categories due to cultural bias (cf. section 6.1.2) led to an increased room for interpretation which might have altered the outcome. However, the fact that no significant connection between prestige and use was found in the quantitative data goes along with the qualitative data. Prestige, and with it the influence or quality of the early user group and media types was rarely stated as a reason why participants would use a new word. However, this could be due to the fact that it is not one of the aspects that comes to mind immediately and that might be quite subconscious.

While assessing the influence and quality of the media and its impact on use has proven to be quite tricky and this study was not free of flaws in its evaluation, it might be possible - especially in the context of the internet, where it is hard to determine the source – that prestige does not play such a great role. Diffusion, and eventually also use, can be promoted by *co-semiosis*, *co-adaption*, and *accommodation* (Giles et al. 1991; Trudgill, 1986, Labov, 1990, Schmid, 2020). All these three processes depict interpersonal processes that primarily take place in face-to-face interaction (cf. section 3.1.1). Thus, while they are often influenced and triggered by prestige, they might apply more to real-life interactions and spoken language rather than to virtual communication. This would contradict Kerremans' findings that the "coiner status celebrity is the most significant factor promoting conventionalisation" (Kerremans, 2015: 151) within her online sample. However, Kerremans' data set also contained counter examples. She found that some words were coined by person of status but nevertheless had a low frequency, in these cases the coiner status did not seem to promote conventionalisation. In fact, there were words of moderate to high frequency with unknown coiners. Therefore, extralinguistic factors or name-unworthiness might override

the coiner status (Kerremans, 2015: 151 f.). While the present study struggled to assess prestige as accurately as it should have, the fact that exclusively written language from the internet was used might be a reason for the lack of significance in the connection of prestige and use.

Another predictor variable that did not show any impact on diffusion and use of neologisms is interest. The assumption, that people who are interested in the area from which a word is taken would also participate in its diffusion, was not supported in the data. This outcome is surprising since participants often stated in the qualitative analysis that the relevance of a word for their life and the applicability would make them use it. Thus, while I would argue that relevance, applicability, and interest are linked, the data does not show this.

One reason for this disparity could be that Kerremans' *Page-level classification scheme* (Kerremans, 2015: 90 f.) was taken as the underlying categorisation for interest. While I denote these categories as fields of interest, Kerremans originally used them to classify different fields of discourse. Moreover, it could be that the classification was too broad for the purpose of the questionnaire study. While a participant might be interested in sports, they would not use *runger* ('the hunger you feel while running'), as they are not into running specifically. A more detailed division or even an open question might have helped to get a clearer picture. This, though, would have prolonged the questionnaire and would have been more suitable for qualitative rather than quantitative analyses.

Further independent variables are *recognisability*, *formal & semantic transparency*, and *formal, conceptual & general appeal*. The data showed that a distinction between the formal and conceptual appeal and the two types of transparency is hard to justify due to their high correlation values. For transparency, this would mean, that neologisms that are transparent in their form are usually also transparent in their meaning and vice versa. While a close link between formal and semantic transparency does seem quite straightforward, formal, and conceptual

appeal are defined as two very different entities. Simply because a speaker likes the form of a word, they might not necessarily like its concept/meaning and vice versa. However, the data implies exactly this. It is also possible though, that participants struggled with the wording, especially for formal appeal. The question 'do you think this word is a good way to say this?' potentially did not prove sufficient to test for formal appeal.

Once the two types of appeal and transparency have been summarised, the impact of the input variables on the output variable *use* was investigated. It was found that all of them – hence transparency, appeal, and recognisability - have a significant correlation with use. Therefore, if participants recognise a word, find it transparent and appealing, they are more inclined to use it and vice versa. These findings are also supported by the outcomes of the qualitative analysis. While familiarity with a word was not amongst the most mentioned motives for use, it was still specifically mentioned by different participants. This goes in line with Lehrer's findings about blends (cf. section 4.1.3). She assumed that words would get better ratings if they were recognised by the participants (Lehrer, 1996: 367). Her study revealed that people who recognised words did indeed rank them significantly higher, hence rated them as 'better' words (Lehrer, 1996: 382). This also coincides with the EC-model insofar, that processing unentrenched utterances, such as neologisms, is much more effortful and potentially slower. Thus, neologisms that are already recognised by speakers are more likely to be used as their activation is less effortful than that of completely new words (cf. section 3.2.1).

Besides this, the study showed that transparency has a positive impact on use. This is supported by the data from the qualitative analysis, where people state that they are more inclined to use a new lexeme when it has a clear meaning, when it is easily understood and therefore facilitates communication. A successful co-semiosis depicts a joint attempt between speaker and hearer and is one of the major aspects of communication (cf. section 3.1.1). Thus, it is not surprising that

transparent neologism, which facilitate successful co-semiosis, are more likely to be used. This is especially straight-forward in regard to semantic transparency. However, since hearers also use analogies and familiar word-formation pattern and processes in order to decipher and process words (cf. 3.1.3), formal transparency is also a facilitating factor for co-semiosis.

Statistically it was shown that appeal and use are also mutually dependent. In the qualitative analysis several subjects stated that they would use a word if they like the concept, if it is a good word to describe something as well as if it is creative, funny, witty - that is, if they find it appealing. Liking a concept might also be linked to the relevance and usefulness of it. It might be that participants like a certain concept as they perceive it as noteworthy, which consequently has a positive impact on conventionalisation (cf. section 3.1.1, 3.1.2, 3.1.3). In regard to the role of humour and its impact on the success of new words, different ideas have been brought forward (cf. section 3.4) The retrieved data supports the idea that humour has a positive impact on the diffusion of neologisms rather than a hindering one.

When analysing these three independent variables closer by bringing them into connection with each other, it was found that *appeal* is the most important predictor for *use* and that it serves as a mediator for the independent variables of *transparency* and *recognisability*. This means that if a new lexeme is recognised and transparent, it will be perceived as more appealing, which ultimately leads to more use. Therefore, it seems that while facilitating communication by using transparent and maybe even recognisable words is important, finding a word appealing is the ultimately decisive factor in the speaker's decision-making to use a new lexical item. Therefore, the main trigger for the establishment of a new word, or at least its usage, is very personal and subjective.

As it was assumed that formal and conceptual appeal are two different variables, the questionnaire also tested for general appeal.

However, the independent variable *general appeal* and the output variable *use* correlated so highly, that they cannot be distinguished. It seems that both questions, which were thought to test different things, did indeed test for the same thing. By asking whether the participants want the neologism in question 'to become a thing', I intended to test for general appeal, in the sense that the participants might not want to use the term themselves but feel that it should generally be used and accepted. The data does not allow this distinction, though. This could either be the result of a wording problem or it might be that only participants who would use a new lexical item themselves, would also consider it worthy to become a thing.

In sum, the questionnaire results did support some of the stated hypotheses but, by far, not all of them. In some cases, there are logical explanations as to why the hypotheses were not supported. Problems with the questionnaire design and its execution could have also played a role. The main issues were unequal distribution of subjects participating in the study, wording issues, overly broad categories and, partially, unevenly allocated representatives for certain categories. Nevertheless, the data did reveal several important outcomes which were supported by the qualitative part of the study. Such are that appeal is the main driving force as to why participants would use a new word and formal and conceptual appeal, as defined so far, cannot be distinguished. These outcomes will be used for further investigation in the second questionnaire study.

7. Questionnaire study II

7.1 Research object and questions

The predominate influence of appeal on the use of neologisms is the most interesting outcome from the first questionnaire study. Therefore, the second study was designed to complement, deepen, and extend the collected data. As the results from the previous study showed, no statistically relevant difference between formal and conceptual appeal was found. This implies that the way formal and conceptual appeal were tested in the first study was not precise enough to find a distinction between different types of appeal. Thus, the aim of this second study is to find a more detailed demarcation, to get a better idea what appeal exactly is and what categories it consists of. To find a more clear-cut subcategorization to analyse, two things were considered: the qualitative outcome of the first study as well as previous research on the different steps towards the successful establishment of new words. Therefore, in a first step suitable subcategories for appeal had to be found in order to allow a closer analysis.

Participants repeatedly stated that a reason for them using a new lexical item is because it is funny, creative, witty, and clever. It stands to reason that this implies that participants like words that 'stick out' and thus exhibit extravagance and salience. While extravagant utterances are salient per se, innovations are not necessarily salient but follow a gradation from salient to non-salient. Completely new creations, amongst them blend neologisms, are located on the salient end of the spectrum (Schmid, 2020) (cf. section 3.1.2; section 3.4). Salience by novelty furthermore impacts the entrenchment processes since speakers fail to match the new item up with the expectations that are activated from the long-term memory (Schmid & Günther, 2016: 2). Extravagance acts as a force that influences both, conventionalisation, and usage. However, extravagance and its perception depend on the speakers' linguistic experience, so that some speakers might regard a lexical item as

extravagant while others are already familiar with it (cf. section 3.3). The five maxims of action discussed in section 3.3. show that extravagance should be applied, and it is often intentionally chosen by a speaker, for instance, to achieve social success. Although extravagance alone is not enough to put usage and the flywheels of the EC-model in motion and keep them up and running, it certainly has a major impact on use, conventionalisation and also entrenchment. Thus, it is assumed that *extravagance*, characterised by a word being *funny*, *creative*, or a *pun* should be considered as a subcategory of appeal within the second study.

Having the opposite effect of extravagance, Keller's maxims of action suggests talking economically and clearly (cf. section 3.3). A great number of participating speakers stated in the first study that they were more inclined to use new lexical items if they were precise, accurate, descriptive, and had a clear meaning, and thus were efficient words. Since the ultimate goal of communication is to be understood and to achieve or at least facilitate co-semiosis, it is not surprising that participants prefer words that increase their chances to reach their communicative aim. This is often supported by using semantically transparent constructions (cf. section 3.3). It saves the speakers unnecessary effort since they do not have to go through the extra hassle of explaining the meaning of the respective new lexeme. Being understood is also one of the major premises of pragmatics. Habermas argues that in order to act communicatively, speakers have to utter something understandably, give the hearer something to understand, make themselves thereby understood, and come to an understanding with another person (Habermas, 1979: 2).

Efficiency can also be increased by analogy, including phonetic, morphologic, or semantic parallels. Novel analogical forms are easier to decipher and process than completely new ones (cf. section 3.3). New products of regular word-formation processes are usually also defined as less salient (cf. section 3.1.2). Thus, although it stands in direct contrast to extravagance, I decided to introduce *efficiency*, with the subcategories

of clear meaning, precise and similar to something that the speaker already knows/analogy, as a second subordinate factor that will be investigated for its suitability to describe appeal in more detail.

Furthermore, participants stated that the relevance of the word and its correspondence to the zeitgeist are important factors as to why they would use a new word. It is sometimes argued that the primary motives for new words are to describe and name a new reality, e.g., to reflect new innovations and progress of science, culture as well as changes in technology, political situation, social trends, etc. (Abu-Algasim Mohamed, 2020: 181). Novel lexical items are often formed based on themes and domains, the extralinguistic reality and audience-related factors (Lipka, 2007: 13). Since I assume that there is a connection between relevance and newsworthiness, and newsworthiness impacts conventionalisation (cf. section 3.1.2), I decided to use *extralinguistic relevance* as a third and last subcategory of appeal. This includes that the word is related to a *current trend, current event, an innovation* and *will be relevant from now on*. It should be noted that relevance in the qualitative part of the previous study did partially entail personal relevance. As this, however, is even more subjective than general extralinguistic relevance, I decided to not take it into account. *Table 10* summarises the subcategories of appeal that are based on the outcome of previous research and study I.

APPEAL		
EFFICENCY	EXTRAVAGANCE	EXTRALINGUSTIC RELEVANCE
Analogy	Funny	Recent trend
Precise	Creative	Recent event
Clear meaning	Pun	Innovation
		Relevance

Table 10 Subcategories of appeal based on the previous study and research

Given these three new subcategories, the hypotheses for the second questionnaire study are:

H1: Appeal can be subcategorised into the distinguishable groups of efficiency, extravagance and extralinguistic relevance.

H2: Neologisms that are more efficient, extravagant and extralinguistically relevant are more likely to be used.

Thus, while formal and conceptual appeal were statistically undistinguishable in the first study, the second study will investigate more detailed subgroups of appeal, based on the outcome of the previous study and theoretical foundations. Besides, the participants will also be asked about their demographics and, very briefly, about their time spent on the internet each day, in order to see whether some of the trends established in the first study (e.g. younger people know more neologisms), can be further supported. Since the qualitative data from the first study depicts an important pillar for the second one, one part of the questionnaire will also deal with verifying them.

7.2 Method

7.2.1 Materials – retrieving neologisms

As in study one, a first step in the design of the second study was to retrieve suitable words. Having already investigated 24 neologisms regarding their appeal and use relationship, I decided to choose a small sample of these words for the second study. Since this follow-up study had more time-consuming rating questions, only six neologisms and three distractors from the previous study were used. The deciding criteria were their indicated usage from study I (high versus low use) as well as their correlation between formal and conceptual appeal in study I (high versus low correlation). For both, the usage as well as the correlation, words with values at the extreme ends were chosen, which left me with the six neologisms depicted in *table 11* (which also shows the distractors that were transferred from the previous study).

	NEOLOGISM	USE	CORRELATION FORMAL AND SEMANTIC APPEAL
Low correlation – low use	HONEYTEER	0.08	0.340
	BRONGERIE	0.04	0.384
Low correlation – high use	CRUFFIN	0.26	0.272
	GLAMPSITE	0.66	0.524
High correlation – low use	BELFIE	0.03	0.674
High correlation – high use	OBLICATION	0.22	0.694
DISTRACTORS			
MOTEL			
BRUNCH			
SMOG			

Table 11 Neologisms chosen for the questionnaire study II

7.2.2 Questionnaire design and participants

Similar to the previous study, the second study started off with a short introduction part, which gave the participants background knowledge about the study and made sure that they were, in accordance with the ethical standards, informed about the nature of the study.⁷³ Next, in order to ensure that the participants did not take part in the previous study and hence already know the presented words, they were asked to indicate whether they have taken part in the first study. This was followed by demographic questions, which exactly match the ones from the previous study. The next part depicted a shortened section about the participants' media usage.

The main part of the questionnaire can be split into two parts. The first part served to find out more about the appeal of the words. The participants were presented with one neologism at a time (without

⁷³ A sample of the second questionnaire study can be found in appendix 2.1

context) and - just like in the previous questionnaire – were asked whether they have come across this new lexeme before. Then they were asked whether they knew the meaning of the word, and if yes, what the meaning is. Subsequently, the questions from study I for formal appeal ('is the word a good way to say this') and conceptual ('do you like the idea/concept') appeal were queried, but rather than a yes/no question they had to be rated on a Likert scale from zero (not at all) to six (absolutely). The newly established subcategories of appeal had to be rated in the same way. For extravagance the participants were asked to rank whether the word is funny, a pun or creative, for efficiency, whether it has a clear meaning, is precise or is similar to something they already know. Lastly, for extralinguistic relevance they were asked whether the neologism depicts a new trend, an innovation, a current event or will be relevant from now on, followed by rating how likely they are to use this new lexical item.

The second part of the study was of a more qualitative, observational nature. While the investigation of appeal is, to a great extent, based on the qualitative outcomes of the previous study, not all reasons for use stated are incorporated in appeal. Thus, the second part of this questionnaire study also included these factors and asked the participants to state (via drag and drop) whether they found any of the presented neologisms socially as well as personally relevant, relatable, useful, familiar, funny, descriptive, facilitating communication, clear and conceptually good. While some attributes of appeal were included here, I was aiming at getting a better understanding of the weighing of these for each individual neologism. Thus, I wanted to get a clearer picture of how the factors for usage stated in the qualitative part of study I describe the individual neologisms.

The drag and drop format implied that participants could chose as many words as they wanted for each question. For example, they were asked 'are any of these words relevant to you' and they could then move as many words as they wanted from the left column to the right column, to indicate their choices (cf. appendix 2.1). There were two reasons for

deciding to use drag and drop: firstly, it served the objective to not prolong the questionnaire, since open questions are usually more time consuming to answer. Secondly, by changing from Likert scale assessment to drag and drop I hoped to break up the monotony of the questionnaire and keep the questionnaire as short as possible and thus keep participants engaged. 143 people participated, excluding those who have participated in the previous study (cf. tbl. 12).⁷⁴ It should be noted that since this study made use of a much smaller sample set of words, no reliable assumptions could be made when it comes to the participants' demographics and their influence on the use of new words.⁷⁵

	School		University		
	female	male	Female	male	
City	15	4	43	22	<25
	0	1	8	11	26-45
	1	0	2	1	>46
Countryside	4	0	15	5	<25
	0	0	3	0	26-45
	0	0	2	0	>46

Table 12 Distribution amongst participants in questionnaire study II

7.2.3 Confirmatory Factor Analysis (CFA)

In order to analyse the data retrieved from the questionnaire study, a confirmatory factor analysis (CFA) was used. CFA is a type of structural equation modelling which deals with measurement models, namely the relationship between observed measures (*indicators*) and latent

⁷⁴ For a complete overview of the demographics in questionnaire study II, including participants who did not want to provide information, refer to appendix 2.2

⁷⁵ The respective statistical non-significant calculation outcomes can be found in appendix 2.3

variables (*factors*). In contrast to exploratory factor analysis (EFA), CFA is used when the researcher already has a firm a priori sense of the number of factors that exist in the data as well as which indicators are related to which factors based on previous research and results (Brown, 2015: 1). Since a presupposition about the factors (extravagance, efficiency and extralinguistic relevance) as well as their respective indicators (such as funny, creative, pun for extravagance) was made, using CFA was appropriate. While the hypothesis at hand is that appeal can be distinguished in efficacy, extravagance and extralinguistic relevance, study I also suggests the possibility of a one factor model (since no difference between formal and conceptual appeal could be detected). Thus, several steps and models were used to investigate whether the factors found differ from each other and whether the assumed model works. These steps will be explained in more detail in section 7.3, along with the results.

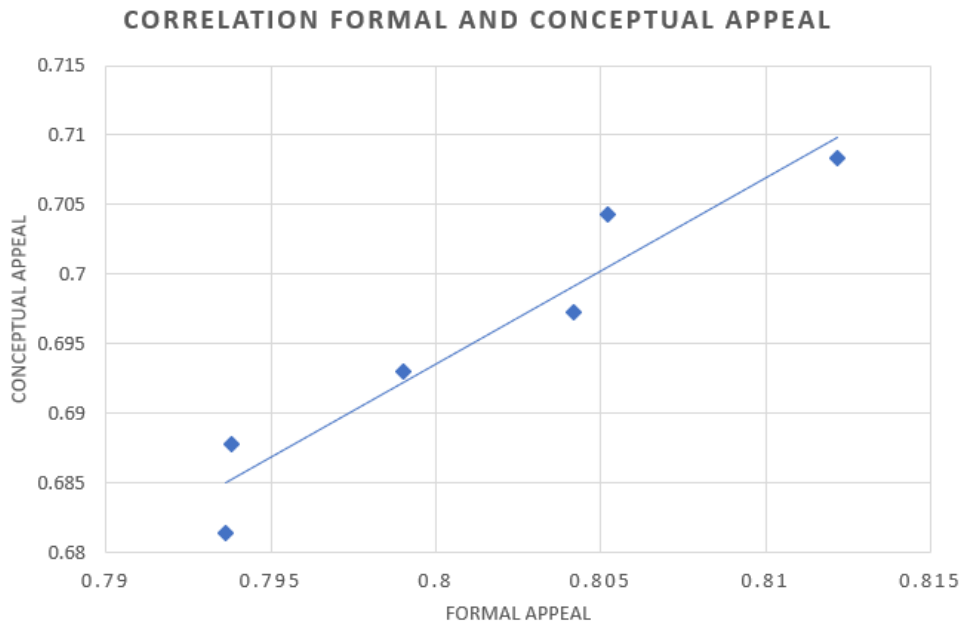
7.3 Quantitative results for appeal

The first study revealed that formal and conceptual appeal, in the way they were assessed, are hard to distinguish in their impact on use, which can be seen from the outcome of the regression calculation. This was underpinned by their high correlation value (cf. section 6.2.5). The questions for formal and conceptual appeal from the first study were reused (this time as a Likert scale) and a similar observation was made. Putting the two factors in connection with use in a regression showed, that in this case formal appeal seems to be a stronger indicator, but a differentiation between the two of them regarding their impact on use continued to be difficult (cf. fig.13). This result was once more reinforced by an extremely high correlation between the two [$r(6)=.960$, $p=.002$] (cf. grf. 16), showing that the way appeal was investigated so far is not detailed and sufficient enough to make clearer assumptions about what appeal entails.

		Coefficients ^a										
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Correlations			Collinearity Statistics		
		B	Std. Error	Beta			Zero-order	Partial	Part	Tolerance	VIF	
1	(Constant)	-14.930	3.227		-4.627	.019						
	formal_appeal	32.855	9.190	2.371	3.575	.037	.834	.900	.663	.078	12.778	
	conceptual_appeal	-15.918	6.596	-1.600	-2.413	.095	.676	-.812	-.448	.078	12.778	

a. Dependent Variable: use

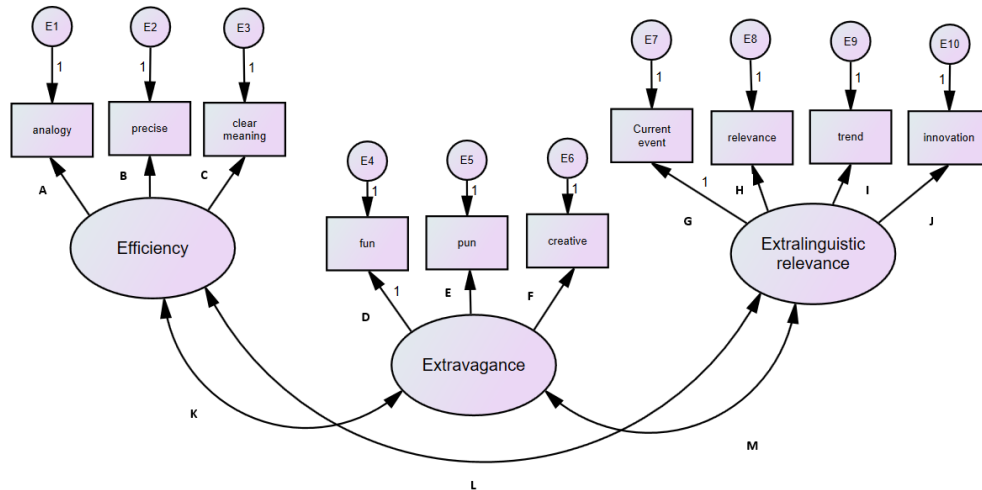
Figure 13 Regression result for use, formal, and conceptual appeal from study II



Graph 16 Correlation between formal and conceptual appeal in study II

Therefore, in order to get a clearer picture of what appeal is, it was subcategorised into extravagance, efficiency and extralinguistic relevance. To find whether this new subcategorisation is more suitable to describe appeal as well as to see whether they are distinct factors and whether they have an impact on use, a CFA was used (cf. section 7.2.3).

As a first step, I put all factors and their indicators into a CFA, exactly the way they were structured in the questionnaire study. Thus, the CFA model shows the factor *efficiency*, with its indicators *analogy*, *precise*, *clear meaning*; the factor *extravagance* with the indicators *fun*, *pun* and *creative*, and the factor *extralinguistic relevance* with its indicators *current event*, *relevance*, *trend*, *innovation* (cf. grf. 17).



Graph 17 CFA (I)

The outcome of a CFA gives several values and indications: the chi-squared test shows the difference between the observed and expected covariance matrices, thus compares the fit of the assumed model to the data (Bryant & Satorra, 2012). The closer this value to zero, the better the model fits. A good model is indicated by a chi-square value rejecting the model, thus with an insignificant p-value. Another indicator for how well a model describes the data in a CFA is the Root Mean Square Error of Approximation (RMSEA) value. It ranges from zero to one, with smaller values indicating better model fit. It is commonly agreed on that a value of .06 or less is indicative of an acceptable model fit (Brown, 2015: 74). By analysing the discrepancy between the hypothesised model, with optimally chosen parameter estimates, and the population covariance matrix, it helps to evade problems with sample size (Hooper et al., 2008: 54). Lastly, the Akaike Information Criterion (AIC) value is useful as a comparative measure if two different models are compared. The lower the AIC, the better the model fit (Bowen & Guo, 2012: 155).

Table 13 shows these four values for the model presented in graph 17, both for all words taken together as well as for each single word. As it can be seen from the values, the overall model fit is quite poor, with a high chi-square (95.1), a significant p-value (<.001) as well as high RMSEA (0.118) and AIC (141.053). Thus, my assumption that the factors and their respective indicators are matching the data was rejected.

To get an insight into which indicators poorly match their factors, I looked at the standardised estimates of the factor loadings of the different paths between factors and indicators. These are labelled with the letters A-M in *graph 17*. The size of the factor loadings determines whether all indicators are reasonable measures of their latent construct, thus the factors. Depending on the empirical context, there are different opinions on what a sufficiently large parameter estimate is (Brown, 2015: 115). In this thesis, I go along with researchers claiming that factor loadings of .70 or above are considered as *good* measures, loadings between .50 to .69 are regarded as *fair* and loadings below .50 are perceived as *poor* measures (Lichtenberger & Kaufman, 2013: 44). To exemplify this: *table 14* shows that the factor loading A (analogy on efficiency) for *belfie* is .44 and B (precise on efficiency) is .92. The first observed variable *analogy* thus explains $(.44^2=0.19)$ 19% of the latent variable variance for *efficiency*. In contrast to that, *precise* explains $(.92^2=0.85)$ 85% of the factor variance *efficiency*. Therefore, *precise* is a more reliable indicator for efficiency than *analogy*. Hence, *table 14* reveals that the two weakest indicators in the model are *analogy* (A) and *innovation* (J).

	CHI-SQ	RMSEA	AIC	p-value
OVERALL	95.1	0.118	141.053	<.001
BELFIE	98.1	0.121	144.116	<.001
BRONGERIE	81.5	0.095	121.118	<.001
CRUFFIN	79.3	0.102	125.336	<.001
GLAMPSITE	72.9	0.095	118.885	<.001
HONEYTEER	80.5	0.103	126.455	<.001
OBLICATION	175.7	0.178	221.723	<.001

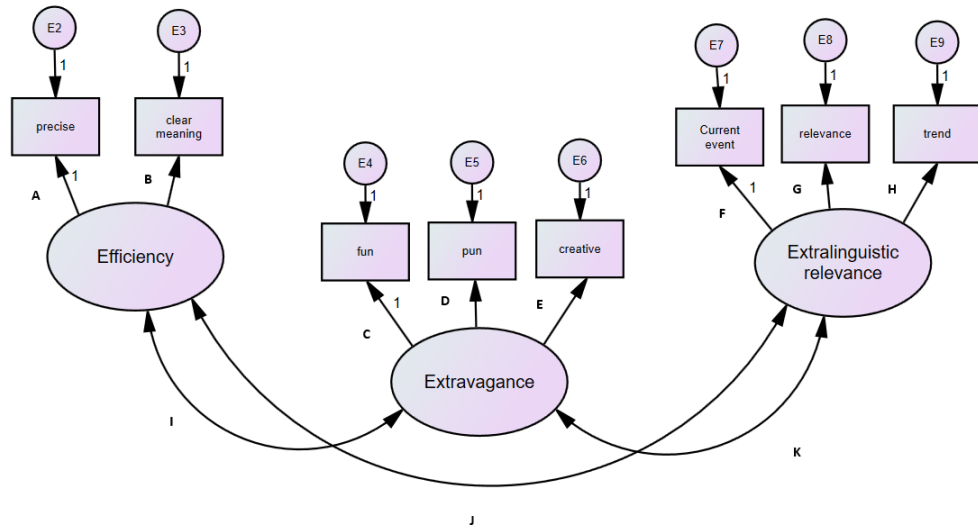
Table 13 Values indicating model fit for CFA (I)

	A	B	C	D	E	F	G	H	I	J	K	L	M
BELFIE	.44	.92	.86	.88	.67	.83	.42	.82	.47	.81	.80	.36	.42
BRONGERIE	.45	.95	.89	.82	.78	.92	.67	.36	.73	.55	.73	.65	.70
CRUFFIN	.32	.87	.89	.30	.73	.91	.43	.88	.78	.50	.78	.66	.78
GLAMPSITE	.07	.71	.88	.64	.46	.87	.52	.68	.78	.74	.25	.29	.52
HONEYTEER	.65	.96	.83	.85	.63	.83	.74	.90	.78	.31	.71	.68	.70
OBLICATION	.14	.79	.85	.34	.57	.78	.32	.91	.84	.10	.70	.68	.68

Table 14 Standardised factor loadings for CFA (I) depicted in graph 17

Based on this, a second CFA without these two indicators was created (cf. grf. 18). Table 15 shows, that when comparing the previous, full CFA I including analogy and innovation - to the reduced CFA II, the values changed drastically. The overall values in table 15 show that the chi-square of the revised analysis is low with 23.8. Moreover, the p-value of .125 shows insignificance and thus implies a good model fit. The RMSEA is, with a value of .053, below .06, the value used as a threshold to indicate good model fit. The AIC with 61.809 is also drastically lower than in the previous model. This means, that the reduced CFA (II) describes the collected data much better than the previous model.⁷⁶ Hence, while I assumed that analogy is a good indicator to describe efficiency, and innovation is suitable for extralinguistic relevance, the data does not confirm this.

⁷⁶ For the standardised factor loadings for CFA II, please refer to appendix 2.4.



Graph 18 CFA (II)

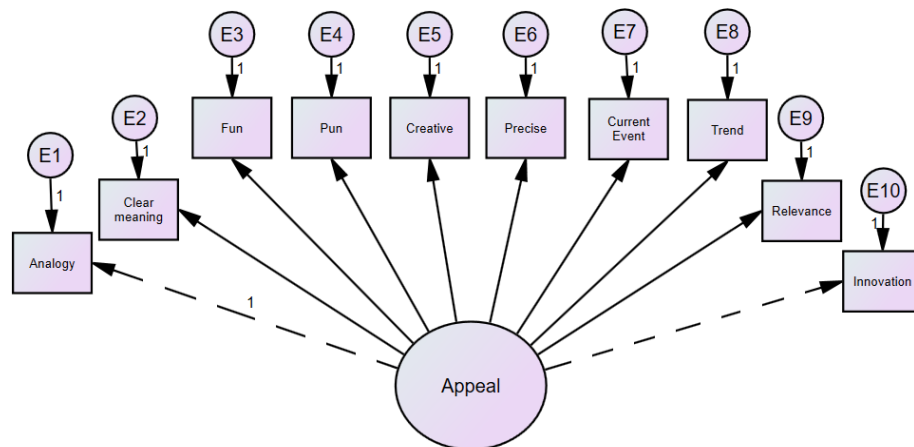
	CHI-SQ (full)	CHI-SQ (reduced)	RMSEA (full)	RMSEA (reduced)	AIC (full)	AIC (reduced)	p-value (full)	p-value (reduced)
OVERALL	95.1	23.8	0.118	0.053	141.053	61.809	<.001	0.125
BELFIE	98.1	20	0.121	0.035	144.116	58.008	<.001	0.274
BRONGERIE	81.5	42	0.095	0.102	121.118	79.998	<.001	0.001
CRUFFIN	79.3	31.7	0.102	0.078	125.336	69.735	<.001	0.016
GLAMPSITE	72.9	19	0.095	0.029	118.885	56.988	<.001	0.329
HONEYTEER	80.5	37.8	0.103	0.093	126.455	75.754	<.001	0.003
OBLICATION	175.7	39.1	0.178	0.096	221.723	77.066	<.001	0.002

Table 15 Values indicating model fit CFA I in comparison to CFA II

While the revised CFA offers a model that describes the data well, it should be noted that not each neologism reacts equally well to the model. While *belfie* and *glampsite* are perfect fits, with extremely low chi-squares of 20 and 19, insignificant p-values of 0.274 and 0.392, as well as RMSEAs way below 0.06 (0.035 and 0.029), the data for other words is not captured in the same reliable way by the model. *Brongerie*, for instance, still exhibits a high chi-square of 42, a significant p-value of 0.001, as well as an RMSEA above 0.06 with 0.102. Nevertheless, even for neologisms like *brongerie*, the reduced model fits much better than the previous, full model. Therefore, CFA II establishes three factors with

their respective indicators, that describe the collected data well and therefore are suitable factors to specify appeal.

So far, the data does not yet give any indication yet as to whether the three factors can be distinguished. While the questionnaire made the assumption of a tripartite factor division, there might be other options, too. Since all indicators are related to appeal, another approach could be a one factor model, with all indicators directly loaded on appeal (both previously mentioned models were taken into account, in- as well as excluding *analogy* and *innovation*). *Graph 19* shows a one factor model that does not include the division between extravagance, efficiency and extralinguistic relevance. The results in *table 16* show, however, that this model is a very poor fit for the data. Both, the reduced as well as the full model exhibit extremely high chi-squares, RMSEAs way above the benchmark of 0.06 as well as high AICs and insignificant p-values. Thus, comparing the results from *table 16* to *table 15* shows that the tripartite system is much more suitable to describe the collected data. This indicates that the three factors of extravagance, efficiency and extralinguistic relevance are distinct from each other.

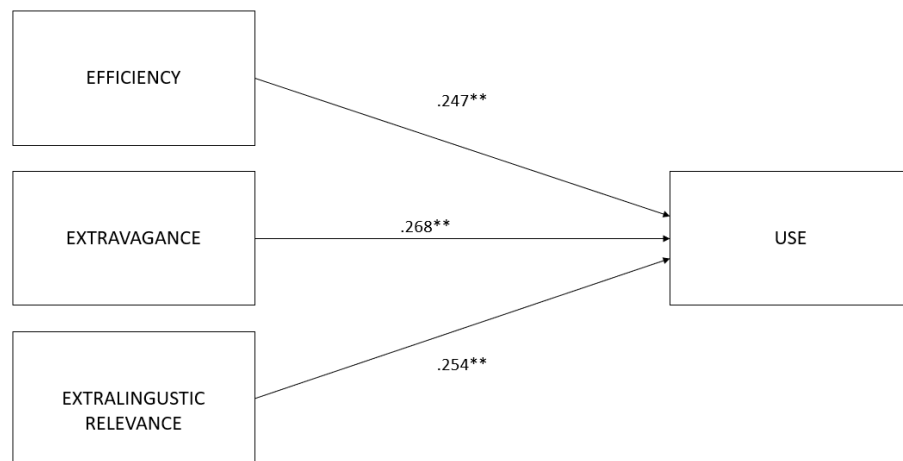


Graph 19 One factor CFA

	CHI-SQ (full)	CHI-SQ (reduced)	RMSEA (full)	RMSEA (reduced)	AIC (full)	AIC (reduced)	p-value (full)	p-value (reduced)
OVERALL	193.5	149.7	0.179	0.146	233.504	112.747	<.001	<.001
BELFIE	206.5	80.7	0.182	0.151	246.540	199.514	<.001	<.001
BRONGERIE	198.2	151.9	0.181	0.215	238.204	183.857	<.001	<.001
CRUFFIN	165.5	115.2	0.162	0.183	205.516	147.242	<.001	<.001
GLAMPSITE	180.3	128.9	0.171	0.196	220.337	160.936	<.001	<.001
HONEYTEER	216.6	167.4	0.191	0.228	256.608	199.365	<.001	<.001
OBLICATION	230.1	98.3	0.198	0.166	270.063	130.343	<.001	<.001

Table 16 Values indicating model fit for one factor CFA

While the data up to now shows that three different factors can be distinguished in order to describe appeal more precisely, it does not yield any information about their impact on use. Thus, as a next step the three factors - extravagance, efficiency and extralinguistic relevance - were put in context with use. A regression analysis revealed that efficiency [$B=.247$, $p=.008$], extravagance [$B=.268$, $p=.005$] and extralinguistic relevance [$B=.254$, $p=.006$] all have a significant independent impact on use (cf. grf. 20)⁷⁷ and, with an $R^2=.465$, explain 46.5% of variances.



Graph 20 Impact of efficiency, extravagance and extralinguistic relevance on use for questionnaire study II

⁷⁷ For data for the average ratings for efficiency, extravagance, extralinguistic relevance and use, see appendix 2.5

7.4 Qualitative results

The aim of the second part of study II was to go beyond the appeal rating of the words and gain more information about how the chosen neologisms can be described best. Hence, the idea was to go past the pure statistical side and try a more descriptive approach. To do so, the participants had to drag and drop the respective neologisms into boxes, denoting them as useful, relatable etc. *Table 17* shows how many percent of participants have dragged and dropped which words in which field. The outcomes differed quite drastically amongst the different neologisms. *Glampsite*, for instance, is not perceived as an extremely funny word, but as being descriptive, relevant – both for society as well as for the single participants – and as having a clear meaning. Further, 70% of participants were familiar with the word and one third of participants stated that it is already in use according to their experience. While one fourth of the participants were familiar with *oblication* and the majority thinks that it is a relevant word for our society, all the other categories receive rather low scorings. *Cruffin* sticks out by being a good concept and a witty word. Nearly 40% of participants perceive this word as funny. It is furthermore regarded as useful. *Honeyteer* behaves similar, just with lower percentages. *Belfie* is primarily seen as being funny but does receive little scoring in all other categories. Lastly, *brongerie* is perceived as a good concept by about 20% and relevant for society by 1/3 of the participants.

When looking at the overall outcome, a general trend can be attested. The words in *table 17* are sorted by their use ratings, with *glampsite* on the top, which received the highest use rating, and *brongerie* at the bottom, with the lowest use rating (cf. appendix 2.5). Within most categories, the scoring of the first three words is often higher than that of the last three. While 73% perceive *glampsite* as a clear word, only 6% state the same for *brongerie*. 76% think that *glampsite* is relevant for society, in case of lower usage words like *belfie* it is 27%. There are exceptions to this, for example, *belfie* is perceived as funny by 31% of

participants, whereas *glampsite* scores lower with 27%. Thus, this shows that every word bears its very own characteristics, so that some are funny and stick out in contrast to more serious words, that might focus more on facilitating communication. When looking at the percentages overall, it becomes apparent that relevance for society is something that is attested to most words in relative high percentages. In contrast to that, personal relevance is usually one of the minor criteria. It seems that, although a word might be perceived as relevant for society, it is not necessarily regarded as useful or relevant on a personal level. Moreover, apart from *glampsite*, all of the words are still relatively little established in the English language.

NEOLOGISMS	personally relevant	relatable	useful	already in use	Good concept	funny	facilitates communication	familiar	descriptive	Clear meaning	Relevance for society
GLAMSPITE	51.75	39.16	48.95	29.37	47.55	26.57	38.46	69.23	71.33	73.43	76.22
OBLICATION	15.38	11.19	10.49	7.69	12.59	20.28	6.99	26.57	20.98	14.68	65.03
CRUFFIN	23.78	27.97	29.37	2.10	48.25	37.06	16.08	4.89	25.87	12.59	48.25
HONEYTEER	18.18	13.99	9.79	4.19	25.17	21.68	8.39	13.99	27.27	10.49	27.27
BELFIE	6.99	4.19	5.59	2.10	12.59	30.77	3.50	2.10	16.08	6.99	27.27
BRONGERIE	6.99	6.29	6.99	2.10	19.58	7.69	5.59	2.10	14.68	6.29	35.66

Table 17 Percentages of how neologisms are perceived by participants

7.5 Discussion

The study confirmed the findings of the first study, namely that formal and conceptual appeal, the way they were measured and assessed so far, were indistinguishable. However, while the first study failed to find suitable categories to assess appeal in a more detailed way, the second study succeeded by finding a tripartite system of efficiency, extravagance and extralinguistic relevance. Thus, these three factors, with their respective indicators, are suitable to describe appeal in a more detailed and zoomed-in way. However, the preliminary indicators had to be adapted slightly, so that *analogy* was deleted as an indicator for *efficiency* and *innovation* as an indicator for *extralinguistic relevance*.

There are different ways to interpret the unsuitability of these indicators. In regard to analogy, one problem is that some of the existing words do not have any similar, existing words that can be used as analogical references. Thus, while for instance a neologism like *Megxit* is formed in analogy to the existing word *Brexit*, the same cannot be said for the majority of chosen neologisms in this study. An exception is *glampsite*, with its analogy to the well-established word *campsite*. Thus, analogy depicted an indicator that measured for something that did not exist for most of the investigated neologisms. Furthermore, analogy was attested for by asking the participants whether the presented novel lexeme is similar to a word they already know. Despite the fact that it is often argued that similarity increases the success of a new word, as the word is easier to decode (cf. section 3.4) and can affect entrenchment (section 3.2.3), the similarity to an existing word and thus analogy, could also lead to misunderstandings. *Obligation*, for instance, was sometimes regarded as a typo of *obligation*. It can be assumed that in writing this should still be less of a problem than in spoken language. Thus, if it already leads to such difficulties and misunderstandings in writing, it is hard to imagine that a successful co-semiosis can be established when using *obligation* in conversation. Therefore, while analogy might work for some neologisms as a pushing factor for an increased usage, it does not

seem to be an extremely reliable guarantor. Further, the factor *efficiency* could have potentially profited from adding the component of morphological clarity to the semantic clarity, attested by clear meaning and preciseness of the word.

Innovation does similarly not seem to be a reliable factor, possibly due to the fact that some of the tested new lexical items rather describe an abstract construct than an innovation, which made testing *innovation* as an indicator problematic. Although the same can also be said for other indicators, like *event* or *trend*, it seems that with *innovation* this depicts a particularly great problem. Hence, one of the main weaknesses of this study was that partially the categories tested did not correspond to the reality of the chosen words. In hindsight it does not seem too surprising that, by excluding these two variables, a model was found that describes the collected data better.

Another, rather general issue with questionnaire studies, that both of my studies were not immune against, is the room for interpretation. In a debrief with a random sample of participants, it was pointed out that some participants did not see a difference between *clear meaning* and *precise*. In my understanding, *clear meaning* refers to the fact, that the meaning is easily understood by the recipient. In contrast to that, *precise*, which refers to whether the word precisely describes what it is meant to describe. However, interpretations of this kind are always subjective and therefore depict a potential danger for questionnaire studies.

The debrief also showed that the variable *use* is quite broad. A possible distinction between *use in real life* as well as *use on the internet, on social media* or *in a hashtag* could have potentially altered the results. Participants might be inclined to use a neologism online but not so much in real life and spoken language, either for stylistic reasons, or because it might be easier to make themselves understood in writing, as metalinguistic information can be added and obstacles like pronunciation, especially with blend neologisms, are not present (cf. section 4.1.2).

The fact that the three subcategories used to describe appeal all have a significant impact on use - meaning that the more efficient, extravagant and extralinguistic relevant a word is, the more likely speakers are to use it – depicts the second important outcome. Whereas the first study already indicated a significant impact of appeal on use, the second study succeeded in a more detailed description of appeal, with each of these categories still significantly impacting the output variable *use*.

This result goes along with previous findings. Within the EC-model, extravagance acts as an outside force that influences conventionalisation, usage and partially also entrenchment (cf. section 3.1.2 & 3.3). Similarly, efficiency, a guarantor for more successful co-semiosis and communication, depicts another force influencing usage (cf. section 3.3). Lastly, extralinguistic relevance, which is mirrored in nameworthiness, promotes the diffusion of a new word and with it its conventionalisation, which further leads to a higher frequency and eventually a higher usage (cf. section 3.1.1 & 3.3). Therefore, it can be said that the empirical findings of this study complement the theoretical background of the EC-model.

The observational part of this study was meant to supplement the quantitative analysis by considering some factors outside of appeal. The results show that different neologisms are perceived differently, some are rather funny and stick out in contrast to others that are less ‘shiny’ but might facilitate communication. Some words were perceived as very relevant for society but at the same time not automatically relevant for and relatable to the speaker personally. The word ratings gave a more detailed description of each neologism.

However, retrospectively it has to be said that the outcome lacks novelty. *Glampsite*, for example, exhibited higher ratings in all areas of the quantitative analysis, extravagance, efficiency, extralinguistic relevance and use. Thus, while the quantitative analysis can be backed-up by the qualitative one, the idea of the qualitative analysis was to get

new insights into the words, which was not achieved. Furthermore, using the drag and drop system allowed me to provide overall percentages for the different categories and words, but it made the extraction of the data much more difficult in a way that I could not easily assess it on the participant level. Thus, while the first part of the study can be seen as well designed and has revealed good outcomes, the second part has been designed poorly and, while it certainly mirrors the results from the first part, no new insights can be taken from it.

8. Questionnaire study III

8.1 Research object and questions

After having established a model that describes the major independent variable for use - appeal - in more detail, I was interested in finding whether this model can also be adapted to not only other speakers, but to speakers from another Germanic language, namely German. There are two reasons for choosing German. The fact that it is my native language made it a rather obvious decision. However, also linguistically there is good reasoning for this choice, as in German linguistics the term *neologism* still suffers problems of demarcation, and definition issues (Braungart et al., 2010: 699). In the mid-90s, the *IDS (Institut für deutsche Sprache)* decided to tackle this deficit by establishing a neologisms dictionary (Herberg, 2002: 195). However, till this day, there is a lack of systematic research of neologisms in German, and the dictionaries that exist these days as well as the monographies about neologisms are often of exemplary nature (Braungart et al., 2010: 699). Thus, the third study transfers a major part of the second study into German, hypothesising:

H1: In German, appeal can also be subcategorised into the distinguishable groups of efficiency, extravagance and extralinguistic relevance.

H2: In German, neologisms that are more efficient, extravagant and extralinguistic relevant are more likely to be used.

Besides these two hypotheses, the German questionnaire is also meant to answer some others. By far not all German neologisms have their root in German. In 2017, an outcry went through the German media as 5000 new words were added to the German reference dictionary *Duden*, amongst which many derived from English (Felder, 2017; Haentjes, 2017). Considering that English words can easily be assimilated regarding their writing, pronunciation, and flexion (Herberg, 2002: 198),

it is not surprising to find many Anglo neologisms in German. Amongst linguists as well as in the media there is an ongoing debate about the pros and cons of Anglicisms in German (cf. Burkhardt, 2013; Pidd, 2011). The mentioned media outcry suggests that not everyone is pleased about the adaption of Anglicisms into German. Since younger people grew up in a globalised, internet dominated world, in which English became the predominantly used Lingua Franca (Polyudova, 2014: 16; Crystal, 2003), it is assumed that English is somewhat more natural to young people, and Anglo neologisms come easier to them, both in regard to memorising and knowing them as well as using them. Vice versa, old people often struggle with Anglicisms in general (Hanisch, 2018: 152). Similarly it is assumed that highly educated people, who are assumed to have a better level of English due to their longer exposure to the language at school might know and use more Anglicisms. Besides this, speakers who live abroad, in a non-German speaking country, are very likely to be bilingual. Even though the second language might not be English, they are maybe in general more receptive to foreign languages and exhibit an increased flexibility regarding language use (Ferlazzo & Sypniewski, 2012: 252). Thus, I hypothesise that the demographic background of the participant might have an impact on the use and diffusion of Anglo neologisms:

H3: Anglo neologisms are more diffused and more likely to be used amongst speakers who are younger, more educated, live in cities and outside of a German speaking country.

Furthermore, the trends found in study I regarding use, demography, and internet usage (cf. section 6.2.1), will also be investigated.

8.2 Method

8.2.1 Materials – retrieving neologisms

Conducting a study in German meant that, as a first step, new lexical items had to be found. In contrast to English, where blend neologisms are very frequent, blending is not a widely used word formation pattern in German (Lohde, 2006). Since compounding is favoured, the German questionnaire study focuses on N+N compounds instead (cf. section 4.2). Further, since one of my hypotheses evolves around Anglo neologisms, I did not only extract German noun compound neologisms, but also Anglo neologisms. After having retrieved a list of preliminary words from OWID (*Online Wortschatz Informationssystem Deutsch*) and *Wortwarte* (Lemnitzer, 2011)⁷⁸, I counterchecked them in the *Duden*, the main German reference dictionary. Since, so far, the only distinguishing criterion between the words was their origin, English or German, another differentiation criterion was needed. The neologisms in the first study differed regarding their frequency, first occurrence and the prestige of the early user group/media types. The results did only indicate that frequency has a significant impact on the use and diffusion of neologisms and, despite the problems frequency brings along, I decided to take this as a second criterion, complementing the distinction between Anglo and German neologism.

In order to gain information about the frequency of the respective neologisms, different tools were used. As a first step, I put the words in Google's customised search tool, starting from 01.01.2012 – the same date when the NeoCrawler started crawling. Two things were counted: on the one hand the total number of websites on which the words occurred, on the other hand the total page number Google provided. However, as Google cannot be regarded as a completely reliable tool (cf. section 5), it had to be accompanied by others. A media monitoring tool

⁷⁸ For the full list of preliminary chosen words, see appendix 3.1

called *brand24*⁷⁹ complemented the Google search. This tool crawls the internet – thanks to customised search for the same time span as I used in Google - and especially focuses on blogs and social media. This was supplemented by data from the *Corpus Search, Management and Analysis System* (COSMAS II), provided by the IDS. Since this software crawls newspapers but not social media, it seemed like a good addition to the other tools. Two things were important: first, only words that exhibit a relatively equal frequency throughout all four tools were considered for the study. Secondly, words on the extreme spectrums of frequency were chosen. This led me to the following twelve words⁸⁰, which are split in German and Anglo neologism with varying frequency (cf. tbl.18).

	Anglo neologisms	German neologisms
High Frequency	Foodporn Foodtruck Clickbait	Schwarmstadt Willkommensklasse Ghettofaust
Low Frequency	Porkday Egosurfer Imageboost	Trinktourismus Bleistiftstemmer Digitalfalle

Table 18 Chosen neologisms for study III

A complicity is the fact that Anglo neologisms exhibit much more frequency in the analysis of social media, blogs and forums (*brand24*). Considering that in early 2020, the most used hashtags in Germany were taken from English (Bauer, 2020), it seems that Anglicisms are much more ‘hashtagable’ than German words, which explains their high-frequency values in this category (cf. tbl. 19).

⁷⁹ <https://brand24.com/>

⁸⁰ The definitions of the chosen neologisms can be found in appendix 3.2

	Google	Google Pages	Brand24	COSMAS II	altogether
FOODPORN	210	40	288	116	654
FOODTRUCK	262	41	346	491	1140
CLICKBAIT	223	39	261	41	564
PORKDAY	4	1	0	0	5
EGOSURFER	14	2	0	4	20
IMAGEBOOST	50	6	0	4	60
SCHWARMSTADT	200	26	8	102	336
WILLKOMMENSKLASSE	210	34	19	326	589
GHETTOFAUST	173	31	56	38	298
TRINKTOURISMUS	28	4	0	3	35
BLEISTIFTSTEMMER	20	4	2	3	29
DIGITALFALLE	20	4	0	0	24

Table 19 Frequency values for the chosen neologisms for study III

8.2.2 Questionnaire design and participants

As this study was meant to be as similar as possible to the second study, its outline is nearly identical. As a first step, I decided to split the questionnaire into two equal studies (like in study I), due to the high number of tested words. Thus, each of the two questionnaires contained six neologisms as well as three distractors. Similar to the previous studies, the distractors served as control words and were meant to keep the participants interested. For the distractors, I decided to not only use German words, but also foreign ones in order to ‘hide’ the Anglo neologisms better. Hence, the six distractors are composed of two relatively new but established German words, two established Anglicisms as well as two Italianisms/Gallicisms. Since Romance language make use of different word formation patterns than Germanic languages, and since compounding is not amongst the most common types of word formation, it was extremely hard to find N+N compounds that are in use in German (Ledgeway, 2016: 514). Therefore, two well-established ADJ+N compounds were used instead (cf. tbl. 20).

QUESTIONNAIRE I	QUESTIONNAIRE II
QUIZMASTER	SCHULDENBREMSE
DOLCE VITA	HAUTE CUISINE
WUTBÜRGER	ROADTRIP

Table 20 Distractors study III

Structurally, the questionnaire was a mirror image of the previous one, just in German.⁸¹ After the introduction, which provided participants with some background knowledge about the study and ensured that they were informed about the nature of the study in accordance with the ethical standards, participants were presented with some demographic questions, dealing with age, gender, education, and location. Since the school system in all German speaking countries is relatively similar, I could subcategorise education in much more detail than I was able to do in the previous questionnaire studies. Thus, in contrast to only two categories previously, education was now split into six different degree classifications. Another new question within the personal data part is the country of residence, aiming at retrieving information about how the country of residency influences the perception of Anglo neologisms versus German neologisms. This was followed by asking the participants about the average time they spend on the internet on a daily basis.

As a next step, the participants were presented with a neologism (without context). Just like in the previous study, they were asked whether they have come across it before, as well as whether they know the meaning and if so, what the meaning is. After this, the participants had to rate the words according to their efficiency, extravagance and extralinguistic relevance, using the respective subcategories, from zero (not at all) to six (absolutely). I tried to transfer the questions as literally as possible into German. Lastly, participants were asked to rank whether they would use the word.

⁸¹ A questionnaire sample can be found in appendix 3.3

A total of 156 participants took part in the study (in a ratio of 70 for one study and 86 for the second). Like in the previous studies, the participants distribution regarding demographics was quite uneven (cf. tbl. 21).^{82 83}

⁸² For a complete overview of the demographics from questionnaire study III, including participants who did not want to provide information, refer to appendix 3.4

⁸³ The same statistical tests that have been used for the demographic variables and *frequency* in study I, will also be utilised in the analysis of questionnaire III, namely ANOVA tests and t-tests. Further, the subcategories of appeal will be analysed by using the CFA (II) established in section 7.3.

		Gesamtschule		Hauptschule		Realschule		Gymnasium		University		
		female	male	female	male	female	male	female	male	female	male	
City	Germany	0	0	0	0	0	0	4	3	2	2	>25
		0	0	0	0	0	0	1	0	14	8	26-45
		0	1	0	0	0	0	1	0	2	2	<45
	Abroad	0	0	0	0	0	0	0	0	2	1	>25
		0	0	0	0	0	0	1	0	46	14	26-45
		0	0	0	0	1	0	0	0	6	5	<45
Countryside	Germany	0	0	0	0	1	0	1	0	0	1	>25
		0	0	1	1	0	1	2	2	5	5	26-45
		0	0	1	0	0	0	1	1	1	2	<45
	Abroad	0	0	0	0	0	0	0	0	0	0	>25
		0	0	0	0	0	0	0	0	4	0	26-45
		0	0	0	0	0	0	0	0	3	0	<45

Table 21 Distribution amongst participants in questionnaire study III

8.3 Quantitative Results

8.3.1 Demographic data: Age, Education, Location, and Internet usage

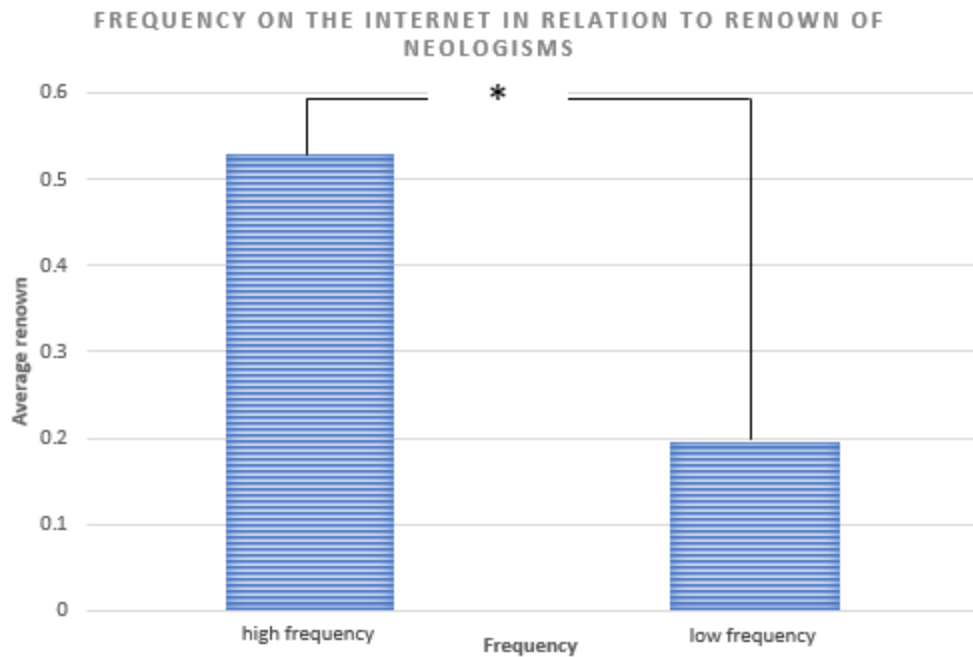
Despite having to face the same problem of an unequal demographic distribution as in the previous two studies, I nonetheless decided to investigate the collected data. Since this questionnaire, in contrast to the second study, contained more words, it seemed that potentially more reliable statements could be made. While the first study suggested a slight connection between *age* and *knowing a new lexeme*, no significant difference between all age groups in regard to knowing a new word could be attested in this data set [$F(2,155)=2.789$ $p=.065$]. However, a trend between the two extreme groups (under 25 versus over 45) was found [$t(42)=2.116$ $p=.036$]. Thus, while this study supports that there is a trend towards younger people knowing more neologisms, no reliable assumption can be made. Apart from this, no statistically relevant proposition can be made for the other demographic information. This includes that I could not attest any significant differences between groups regarding whether they know and use more Anglo neologisms.⁸⁴

8.3.2 Frequency

Similar to the first questionnaire study, also this study made use of frequency as a distinguishing criterion for the chosen neologisms. The previous outcome showed that frequency is an important dependent and independent variable (cf. section 6.2.2). In the scope of this study, these results were partially supported. The frequency attested online once more mirrored how widely the neologism was already diffused amongst speakers [$t(10)=2.428$, $p=0.036$] (cf. grf. 21). Thus, words that exhibit a high frequency on the internet were also more frequently known by the participants of the study. This confirms that the assessed online frequency is a reliable predictor that mirrors how much a neologism is diffused offline.

⁸⁴ Compare appendix 3.5 for the complete list of statistical calculations for all demographic factors.

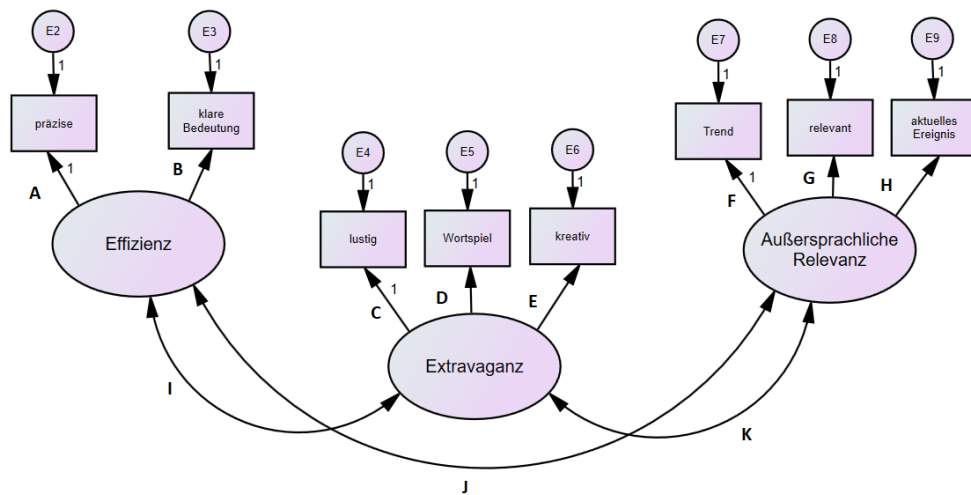
However, while study I showed that more frequent words also tend to be used more, the same was not found in this study [$t(10)=1.449$, $p=0.178$]. Thus, words that are more frequent on the internet are also more widely known offline but not necessarily used.



Graph 21 Frequency of the neologisms on the internet in relation to renown of neologisms amongst participants

8.3.3 Appeal

The main aim of this study was to find out whether the established model from the second study, which offered the possibility to describe appeal in more detail with three distinguishable factors, can also be transferred into German. Thus, the final model CFA (II) (excluding analogy and innovation) from the second study was applied to this data set (cf. grf. 22).



Graph 22 CFA II adapted to the German questionnaire study

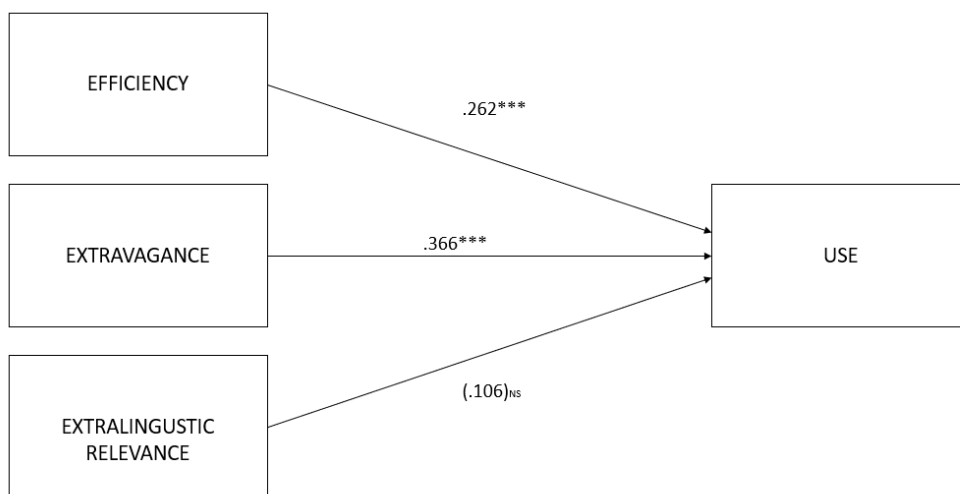
Table 22 shows that the model overall describes the collected data well.⁸⁵ With a chi-square of 19.9, it is even lower and thus a better fit than for the data of the second study (cf. tbl. 14). The insignificant *p*-value of 0.277, the low AIC of 57.94 as well as the RMSEA, which is with 0.033 below the benchmark of 0.06, all indicate good model fit. Similar to the previous study, differences between the respective neologisms can be attested. The neologisms *egosurfer*, for instance, matches the model extremely well with a chi-square of only 9.5, an insignificant *p*-value of 0.927 and a RMSEA of 0. In contrast to this, *foodtruck* is a less good fit, with a relative high chi-square of 40.2, a significant *p*-value of 0.001 and a RMSEA way above 0.06. Nevertheless, overall, the model describes the collected data well.

⁸⁵ More details about the model fit can be found in the appendix. While appendix 3.6 contains the values for the full model (CFA I) in contrast to the reduced model (CFA II), appendix 3.7 and 3.8 show the factor loadings for the respective confirmatory factor analyses.

	CHI-SQ	RMSEA	AIC	p-value
OVERALL	19.9	0.033	57.94	0.277
FOODTRUCK	40.2	0.141	78.228	0.001
WILLKOMMENSKLASSE	25.3	0.084	63.282	0.089
EGOSURFER	9.4	0	47.383	0.927
IMAGEBOOST	17.6	0.023	55.606	0.414
SCHWARMSTADT	20.1	0.052	58.141	0.267
FOODPORN	25.4	0.076	63.386	0.086
TRINKTOURISMUS	17.5	0.019	55.504	0.421
GHETTOFAUST	24	0.069	61.977	0.12
PORKDAY	37.4	0.119	75.415	0.003
DIGITALFALLE	12.9	0	50.917	0.742
CLICKBAIT	25.8	0.078	63.83	0.078
BEISTIFTSTEMMER	19.9	0.05	57.928	0.278

Table 22 Values indicating model fit for German data for CFA (II)

Study II furthermore revealed that all three factors used in the CFA with their respective indicators are not only distinct from each other, but that they all have a significant impact on use. When calculating the same regression with the data set from this study, it was found that efficiency [$B=.262, p=.001$] and extravagance [$B=.366, p<.001$] have a significant and independent impact on use, exactly as it was observed in the previous study. However, extralinguistic relevance [$B=.106, p=.117$] does not exhibit a significant impact on use (cf. grf. 23).



Graph 23 Impact of efficiency, extravagance and extralinguistic relevance on use for questionnaire study III

8.4 Discussion

The main outcome of the third study is that the model established in the previous study to describe appeal in more detail also matches this data set. Thus, the tripartite denotation into efficiency, extralinguistic relevance and extravagance as the major subcategories for appeal is not only described well by the English but also by the German data. Therefore, it seems that appeal was much more successfully subcategorised in study two and three in contrast to study one.

However, while in the previous study all subcategories of appeal had a significant impact on use, this data set only attributes this to efficiency and extravagance. Therefore, extralinguistic relevance does not seem to be an indicator for use in the German speaking data set. One explanation for this is a possible mismatch between indicators and words. Not every neologism used denotes a current event or a trend. It is possible that the words chosen for the English study were a better fit for the indicators of extralinguistic relevance than the ones for the German study. Comparing the factor loadings for all words in German and English (cf. appendix 2.4 & 3.8), it becomes apparent that the three indicators for extralinguistic relevance describe the factor better in English than in German. Thus, this can be an explanation for the outcome in relation to the factor's influence on use. Hence, while extralinguistic relevance is a useful subcategory of appeal, the used words might lack a connection to it.

Further, it could be a language or translation issue, so that German speakers might interpret the respective categories differently to English speakers. Another interpretation would be based on cultural differences. While English speakers might be more inclined to use a word in order to express something extralinguistically relevant, the focus of German speakers might be different. While for English speakers it might not be enough to communicate efficiently and to use words that stand out through their extravagance, a word maybe also needs to be nameworthy and has to describe an existing reality in order to be used. In contrast to

this, it could be that German speakers are more responsive to efficient and extravagant words, implying that for them the main aims are to stand out, but be understood at the same time. This fits the fact that efficiency is often stereotypically attested to Germans and, while this surely does not apply to all German speakers, it might be that efficiency, as an important construct of society, reaches all sorts of areas, including language.

Apart from this, hardly any statement can be made about the demographics in this study. Besides a trend at the extreme ends of the input variable *age* on the output variable *knowing a neologism*, nothing significant could be attested. A new hypothesis investigated in this study was the impact of the input variables such as *age*, *education*, *location* as well as *country of residency* on the output variables *use of Anglo neologisms*. It was assumed that for instance younger people might use and know more Anglo neologisms, the same was thought for more educated people, people in cities and German speakers who live outside of Germany. However, none of these assumptions were supported by the data. While the city-countryside gap lacks relevance due to the internet being available anywhere, the other outcomes need to be looked at more closely.

Since younger people grow up with the internet and use a greater variety of social media platforms (cf. section 6.2.1), it seemed logical that they would know and use more ‘hashtagable’ Anglo neologisms in contrast to older people. However, since the questions of whether participants would use a neologism is rather broad, it is possible that maybe younger participants are more inclined to use an Anglo neologism online, on social media or as a hashtag but not so much in spoken language. Thus, once more a clearer definition or distinction of use – such as would you use it as a hashtag, in spoken language, in a blog etc. – could have revealed a different outcome. Furthermore, while younger people certainly are influenced more by the English language nowadays, it could be a false assumption that most of them appreciate English words and thus would know or use them.

Besides age as a factor, highly educated people as well as participants who live outside of Germany were assumed to be more open towards Anglo neologisms. With regards to the notion that people who live abroad might be more inclined to use Anglo neologisms, it is possible that these people do not follow German media, social media etc. that much anymore and therefore potentially miss out on both, German and Anglo neologisms. Further, these people might speak the language of the country they live in for most of the day and therefore, the opportunities for a new German word to become conventionalised and entrenched as well as used by these speakers might be lower due to a lack of language use. Thus, it could be that even if expats stumble over a German or Anglo neologism, they would not make use of it and eventually will forget about it, since they hardly make use of German in general.

Moreover, education might improve the speakers' English skills, however, there might be other – potentially social – factors, such as prestige or stigmatism involved, that prevent them from using Anglo neologisms.

All in all, several issues could have imposed a problem on the assessment of demographics and their influence on neologism use. Just like in the first study, the output variable of *use* might be too broad and should better be subcategorised. Another reason could be - once more - the unequal distribution in the data set.

Apart from demographics, the study also assessed the frequency and its impact on use. It showed that frequency online reflected frequency offline. Thus, similar to study I, it seems that the online estimation of frequency accurately depicts how widely a word is already diffused and known offline. Therefore, frequency seems to be a reliable dependent variable. However, in contrast to the previous findings of study I, frequency did not show any impact on use in this data set and thus did not prove to be an independent variable that influences the dependent variable of *use*. The EC-model implies that more frequent words are already more diffused by definition, but this does not necessarily mean

that they are usualised and entrenched at the same speed. Especially mass media often pushes diffusion and makes words 'go viral' but does not have the same quick effect on usualisation and entrenchment (cf. section 3.1.1).

While in the first study a higher frequency, and thus a higher diffusion also meant that people tended to use the respective words more, it seems that in this study frequent words have started to diffuse but have not gone beyond this and usualisation and entrenchment have not taken place yet. Questionnaire studies are snapshots, which implies that the some of the already diffused words might get more usualised and entrenched in the future. However, for them to do so, they need to be appealing to the participants. Thus, it can be assumed that to a certain degree appeal overrules frequency when it comes to why people would use a new word. Whereas frequency of repetition leads to diffusion, for the diffusion and usualisation wheel to spin further and to move the conventionalisation and eventually the entrenchment feedback loop and usage wheel, appeal seems to be needed.

9. Explorative follow-up Investigation

9.1 Research questions

As various new insights were gained from the previous studies, I decided to try to put some of these into practice in form of a small-scale, explorative follow up investigation, consisting of two different parts.

Study I and III showed that frequency within the internet is mirrored by frequency amongst the participants of the respective studies and thus it can be said that a neologism's diffusion on the internet equals its diffusion offline. While study I showed a significant connection between frequency and use and made frequency a significant predictor for use, study III did not confirm this (cf. section 6.2.2 & 8.3.2). The fact that there is a potential disparity between use, frequency, and diffusion goes hand in hand with the findings that diffusion, usualisation and entrenchment do not inevitably happen at the same time and with the same speed. Especially mass media can fuel diffusion drastically but not so much usualisation and entrenchment (cf. section 3.1.1). This implies that diffusion and use are not necessarily synchronous, however, they are still mutually dependent on each other as the EC-model shows (cf. fig. 4).

Further, it became apparent that a model developed to characterise appeal describes the collected data in English and German well and that it is a good predictor for use. Thus, a more appealing word is more likely to be used. While the interaction between frequency and use did not prove to be reliable, it is assumed though, that usage as the central wheel of the EC-model, impacts diffusion/frequency and usualisation (cf. section 3.1.1). Bearing in mind that neologisms are always bound to a certain time frame (cf. section 2.1) and their establishment is highly dynamic and subject to constant change, it can lead to some stability with all wheels constantly in motion. Thus, I assume that:

H1: Neologisms that exhibited high diffusion, high appeal ratings and high use in study I and II, will still be more diffused now (2 years later) than neologisms that were ascribed little diffusion, appeal, and use.

In contrast to this rather long-term approach, the second part of this small follow-up investigation will look at the short-term future fate of some very recent neologisms, trying to give a snapshot within the neologisms' life. It is assumed that:

H2: Recent neologisms that are perceived as appealing are more likely to diffuse and be used in the near future.

Thus, the aim of this small project is to provide some brief insight into both, the long-term diffusion development of neologisms as well as the instant, short term one.

9.2 Method

9.2.1 Materials – retrieving neologisms

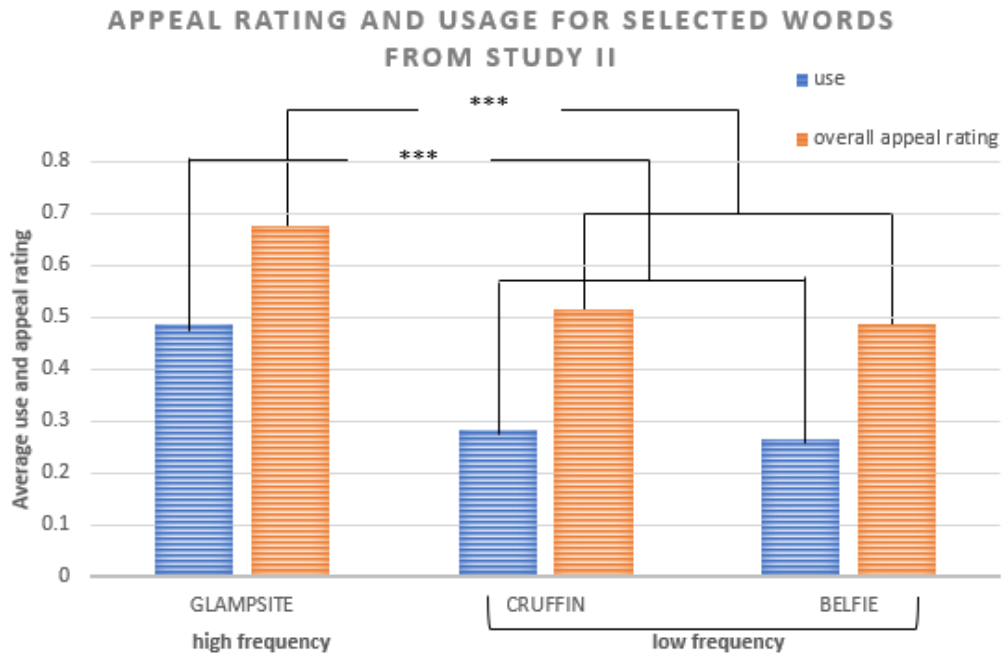
As a first step, I had to retrieve suitable neologisms for the investigation. To address the first hypothesis, I needed to choose some of the neologisms that I examined from the very beginning. Besides, they also needed to be included in the second study, since this allowed me to gain information about their appeal ratings. Suitable words were meant to differ regarding their perceived appeal, their online diffusion (frequency) as well as their use. When looking at the available data, a prime example for a highly appealing, highly frequent/diffused and highly used word was *glampsite*. At the time it was extracted from the NeoCrawler, it was already classified as highly frequent and thus diffused, both on the internet as well as offline (cf. tbl 6). Apart from this, it exhibited a high use in questionnaire study I (cf. appendix 1.15), as well as a high use and appeal rating (in all three different subcategories) in study II (cf. appendix 2.5). Therefore, it seems as if *glampsite* units all the factors investigated in this thesis that fuel a neologism's success. Thus, the data hints

towards a virtuous circle for *glampsite* and I decided to choose it as a word that could potentially be still highly diffused two years down the line.

On the opposite end of the spectrum, I found *cruffin* and *belfie*. They exhibited a lower frequency/diffusion when they were extracted from the *NeoCrawler* (cf. tbl 6).⁸⁶ Further, their usage ratings in study I were rather low (cf. appendix 1.15), a trend that was confirmed in study II (cf. appendix 2.5). When putting *cruffin* and *belfie* in context with *glampsite* in t-tests, it can be said that they were perceived as significantly less appealing (for *cruffin* $t(2286)=16.256$, $p<.001$, for *belfie* $t(2286)=17.520$, $p<.001$) as well as less likely to be used (for *cruffin* $t(284)=6.697$, $p<.001$, for *belfie* $t(284)=7.590$, $p<.001$) (grf. 24). Hence, the clear gap in their perceived appeal, their use and their initial online diffusion made them suitable candidates for this explorative investigation.

Another factor that made them eligible is the fact that their first occurrences go back a long time, implying that all these words have been around for a while, which is necessary for making observations about their long-term development. *Glampsite* and *cruffin* first occurred in 2009, *belfie* in 2013. Their first assessment in the framework of this study took place at the end of 2017, early 2018. The appeal rating in questionnaire study II was conducted at the end of 2018. Being 2020 now, nearly two years have passed since these words have been last assessed.

⁸⁶ While *belfie* depicted a borderline case in regard to its frequency when extracted from the *NeoCrawler* for the first study, the fact that its use and appeal were significantly lower rated in contrast to *glampsite* in study I and II makes it a suitable candidate.



Graph 24 Appeal rating and use for selected high- and low-frequency words from study II ⁸⁷

As a first step to investigate the second hypothesis - whether newly invented neologisms that are seen as efficient, extravagant and extralinguistically relevant are more likely to diffuse online in the near future - I needed to find such neologisms. To do so I looked through the aforementioned *About words*, a blog provided by the *Cambridge Dictionary* that lists recent neologisms (cf. section 5) and extracted suitable candidates (cf. appendix 4.1). As a next step, I searched all selected words on Google and Twitter, to find whether they might have been in use for a longer period already. The most recent words that I found are the following:⁸⁸

⁸⁷ The appeal score used here is the average ranking of efficiency, extravagance and extralinguistic relevance put together

⁸⁸ Their definitions can be found in appendix 4.2

	Cambridge Dictionary Blog	Google	Twitter
BRADIGAN	-	2019	2019
WALKUMENTARY	2019	2018	2017
RANDONOUT	2019	2019	2019
JANXIETY	2019	2019	2020
CORKITECTURE	2020	2019	2019
MEGXIT	-	2018	2017

Table 23 New neologisms chosen for study IV

As stated, while I was mainly looking for N+N blends for greater comparability, amongst the six words, two do not follow the N+N blending pattern. While *walkumentary* depicts a borderline case that can be argued to be a blend of *walk* + *documentary* instead of *walking* + *documentary*, *randonaut* follows an ADJ+N pattern, being formed by blending *random* and *astronaut*. Since finding very recent neologisms that have just been added to the English language has proven to be quite difficult, I decided to keep these two words in for now and see whether they behave differently in the analysis.

9.2.2 Assessment and questionnaire design

After having decided on suitable neologisms for the two different parts of the explorative investigation, I assessed them in a next step. For the 'older' neologisms, which have already been investigated in study I and II, two values were needed. First, I reused the previous overall frequency assessment from study I (cf. tbl 6). Secondly, I made use of the *NeoCrawler* in order to get current information about the words' diffusion. I extracted their frequency over four months in 2020 - starting from week 4 (21.01.202-27.01.2020) until week 20 (12.05.2020-18.05.2020). As in the first study, page frequency was used to assess frequency. Similar to the process in study I, I needed to make sure that I eliminated false hits, typos etc.

The recent neologisms had yet to be analysed. Therefore, I added them to the *NeoCrawler* and investigated their development over the same time span as the other ones, namely from week 4 in 2020

(21.01.202-27.01.2020) until week 20 (12.05.2020-18.05.2020). Once more, the page frequency was counted.

Since these recent neologisms have not been assessed yet in relation to their efficiency, extravagance, and extralinguistic relevance, I needed to get them rated by native speakers. For this purpose, I created one last, small-scale questionnaire. After a short introduction to the questionnaire and questions about the demographics of the participants – a mere standard process in this questionnaire - the participants were presented with one neologism at a time (without context). They were asked whether they know the neologism and afterwards were presented with a definition of the new lexeme. Then they had to rate it for the different indicators of appeal, from zero (not at all) to six (absolutely). In the last question they had to state how likely they are to use the word from one to six.⁸⁹ This way I hoped to get a comparable rating to study II, even though on a much smaller scale. In the end 22 participants took part in the study.⁹⁰

9.3 Results

For the first part of the investigation, I added up and averaged the page frequency during the chosen time span. *Table 24* shows the average observed frequency in 2020 as well as the frequency that was attested to the lexemes when they were first assessed for study I.

NEOLOGISMS	AVERAGE OBSERVED FREQUENCY	AVERAGE FREQUENCY AROUND FIRST OCCURRENCE
GLAMPSITE	3.9	9.4
BELFIE	2.6	4.2
CRUFFIN	6.4	2.2

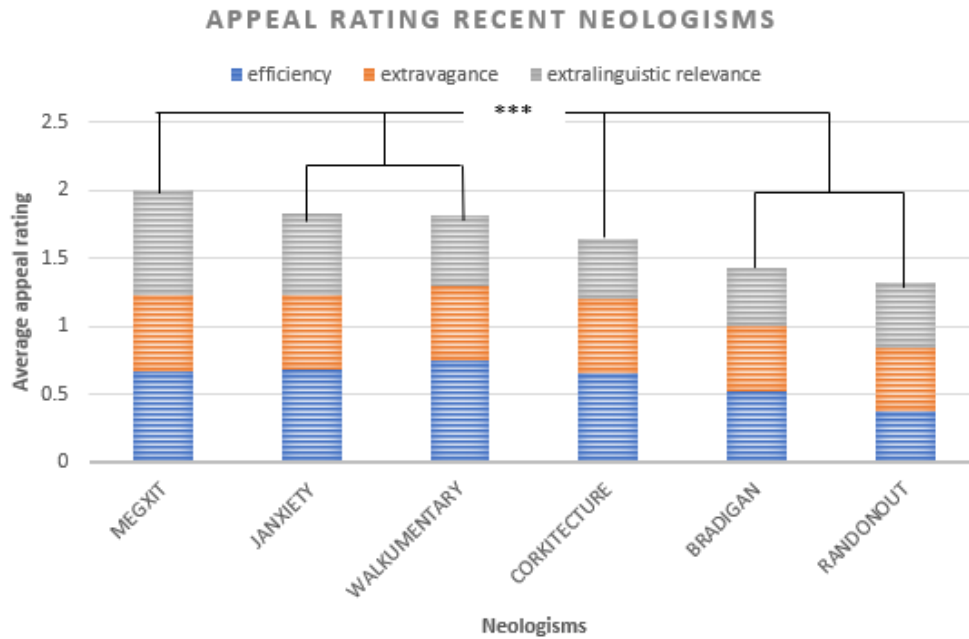
Table 24 Average frequency score for neologisms from study I & II during their first occurrence and in 2020

⁸⁹ A complete sample of the questionnaire can be found in appendix 4.3

⁹⁰ For a complete overview of the demographics, including participants who did not want to provide information, refer to appendix 4.4

The table shows that, while *glampsite* had an average frequency of 9.4 when assessed in 2017/18, it now shows an average of less than four pages per week. *Cruffin* follows a reverse trend, with only 2.2 occurrences per week in the past, and 6.4 in 2020. The frequency of *belfie* decreased between 2017/18 and 2020 from 4.2 to 2.6 pages per week.

For the investigation of the recent neologisms, two things were assessed. The neologisms' appeal on the one hand, and their frequency and thus diffusion in early 2020 on the other hand. For investigating the former, a small-scale questionnaire study was carried out. For a better comparability of the six words, I added up their scores for efficiency, extravagance and extralinguistic relevance (cf. appendix 4.5). Comparing these values with each other in an ANOVA showed that a significant difference between groups/words can be detected [$F(5,1050)=21.955$, $p<.001$]. When zooming into this outcome by using t-tests, I realised that no significant difference in appeal can be attested between *janxiety* and *walkumentary* [$t(350)=-.713$, $p=.476$] as well as *bradigan* and *randonout* [$t(350)=.961$, $p=.337$]. Thus, this leaves me with four distinguishable groups: high appeal (*megxit*), relatively high appeal (*janxiety*, *walkumentary*), relatively low appeal (*corkitecture*) and low appeal (*bradigan*, *randonout*) (cf. grf. 25). Due to the fact that *walkumentary* and *randonout* depict rather imperfect candidates since they cannot be classified (clearly) as N+N blends, and since their perceived appeal does statistically not differ significantly from the one of *janxiety* and *bradigan*, I decided to exclude them from the following analysis.



Graph 25 Appeal rating for recent neologisms

Once these four groups were established, I investigated the frequency of the underlying neologisms in the NeoCrawler. I added up all countable hits during week 4 (21.01.202-27.01.2020) until week 20 (12.05.2020-18.05.2020) and determined an overall average frequency for each word (cf. tbl. 25). The outcome shows that *mexxit* is highly diffused with an average value of 9.1 pages per week, in contrast to all other words. Within the other groups, *corkitecture* is the least frequently used neologisms with an average of 0.3 pages per week.

NEOLOGISMS	AVERAGE OBSERVED FREQUENCY
MEGXIT	9.1
JANXIETY	1.9
CORKITECUTRE	0.3
BRADIGAN	0.5

Table 25 Average frequency score in 2020 for recent neologisms

9.4 Discussion

The results of this small-scale explorative investigation show two different things. Regarding the ‘older’ neologisms, it was found that no reliable statement can be made about their diffusion two years after their first assessment. It was assumed that highly appealing, diffused and used words stand a high chance of survival and, by having all wheels in motion should stabilise in their diffusion. This cannot be confirmed.

In contrast to my hypothesis, the word with particular prospects for success, *glampsite*, records a decrease in frequency and thus diffusion. In contrast to this, *cruffin* seemed to have been doomed for obscurity when first analysed but is currently more frequent and diffused on the internet than *glampsite*, despite its poorly perceived overall appeal, use and diffusion back in 2017/18. Only *belfie* seems to follow the pattern, being a word that was attested little chance of survival due to a lack of diffusion, use and appeal. The word’s frequency decreased but nonetheless it is still in use. Thus, the hypothesis that neologisms that exhibit promising characteristics at a certain point in time will also do so some years down the line could not be confirmed. It seems that appeal, diffusion and use of a word at a certain time are not reliable predictors for a word’s success in the long term. It was mentioned that neologisms are dependent on time (cf. section 2.1) and conventionalisation and thus the diffusion and usualisation can also depend on topicality and nameworthiness (cf. section 3.1.1). These two aspects partially account for the appeal factor of extralinguistic relevance.

Taking this into account, various interpretations for the mentioned results can be brought forward. The decrease in use of *glampsite* might be explained seasonally, similar to what has been found in the case of *cherpumple* (cf. section 3.1.2). The frequency assessment for this follow-up investigation took place between January and May 2020 (in contrast to late summer/autumn in the first study), thus maybe not the prime season for *glamping*. It is possible that if the neologisms had been re-examined in a different season, the frequency of *glampsite* could have

been higher. Another explanation could be that glamping and with it *glampsites* might have been a trend some years ago, but its popularity is decreasing. In the case of *cruffin*, the opposite might apply. The increased diffusion could be the outcome of an enhanced presence of the invention in our world.

While extralinguistic changes are one possible explanation for a decreased frequency, appeal does also constitute of efficiency and extravagance. During the establishment process of new words, the new lexical items often lose ambiguity with increased frequency. The words develop a limited, specialised, and fixed meaning by means of hypostatisation (cf. section 3.1.3). Thus, once this occurred, the word proves to be more efficient for communication as ambiguity is reduced. This also implies that the neologism becomes more and more semantically transparent, which, according to study I has a significant impact on the use of a new word. Hence, frequency, diffusion, use and transparency form a virtuous circle that condition each other.

It is possible that *cruffin* entered such a circle and by potentially becoming a more efficient word over time, it also became more frequent. The trigger for this development is hard to determine. It could be extralinguistic, as stated, however, it could be of another, unknown nature, too. Another part of this circle is that increased efficiency often goes along with a decreased extravagance. The more habitual, predictable, and 'normal' a word becomes, the less extravagant it will be (cf. section 7.1). This would imply that during the establishment of a word, extravagance, and efficiency balance each other out with one of them increasing and the other one decreasing.

With the recent neologisms observed, it behaves slightly differently. Since they have been rated shortly after coming into existence and their progress was monitored briefly afterwards, they did exhibit a usage behaviour that was in close proximity to the expectations. Even though not each of the four established rating groups behaved differently, extremely highly appealing words showed a much higher frequency on

the internet in contrast to lowly rated words. This suggests that appeal might be a good predictor for the short-term diffusion of neologisms, however, this conclusion must be treated with caution due to the small number of investigated words.

All in all, the analysis of the 'older' neologisms showed that appeal, frequency/diffusion, and use might not necessarily be good predictors for the long-term success of a word. The recent neologisms, however, imply a connection between diffusion and appeal. It seems that the subcategories of appeal are good predictors for a stabilised diffusion in the near future, not so much in the long-term. Thus, the summarising observation is that appeal seems to be an unstable construct over time and a change in appeal could cause a change in frequency/diffusion and eventually also use. Therefore, while high efficiency, extravagance and extralinguistic relevance can be regarded as good predictors for use at a specific time and potentially the close future, their changeability makes a prediction of how successful a word will be in the future difficult. Appeal can give a word an additional boost in its early stages and positively impact its diffusion. However, if this boost does not result in other factors, such as an increased usualisation of the word, the neologism is unlikely to survive since the conventionalisation and entrenchment feedback-loops will not pick up pace and spin, which eventually means that also repeated usage will not happen.

The fact that only a very small sample of words was used and that it only served an observational, rather than quantitative purposes means that the findings cannot be regarded as finite results but rather as observations that have to be treated with caution. However, the outcome shows that neologisms cannot be seen as timeless, contextless words, but have to be embedded in a certain era or timeframe. The fact that all appeal factors are subject to constant change shows once more how dynamic language is and how all wheels in the EC-model are linked to each other. All of them need to be in constant swing for a new lexical item to get established by means of conventionalisation, entrenchment, and usage.

10. Summary and conclusion

10.1 Overall summary of the results

Over the course of three years, several studies have been carried out and various neologisms have been investigated. While the design and execution of the studies were not free of flaws, some promising results have been found. I want to start this summary by pointing out the main problems I encountered during the development, implementation, and analysis of the studies before moving to the most interesting results.

One of the clearest issues I had to face during the conduction and consecutive analysis of the studies was the unequal distribution of subjects throughout all questionnaires. This implies that making reliable statements about the demographics and their impact on the diffusion and use of neologisms is difficult. Therefore, while the findings exhibit some trends – such as younger people being familiar with more neologisms – they can only be understood as exactly that: trends. While this is not an ideal outcome, the fact that most of the analyses of the studies were made on the word level rather than the participant level, rendered this a bearable flaw. Nevertheless, it would be interesting to have a closer look at demographics and diffusion in future, with a more equal distribution amongst participants.

Another issue, which especially affected the first study, was wording problems. The analyses showed that some questions did not test what they were meant to test for, since they were perceived and interpreted differently by the participants. This became very visible in the case of *general appeal*. The high correlation value with *use* shows that *use* and *general appeal* are indistinguishable and were perceived as the same thing by the participants, although I intended to test two different variables. While this could be avoided by conducting large scale pilot studies (my conducted pilot study might have been too small and did not raise the issue), the outcomes can still be used for follow-up studies.

Apart from a more large-scale pilot study, using interviews in addition to the studies could have been a possible improvement and is something that should be considered for the future. A potential problem with wording also became apparent in the debrief for my second questionnaire study, where it was suggested that the variable *use* might be too broad. Thus, participants could interpret this as use in real-life, the internet, on social media etc. Therefore, a more precise and clear-cut description of use could have potentially revealed some more insight into the matter.

Another issue of some of the studies was inaccurate categorisation. Hence, some categories and variables tested did only partially fit the data or were unequally represented. When looking into the quality of the media, for instance, the fact that some categories were rather unevenly advocated potentially led to a distortion in the results. Similarly, not all indicators for the appeal factors represented the chosen neologisms. Hence, testing for something that is not inherent to the data depicted an issue. Despite trying to make sure that this does not happen, some of this only became apparent during the evaluation of the data. The change from English to German might have yielded similar problems, as it seems that some subcategories for appeal were more accurate for the English rather than the German neologisms. Hence, some more thoroughness would have been needed in these cases. Extensive pilot studies could once more have been a way to prevent these discrepancies. However, due to the fact that each of the studies was based on a previous one and some of them had to be as comparable as possible, some of the issues stated could not be changed throughout the course of the different studies.

Despite the weaknesses mentioned, the studies succeeded in increasing our knowledge about the diffusion and use of neologisms. Establishing a new word and keeping the different wheels of the EC-model afoot once an innovation was made is a highly complex and dynamic process. Various aspects have been investigated to try to get a clearer picture as to what brings the wheels (and particularly the conventionalisation and usage wheels) of the EC-model into swing. A

highly debated force that pushes all three wheels – conventionalisation, entrenchment, and usage – is frequency. While the first and third study showed that frequency online mirrors the neologisms' diffusion offline, a higher frequency did not prove to be a reliable predictor for an increased use (cf. section 8.3.2). Thus, more diffusion does not automatically imply more usage, which seems to be especially true for mass media that reaches a large number of people and thus spurs on diffusion. This outcome exemplifies the fact that the interactions between the two feedback loops and usage can be rather complex and uneven.

The first study further revealed that, while transparency, recognisability and appeal all have a significant impact on use, appeal is the strongest predictor out of the three. This outcome formed the basis of the second study. Since the first study was able to identify appeal as the best indicator for use but failed to get a finer and more detailed grip on it, the main aim of the second study was to get a better idea of what appeal denotes.

Considering the outside forces that set the usage wheel within the EC-model in motion, efficiency and extravagance were selected as two important subcategories of appeal. Although diffusion/frequency was not found to be a good predictor for use, the conventionalisation feedback loop and the usage wheel are still dependent on each other. Therefore, when looking at forces promoting conventionalisation, previous research found that newsworthiness and topicality can positively impact conventionalisation. Hence, the third category of extralinguistic relevance was assigned to appeal. While this subcategorisation has proven to be a good way to describe the investigated neologisms, some amendments regarding their respective subcategories had to be made. *Innovation*, for instance, was found to not describe extralinguistic relevance well for the chosen words, potentially due to the fact that most neologisms investigated were not innovations. Once this 'fine-tuning' was done, the tripartite division of appeal described the investigated neologisms well and exhibited a significant impact on use.

The third study confirmed the outcome of the second study in German. Thus, the way appeal was defined in study II also described the chosen German neologisms well. Furthermore, two out of the three subcategories had a significant impact on use. Therefore, the three conducted studies worked towards getting a better idea of what appeal is and succeeding in doing so. The indicators of the subcategories of appeal are partially forces that drive conventionalisation and usage, which links the outcome of the studies directly to the EC-model. Therefore, the results are a supplement to the EC-model that empirically tested some of the forces which are assumed to influence the movement of the wheels of the Tinguely machine.

In a very last step, the acquired new insights were investigated in a rather observatory approach, using both completely new neologisms as well as some of the previously investigated ones. While the sample of tested words is unrepresentatively small, the observation made is that appeal as a predictor for a neologism's success – here in case of its diffusion – only seems to work in the near future. Over various years, it does not seem as if appeal can predict the diffusion of new words. This is not surprising insofar as neologisms are highly time sensitive and always have to be embedded in a certain context and a specific time frame. Therefore, while these studies served to define appeal and its influence on diffusion, usualisation and use of neologisms in more detail, the fact that language itself and the establishment of innovations is so dynamic and dependent on so many different, constantly changing factors, implies that the studies conducted only provide a snapshot of the state of the investigated neologisms in a very specific moment during their circle of life.

10.2 Outlook

While the presented studies have revealed some interesting aspects and although some of the outcomes fit nicely into the theoretical discussion, there are still many open questions concerning neologisms in general but also regarding their appeal and 'attractiveness'. Therefore, I will briefly

outline some aspects of the findings that might be worth investigating further in the future to supplement what has been found so far.

Although the first study identified appeal as one of the major forces for the use of neologisms, other important factors, such as transparency and recognisability, also exhibited a significant influence on use. Some aspects of transparency were already incorporated in the subcategory of efficiency. Nevertheless, it would be interesting to go deeper into this matter in further studies. While the connection between transparency, recognisability and blends has already been partially investigated (cf. Lehrer, 1996), the connection between these factors and neologism blends has not been subject to research yet. Like in the case of appeal, study I showed that the assumed distinction between formal and semantic transparency cannot be supported statistically. Hence, a more zoomed-in approach with better subcategories of formal and semantic transparency and subsequently their influence on the use of neologisms could be investigated.

The term of *use*, within this thesis, has been used very broadly. Only throughout the studies and by receiving feedback it became apparent that a more detailed subcategorisation of the variable *use* would have been of interest. Thus, another possible project that could evolve out of the results of these studies could be to investigate whether there is a difference in the use of neologisms when used in *real-life*, on the *internet* or as a *hashtag*. Therefore, investigating different types of use could prove to be interesting.

While the research conducted in this thesis mainly focused on the social aspect of conventionalisation, it would also be interesting to look more closely at the feedback loop of entrenchment and its connection with transparency, appeal, and recognisability. While some research into the behaviour of neologisms and processing has already been conducted (cf. de Vaan, Schreuder & Baayen, 2007), and whereas semantic plausibility, frequency and neighbouring effects have not proven to influence the processing of novel blends (cf. Lehrer, 2003), it would be

interesting to investigate this further and deeper. Another facet could be whether blend neologisms are processed differently to other neologisms, which would mirror the findings that generally blends seem to be processed differently from non-blend words (cf. Juhasz et al., 2017). Hence, a more detailed examination of the entrenchment feedback-loop regarding various aspects of how neologisms and blend neologisms are processed would complement the findings from this thesis that focused on the conventionalisation feedback-loop.

Moreover, various shortcomings have been identified during the analysis of the conducted questionnaire studies. Therefore, it might be worth going back to these issues and tackling them in follow-up studies. Implementing further studies with a focus on equal demographic distribution could reveal more reliable outcomes on whether the speaker's background has any impact on the use and diffusion of neologisms. Another aspect that has not been accounted for satisfactorily is the impact of prestige – here in the form of media types and early user groups – on use. While the theoretical framework assumes that power and prestige are forces that can impact usage (cf. Schmid, 2020) and although previous studies have found a positive impact of the coiner status on conventionalisation (cf. Kerremans, 2015), the insufficient way of measuring this factor within this thesis did not allow to draw any conclusions. Hence, together with demography, the influence of power and prestige is a short coming that would be worth investigating further.

Penultimately, the word formation processes, which have not really been properly considered in this thesis, could be looked at in the future. While the main criteria in this thesis was to use N+N blends, they have not been sorted or examined according to their internal structure. Since blends can be formed following different patterns, such as clipping and overlapping (cf. section 4.1.2), it is possible that the way they are formed and the amount of material that is deleted might have an impact on how transparent but also as to how appealing they are to be perceived. In the framework of this study, I could not make any prediction whether appeal and transparency might be governed by the internal

structure of blends and thus whether the likeness of a neologism to be used is ultimately also subject to the words' structure. The same might be true for compounds. While this is a very different approach to the matter, it could reveal interesting additional information.

A last suggestion for the research to come is a long-term study of the lives of neologisms. Once we have an even better picture on how not only appeal but also other factors such as transparency and word-formation processes influence different types of use, some recent neologisms could be investigated over several years. The words could be reassessed ever so often in relation to the factors that are assumed to influence the different types of use (*online, offline, as a hashtag*), so that it would become clear whether there is a constant dependency between the examined factors and the different use types or whether it is impossible to link any factors to the use of neologisms over a longer time span.

Therefore, there is still plenty of follow-up research that could and should be done. While both neologisms and blends used to be, and still partially are, underdogs in the world of linguistics and have been marginalised for a long time, they experience more attention lately. Therefore, it can only be hoped that this trend will continue and that some of the findings of the studies presented will serve as the basis for other studies to come. The internet not only offers an increasing number of neologisms to investigate, but it also provides various means with which to examine them. While neologisms were often solely studied by using corpus data, this will hopefully be complemented by more field work such as questionnaire studies or interview-based experiments in the future to get a clearer picture of how neologisms behave, both in the 'real-world' as well as on the internet.

Creativity is one of the most fascinating things that language has to offer. There are hardly any boundaries as to what we can create and come up with. Even the worst realities, disasters, and crisis situations, such as the current pandemic and the associated restrictions, that keep

the whole world in suspense, can be described through the creativity of the human language. From *Coronavirus* and *covidiot* all the way to *social distancing* and *quarantine*, language, and in this particular case lexis, proves again and again how creative, flexible, dynamic, and inexhaustible it is, and this is surely worth researching.

Appendices

Appendix 1 Questionnaire study I – Materials and data

Appendix 1.1 Preliminary neologism list for study I

BAECATION	HONEYTEER
BEARDRUFF	INFOTAINMENT
BEDIQUETTE	INSTASHAM
BEGPACKING	LINNER
BELFIE	MANTRUM
BLEISURE	MARANOIA
BONESPIRATION	MEATMARE
BREADATARIAN	MUMSPIRATION
BROFLAKE	MOMSTER
BROGA	NEWSJACKING
BROGRAMMER	OBLICATION
BRONGERIE	PRESSTITUTE
BURKINI	RAPEFUGEE
CANCERVERSARY	REDDIQUETTE
CATIO	RUNCH
CRIMMIGRATION	RUNCHIES
CRONUT	RUNGER
CRUFFIN	SARGASM
CUPCAKERY	SHARENTING
CYBERREHEA	SITATUIONSHIP
DOGA	SMONDAY
DOGTOR	SNACCIDENT
DRONFIE	STAGFLATION
DRONOGRAPHY	SUBLING
FITSPIRATION	THRUPLE
FLEXTING	TINDERELLA
FOODVENTURE	TRUMPANZEE
GLAMPSITE	VEGANUARY
GYMSPIRATION	WINTERSCAPE
HAMDOG	YOUTHQUAKE
HOTNUM	

Appendix 1.2 Definitions of the chosen neologisms for study I

BAECATION	Hanging out with the bae for any length of time really. although usually used for extended periods of time, like a vacation. (Urban Dictionary)
BEGPACKING	Backpackers busking or begging to fund their travel. (Urban Dictionary)
BELFIE	A 'bottom selfie' - a photographic self-portrait featuring the buttocks, usually posted by female celebrities on social media networks (Urban Dictionary)
BLEISURE	Business and leisure combined, commonly used in the hospitality industry (Urban Dictionary)
BREADATARIAN	1) Someone who only eats bread (Urban Dictionary) 2) someone who consumes a large amount of bread: a lover of all types of breads I consume all types of bread, so I consider myself a breadatarian (Webster, 2016)
BROGA	1) yoga for bros (Urban Dictionary) 2) It's when you show up to yoga only to find the class is full of guys (Urban Dictionary) – not so common
BRONGERIE	Sexy lingerie made for men (The Sun)
CATIO	1) a cat patio (Urban Dictionary) 2) an outdoor enclosure for cats (Wiktionary)
CRONUT	A deep fried and glazed croissant aka a mix between a donut and a croissant (Urban Dictionary)
CRUFFIN	The food of godliness that combines the 3 greatest foods of all time - the Pie, Muffin, and Cookie. It consists of a Pie shaped mass of cookie dough, baked until its hard and cookie like, followed by a delicious centre of chocolate chip or double chocolate chip muffin stuffing.(Urban Dictionary)
DOGA	1) 'Doga is the practice of yoga with pet dogs.' (Wikipedia, 14.09.2017) 2) the practice of yoga by dogs (Urban Dictionary)
FITSPIRATION	Fit + inspiration. A healthier alternative to thinspiration, fitspiration is using examples of good fitness (people, photographs, skinny jeans, etc) as inspiration to attain a fitness goal (Urban Dictionary)
GLAMPSITE	A place to do glamping (NeoCrawler)
HAMDOG	1) A hotdog wrapped in a hamburger patty that is then deep fried and served on a hoagie roll with chili, cheese, onions, bacon, and a fried egg (Urban Dictionary)

	2) a combination of a hamburger and a hot dog (NeoCrawler)
HONEYTEER	A honeymoon spent doing voluntary work, usually abroad (Cambridge Dictionary Blog)
MANTRUM	When a grown man throws a tantrum when he can't have his way (Urban Dictionary)
MARANIOA	Fear of something going wrong (illness, injury, etc.) in the weeks before a marathon. (WordPress)
OBLICATION	1) A vacation taken out of obligation and not for fun, enjoyment or relaxation (Webster online dictionary) 2) A required or expected visit to family/in-laws during your vacation time. Not exactly what would qualify as a holiday (Urban Dictionary)
PRESSTITUDE	Is used on social media to refer to a media organization or an individual who claims to be unbiased but reports the news in a way so as to serve someone's hidden purpose (Techwelkin)
RUNGER	The hunger felt while running (Urban Dictionary)
SHARENTING	When parents share too much of their children's information, pictures and private moments online, mostly on Facebook (Urban Dictionary)
SNACCIDENT	When food (a snack) is consumed in an accidental, often regrettable way. This can refer to accidentally eating food of questionable quality and/or quantity (Urban Dictionary)
TRUMPANZEE	A fully respecting, unquestioning, loyal republican to Donald Trump. These people get triggered at even the slightest negative thing towards him (Urban Dictionary)
VEGANUARY	Veganuary is a New Year's resolution fad where someone decides only eat vegan meals for the month of January (Urban Dictionary)

Appendix 1.3 Neologisms according to their division into private and professional and their first appearance

NEOLOGISM	private	
	First occurrence 2001-2009	First occurrence 2010-2017
SNACCIDENT	Urban Dictionary 2007 Twitter 2009	
CATIO	2005 Blogs 18/07/2008 Twitter (private person)	
FITSPIRATION	24.01.2006 Blog (http://blog.muschamp.ca/2006/01/24/calf-crazy/) 2007 Video from YouTube (https://www.youtube.com/watch?v=SNsVCp9FIFQ)	
BREADATARIAN	2004 Blogs 2007 Urban Dictionary	
RUNGER	29/10/2009 Twitter (private person) Forum discussions, blogs	
BROGA	2005 Gym (https://www.kelseykerridge.co.uk/fitness-classes/yoga-classes) 2007 Urban Dictionary 2009 Twitter	
PRESSTITUDE		15/08/2011 Twitter (private person) forum discussions (http://www.zerohedge.com/contributed/2012-11-01/when-you-wish-upon-death-star)
BLEISURE		2010 Twitter (private person) 2013 Blog (http://blog.shawcontractgroup.com/tag/bleisure/)
TRUMPANZEE		2015 Twitter (private person)

		2016 Urban Dictionary
CRUFFIN		2009 Urban Dictionary 2011 Internet (http://www.americancupcakeabroad.com/yummy-places/bubble-tea-fever-hits-london) 2011 Twitter (private person)
HONEYTEER		2010 Twitter (private person) 2010 Adverts/blogs (https://www.experiencethevillage.com/honeyteer-with-the-village-experience/ https://thevillageexperience.wordpress.com/page/8/)
BEGPACKING		03.11.2016 Twitter private (https://twitter.com/search?l=en&q=begpacking%20bagpacking%20since%3A2015-07-31%20until%3A2017-12-31&src=typd&lang=en) 2017 Urban Dictionary 2017 Travel blogs (http://www.globetrotterguru.com/begpacking/ https://mothership.sg/2017/11/begpacking-singapore-money/ http://travellingclaus.com/begpacking/)

NEOLOGISM	professional	
	First occurrence 2001-2009	First occurrence 2010-2017
VEGANUARY	<p>2014 newspaper (http://metro.co.uk/2014/12/26/will-you-go-vegan-for-january-veganuary-is-the-latest-bossy-month-with-a-stupid-name-but-it-has-a-point-4996761/) https://www.theguardian.com/sustainable-business/veganuary-campaign-sustainable-eating-vegan-diet http://www.independent.co.uk/news/uk/this-britain/take-a-punth-on-it-how-each-month-of-the-year-became-a-charity-endeavour-9033226.html)</p>	
GLAMPSITE	<p>04/05/2009 Internet (http://www.greentraveller.co.uk/blog/luxury-campsites-and-glamping) https://travellingcontent.wordpress.com/tag/glamping/ http://www.tntmagazine.com/travel/top-guides/glamping-sites-uk-eight-of-the-best http://justimagine.jp/WordPress/glamping https://festivalbrides.co.uk/planning-a-hen-do-honeymoon-or-wedding-how-about-a-spot-of-glamping-part-2/)</p>	
HAMDOG	14/02/2005 CBS News	
MANTRUM	<p>2006 Urban Dictionary 2009 Internet homepages/magazines (http://minivanmonologues.blogspot.co.uk/2009/11/dont-have-man-trum-over-car-maintenance.html) http://www.everseradio.com/return-of-the-top-five-words-that-have-been-altered-so-they-refer-to-men-but-which-didnt-really-need-to-be-altered-at-all/ http://hollywoodlife.com/2009/12/14/)</p>	

	jay-seans-advice-to-chris-brown-be-more-humble/)	
OBLICATION	2002 Slang Dictionary 2009 CNN	
DOGA	2003 news (http://www.dogsonly.org/dog_news.html) 2004 newspaper (http://www.telegraph.co.uk/news/uknews/1466119/Beware-of-the-Doga.html) http://news.bbc.co.uk/1/hi/uk/3870273.stm http://query.nytimes.com/gst/fullpage.html?res=9C0CEFDA1539F932A1575AC0A9629C8B63)	
BAECATION		2014 WordPress (https://dopedecisions.wordpress.com/2014/09/23/i-need-a-baecation/) 2015 tabloids and press (http://www.dailymail.co.uk/tvshowbiz/article-3132945/Amber-Rose-enjoys-baecation-Machine-Gun-Kelly-claiming-busy-romance.html) http://www.caramiawhy.com/baecation/ https://travelbyky.com/2015/06/06/are-you-ready-for-a-baecation/)
SHARENTING		2011 Newspaper (http://www.healthywomen.org/content/blog-entry/moms-are-you-guilty-sharenting) http://parentinfo.org/article/should-you-share-pictures-of-your-children-online)
CRONUT		2013 Newspaper (http://www.businessinsider.com/what-is-the-cronut-2013-5?IR=T) http://www.independent.co.uk/news/uk/greggs-launches-their-own-version-of-a-cronut-the-greggsnut-8800041.html http://www.dailymail.co.uk/femail/article-2332287/Would-pay-40-cronut-Manhattans-new-pastry-craze-hits-Craigslist-croissant-donut-hybrid-resold-EIGHT-TIMES-retail-value.html)

MARANOIA		04/10/2012 Twitter (Runner's World: page for training tips etc (US page)) 04/10/2012 WordPress
BRONGERIE		2017 tabloid newspaper (https://www.thesun.co.uk/fabulous/4814358/man-wore-sexy-lingerie/ http://www.ladieswantmore.com/tag/brongerie/ https://www.oddnaari.in/life/story/lacey-lingerie-for-men-is-now-a-thing-brongerie-126633-2017-11-05 https://firenewsfeed.com/incident/700003)2010 Adverts/blogs)
BELFIE		2013 tabloid newspaper (https://www.irishmirror.ie/rise-side-bum-latest-celebrity-trend-2795257 http://metro.co.uk/2013/12/13/from-diane-kruger-to-lily-collins-our-best-dressed-celebs-of-the-week-4229518/ https://www.beaut.ie/life/bottom-selfies-belfies-latest-celebrity-trend-make-feel-like-aul-wan-87386)

Appendix 1.4 Questionnaire study I - sample

Neologism

0%

[Exit Survey](#) ↗

Dear Participant,

Many thanks for participating in this survey on new words. The study investigates the connection between new words, their usage amongst native speakers and the role of the media.

This is not about testing your linguistic proficiency. You cannot get it wrong! All we want to find out is a bit about your usage of media, your personal interests and whether you have come across some words that have entered the English language a short while ago.

The questionnaire lasts about 20 minutes. As a thanks for participating there will be a raffle for several prizes! At the end of the questionnaire you will get a link that will enable you to participate in the raffle! The link will also give you the option to receive more information about the content of the study and its outcome.

If you have any questions before, while or after completing the questionnaire please feel free to contact Sabrina.link@abdn.ac.uk

Next

Neologism

0%

Questions marked with a * are required

[Exit Survey](#) ↗

* Before you start, please read the following:

- I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason
- I understand that the information provided by me will be held anonymously so that it is impossible to trace this information back to me individually. In accordance with the Data Protection Act this information may be retained indefinitely.

Next

Participant's information

*** Sex**

- Female
 - Male
 - Non-binary
 - I'd prefer not to say
-

*** Age**

- <25
- 26-45
- 46-65
- >66
- I'd prefer not to say

* What is your highest level of education?

- School education
 - University degree
 - None of the above
 - I'd prefer not to say
-

* Where do you live?

- City/suburbs
 - Countryside/village
 - I'd prefer not to say
-

* What is your native language?

Media usage

* Which type of media do you use on an average day?

- Television
- Radio
- Print newspaper
- Print magazine
- Newspaper website
- Magazine website
- Blogs

Social media platforms

- Twitter
- Facebook
- SnapChat
- YouTube
- Instagram
- LinkedIn
- Other

* Wieviel Zeit verbringen Sie durchschnittlich am Tag im Internet (inklusive auf sozialen Medien)?

- Weniger als eine Stunde
- 1-3 Stunden
- 3-5 Stunden
- Mehr als 5 Stunden
- Anderes

How do you rate the quality of information of the following media types?

	very good	good	average	slightly poor	poor	no opinion
* Broadsheet	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Tabloid journalism/Yellow press	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Magazine	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Scientific journal	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Online Blog	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Social media platforms	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Online dictionaries and encyclopaedia	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
* Online discussion forums	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

How do you rate the influence of the following people?

	very influential	influential	slightly influential	not influential	no opinion
Professional broadsheet journalists	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional tabloid/yellow press journalists	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Professional magazine journalists	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Scientists	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Private users of social media	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Private bloggers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Private users of forums	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Private authors of encyclopedia and online dictionaries entries	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

*** What are your main interests?**

- Sports
- Law
- Politics
- Computing/Internet (gaming/technology)
- Entertainment (radio, TV, movies, music)
- Food and drinks
- Animals
- Celebrities
- Fashion
- Health and Beauty
- Business
- Advertising
- Science
- Travelling
- Other

Please answer the following questions referring to the respective words in the green box.

SNACCIDENT

* Have you come across this word before?

- Yes
- No

* Do you know the meaning of the word?

- Yes
- No
- I can guess it

* What is the meaning of the word?

OR

* Please try to guess the meaning of the word.

* Which two words do you think does the word consist of?

Word 1

* Word 2

* The word describes '*food that is consumed in a accidental, often regrettable way, both regarding quality and/or quantity*'. Do you think SNACCIDENT is a good way to say this?

- Yes
 - No
-

* Do you like the idea/concept of the word?

- Yes
- No
- I don't know

* Do you think you could find yourself in a situation where you would use this word?

- Yes
- No
- I don't know

*** According to you, in which of the following media types can this word most likely be found?**

- | | |
|---|---|
| <input type="checkbox"/> Broadsheet newspaper | <input type="checkbox"/> Online Blogs |
| <input type="checkbox"/> Yellow press/tabloid newspaper | <input type="checkbox"/> Social media platforms |
| <input type="checkbox"/> Magazines | <input type="checkbox"/> Online forums |
| <input type="checkbox"/> Online dictionaries and encyclopedia | <input type="checkbox"/> Scientific journals |
| <input type="checkbox"/> Other | |

*** SNACCIDENT was initially mainly used by private people on social media. Would you use this word?**

- Yes
 No

*** Why would you use it?**

OR

*** Why would you not use it?**

* Do you think SNACCIDENT deserves to become a thing?

- Yes
- No

Appendix 1.5 Neologisms and distractors split in questionnaire I & II within study I

QUESTIONNAIRE I		QUESTIONNAIRE II	
Neologisms	Distractors	Neologisms	Distractors
SNACCIDENT	MOTEL	FITSPIRATION	EMOTICON
CATIO	WORKAHOLIC	VEGANUARY	CHOCOHOLIC
BREADATARIAN	SMOG	GLAMPSITE	SPANGLISH
MANTRUM	SHOPAHOLIC	PRESSTITUTE	ROMCOM
OBLIGATION	SITCOM	TRUMPANZEE	CAMCORDER
BELFIE	BRUNCH	BAECATION	BOLLYWOOD
CRONUT		DOGA	
HAMDOG		CRUFFIN	
BRONGERIE		HONEYTEER	
BLEISURE		MARANOIA	
SHARENTING		RUNGER	
BEGPACKING		BROGA	

Appendix 1.6 Complete overview of the demographics of all participants in study I

	School				University				N/A or other				
	Female	Male	Non-binary	N/A	Female	Male	Non-binary	N/A	Female	Male	Non-binary	N/A	
City	23	10	3		17	7						1	<25
	2	0			18	13							26-45
	1	0			4	3							>46
Countryside	12	2			9	3				1			<25
	0	0			3	0							26-45
	2	0			4	5			1				>46
Prefer not to say	1												<25
													26-45
													>46
												1	N/A

	Criteria	Participants
Gender	female	99
	male	43
	Non-binary	1
	Prefer not to say	1
Education	School	54
	University	88
	Prefer not to say	3
	None of the above	1
Location	City	104
	Countryside	40
	Prefer not to say	2
Age	<25	90
	26-45	35
	>46	20
	Prefer not to say	1

Appendix 1.7 Overview of all outcomes regarding the demographics of study I

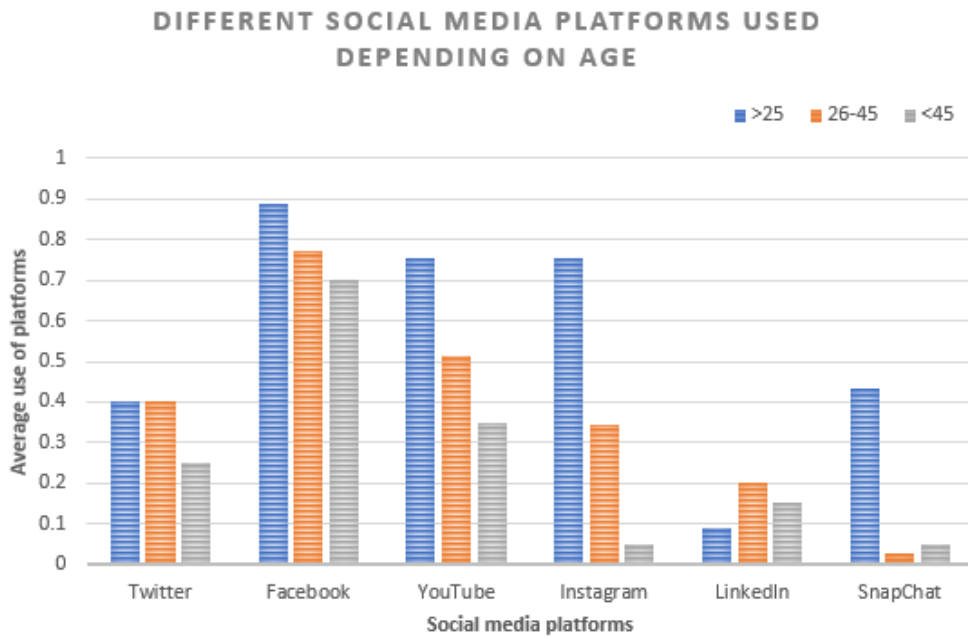
IN RELATION TO USAGE OF NEOLOGISMS:

AGE	$F(2,142)=1.941, p=.147$
LOCATION/PLACE OF LIVING	$t(142)=-.673, p=.502$
EDUCATION	$F(2,140)=.152, p=.859$
TIME SPENT ON THE INTERNET	$F(3,142)=.559, p=.643$
GENDER	$t(140)=.663, p=.509$

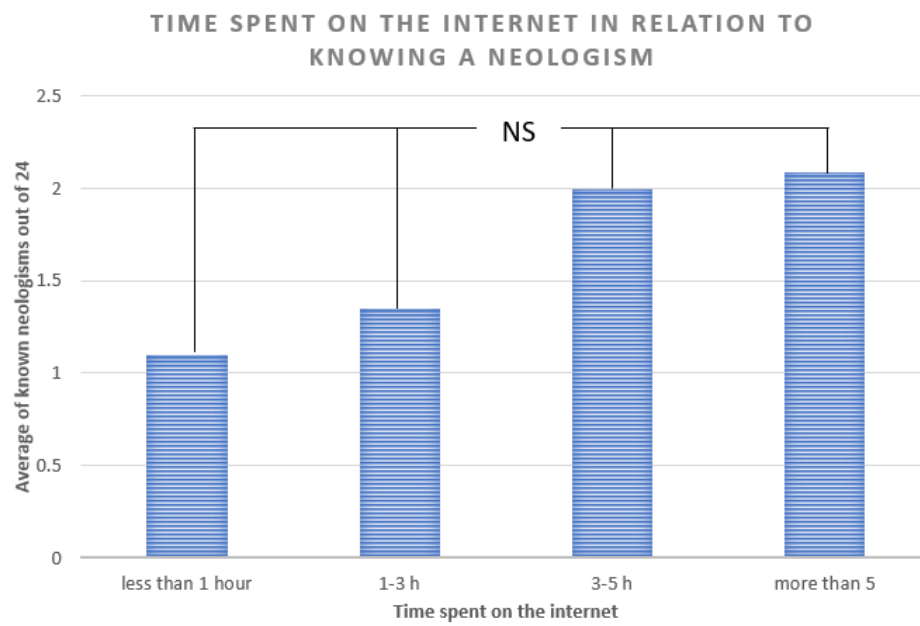
IN RELATION TO KNOWING A NEOLOGISMS:

AGE	$F(2,142)=5.122, p=.007$
LOCATION/PLACE OF LIVING	$t(142)=-.763, p=.447$
EDUCATION	$F(2,140)=.643, p=.527$
TIME SPENT ON THE INTERNET	$F(1, 145)=3.882, p=.051$
GENDER	$t(140)=3.037, p=.003$

Appendix 1.8 Different social media platforms used within different age groups



Appendix 1.9 Time spent on the internet in relation to knowing a neologism – graph



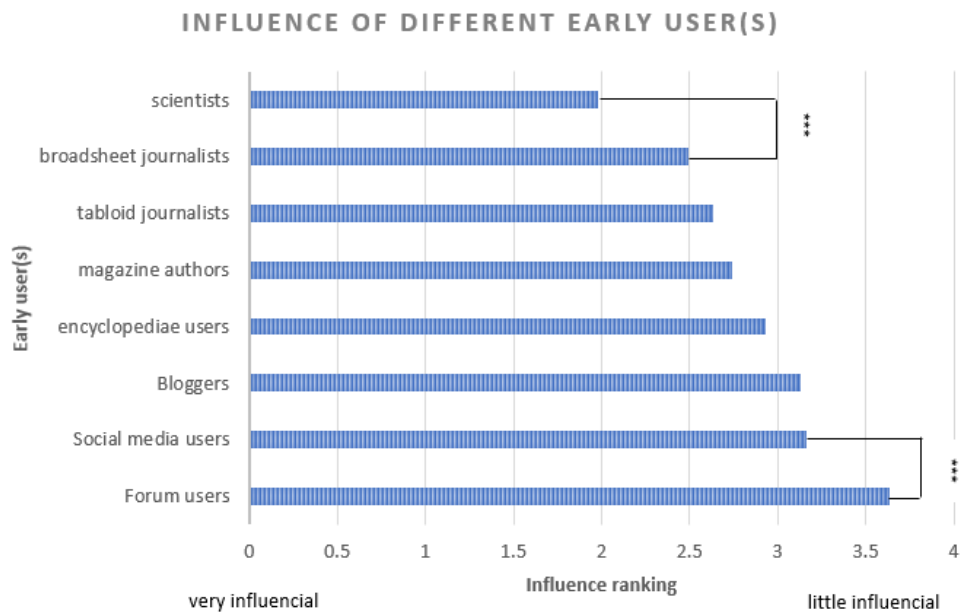
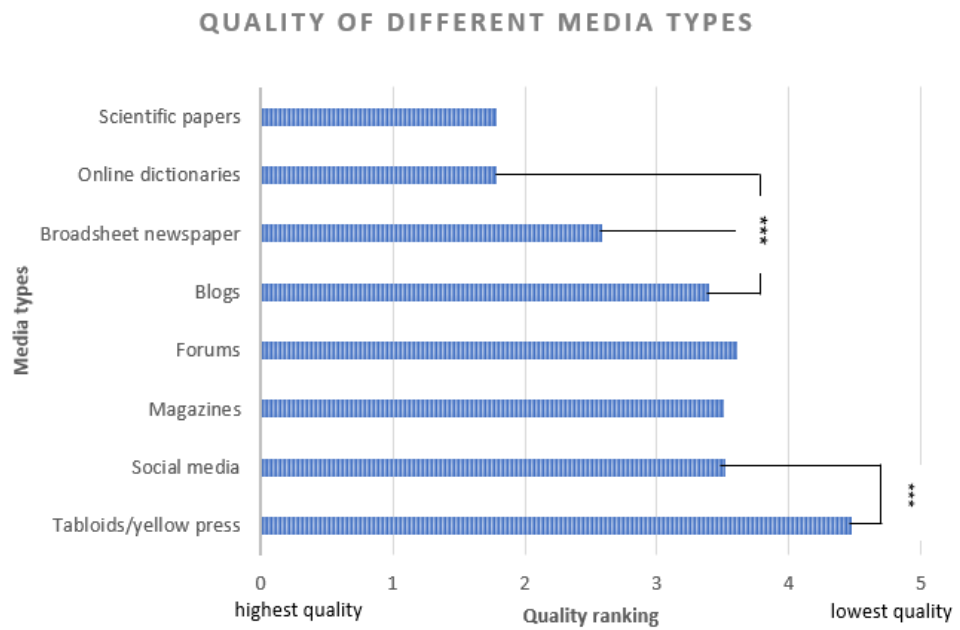
Appendix 1.10 Media types and early user group of neologisms

NEOLOGISM	MEDIA TYPE OF FIRST APPEARANCE(S)	EARLY USER GROUP
SHARENTING	Medical journals and magazines	Scientists
MANTRUM	Tabloid/yellow press, celebrity magazines	Tabloid journalists
BRONGERIE	Tabloid/yellow press	Tabloid journalists
BELFIE	Tabloid/yellow press	Tabloid journalists
DOGA	Broadsheet newspaper	Broadsheet journalists
VEGANUARY	Broadsheet newspaper	Broadsheet journalists
OBLICATION	Broadsheet newspaper	Broadsheet journalists
CRONUT	Broadsheet newspaper	Broadsheet journalists
HAMDOG	Food homepages, broadsheet newspaper	Broadsheet journalists
BAECATION	Travel magazines/pages	Magazine journalists
GLAMPSITE	Travel magazines/homepages	Magazine journalists
MARANOIA	Sports magazines/pages	Magazine journalists
BREADATARIAN	Urban Dictionary, Forums	Encyclopaedia authors
CRUFFIN	Urban Dictionary, Twitter	Encyclopaedia authors
BROGA	Twitter	Social media users
TRUMPANZEE	Twitter	Social media users
HONEYTEER	Twitter	Social media users
SNACCIDENT	Twitter	Social media users
FITSPIRATION	Blogs, Twitter	Blogger
BLEISURE	Blogs, Twitter	Blogger
CATIO	Blogs, Twitter	Blogger
BEGPACKING	Blogs	Blogger
PRESSTITUTE	Forum, blog, social media	Forum users
RUNGER	Forum, blog, social media	Forum users

BLUE = neologisms firstly used in professional contexts

RED = neologisms firstly used in private contexts

Appendix 1.11 Media types and early user groups ranked by participants according to their quality and influence respectively



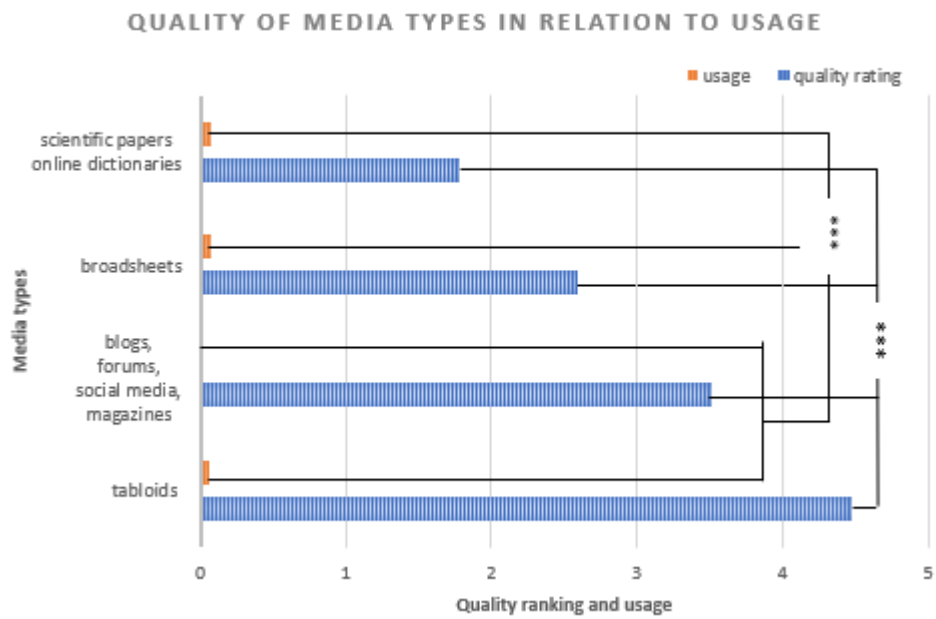
Appendix 1.12 Statistical details for categorisation of quality of media and influence of early user groups

Different media types compared	Outcome t-test
Scientific papers – online dictionaries	$t(290)=.058, p=.954$
online dictionaries - broadsheet	$t(290)=5.434, p<.001$
Broadsheet - blog	$t(290)=-4.760, p<.001$
Blog - forum	$t(290)=1.473, p=.290$
Forum - magazines	$t(290)=.762, p=.447$
Magazines – social media	$t(290)=.055, p=.956$
Blog – social media	$t(290)=.861, p=.390$
Blog/forum/magazine/social media – yellow press	$t(728)=-9.243, p<.001$
Social media- yellow press	$t(290)=-7.670, p<.001$

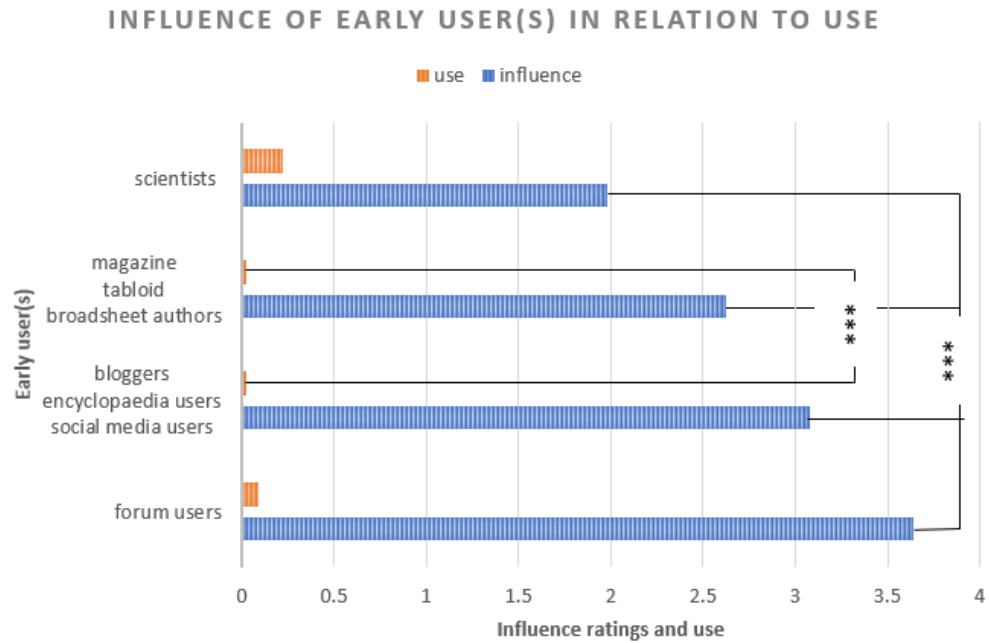
Different media types compared	Outcome t-test
Scientists - broadsheet	$t(290)=4.399, p<.001$
Broadsheet – tabloid journalists	$t(290)=-1.091, p=.276$
tabloid – magazine journalists	$t(290)=-.846, p=.398$
Magazine – encyclopaedia writers	$t(290)=-1.663, p=.097$
encyclopaedia writers - bloggers	$t(290)=1.740, p=.083$
Bloggers – social media users	$t(290)=.328, p=.743$
social media users – forum users	$t(290)=-5.162, p<.001$
Broadsheet/tabloid/magazines - encyclopaedia writers/bloggers/social media users	$t(874)=-6.632, p<.001$

Appendix 1.13 Graphs for quality and influence ratings in relation to use

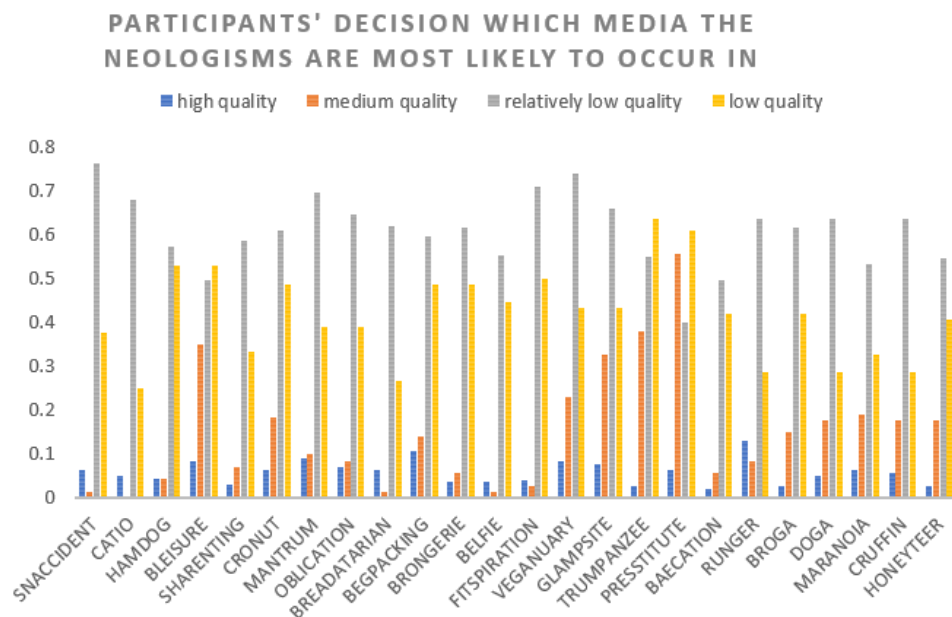
The blue bars indicate quality, with lower values implying higher quality. The orange bars show the summary of the overall usage for all neologisms belonging to the respective quality group, with higher values indicating more usage.



The blue bars indicate influence, with higher values standing for lower influence, the orange bars show the average use for the neologisms in the respective categories, with higher values indicating higher use.



Appendix 1.14 Participants' perception of media the neologisms mainly first occurred in



Appendix 1.15 Categorisation of neologisms for interest

ANIMALS	CATIO DOGA
CELEBRITIES	BELFIE FITSPIRATION MATRUM
FASHION	BRONGERIE
FOOD AND DRINKS	BREADATARIAN CRONUT CRUFFIN HAMDOG SNACCIDENT VEGANUARY
HEALTH AND BEAUTY	BELFIE BROGA FITSPIRATION SHARENTING SNACCIDENT VEGANUARY
POLITICS	TRUMPANZEE PRESSTITUTE
SCIENCE	SHARENTING MANTRUM
SPORTS	RUNGER MARANOIA BROGA DOGA FITSPIRATION
TRAVELLING	BEGPACKING GLAMPSITE BAECATION BLEISURE HONEYTEER OBLICATION

Appendix 1.16 All overall values for recognisability, transparency, appeal, and use

	RECOGNISABILITY	SEMANTIC TRANSPARENCY	FORMAL TRANSPARENCY	CONCEPTUAL APPEAL	FORMAL APPEAL	GENERAL APPEAL	USE
BAECATION	0.17	0.12	0.76	0.36	0.5	0.27	0.20
BEGPACKING	0.028	0.01	0.67	0.43	0.65	0.33	0.28
BELFIE	0.21	0.15	0.18	0.08	0.17	0.04	0.03
BLEISURE	0.03	0.03	0.12	0.28	0.26	0.18	0.11
BREADATARIAN	0.07	0.07	0.92	0.28	0.44	0.21	0.17
BROGA	0.04	0.03	0.76	0.23	0.5	0.23	0.13
BRONGERIE	0.01	0	0.5	0.22	0.21	0.18	0.04
CATIO	0.03	0	0.68	0.33	0.47	0.25	0.14
CRONUT	0.51	0.37	0.64	0.58	0.67	0.58	0.58
CRUFFIN	0.13	0.09	0.34	0.38	0.54	0.36	0.26
DOGA	0.19	0.16	0.62	0.43	0.53	0.36	0.16
FITSPIRATION	0.46	0.38	0.99	0.47	0.74	0.30	0.22

GLAMPSITE	0.69	0.73	0.80	0.76	0.92	0.70	0.66
HAMDOG	0.04	0.06	0.58	0.36	0.62	0.36	0.31
HONEYTEER	0	0	0.09	0.28	0.17	0.16	0.08
MANTRUM	0.14	0.19	0.93	0.53	0.74	0.49	0.43
MARANIOA	0.01	0	0.01	0.24	0.22	0.23	0.05
OBLICATION	0.01	0.01	0.46	0.36	0.30	0.31	0.22
PRESSTITUTE	0.04	0	0.84	0.34	0.26	0.20	0.18
RUNGER	0.03	0.01	0.27	0.31	0.38	0.34	0.18
SHARENTING	0.15	0.03	0.57	0.32	0.32	0.29	0.22
SNACCIDENT	0.33	0.25	0.83	0.54	0.72	0.44	0.32
TRUMPANZEE	0.08	0.04	0.96	0.53	0.54	0.28	0.27
VEGANUARY	0.62	0.62	0.99	0.54	0.74	0.57	0.49

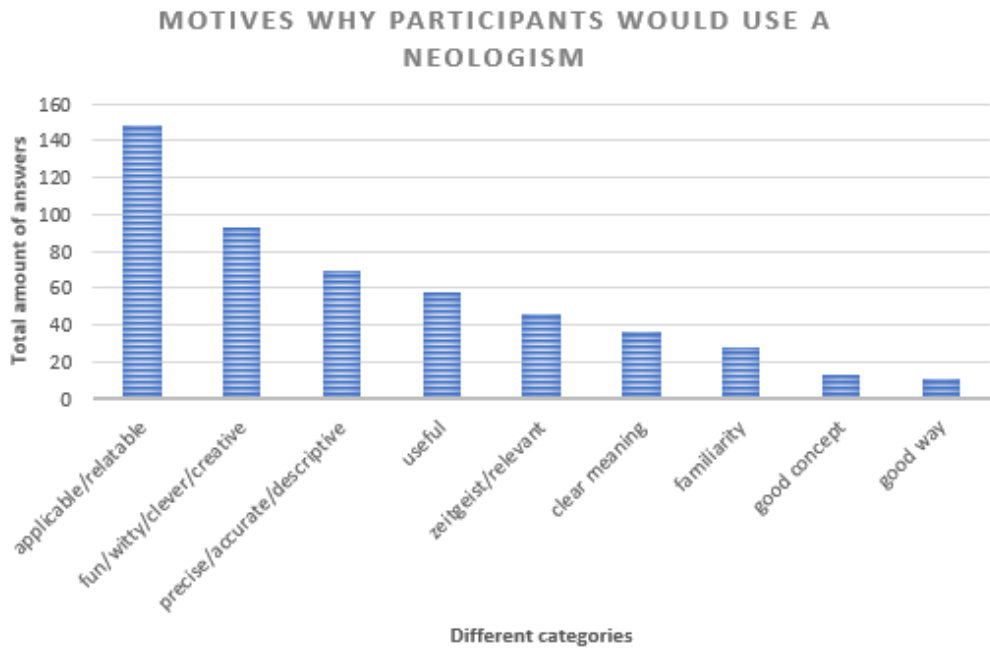
Appendix 1.17 Data sample *glampsite*

MOTIVES FOR USE

- [18972729](#) When I camp, I would like to glamp
- [18689119](#) well if we say glamping then glampsite just makes sense since it's like campsite
- [18541708](#) Gets point across
- [18502020](#) it's the only way to describe the concept
- [18473732](#) no strong feelings about this word
- [18433103](#) It is widely accepted and accurately describes this new trend.
- [18404949](#) May not use it personally as would be unlikely to go 'glamping', but think enough people already recognise it as a term that you could use it in conversation
- [18402834](#) I use it myself.
- [18347344](#) It is the best way to differentiate between regular camping.
- [18343059](#) it is the only type of camping I would do and glamping is becoming more popular
- [18313795](#) only if i ever needed to discuss glamping
- [18310878](#) Glamping is becoming more popular and so it is a good idea to have a word specifically to describe this type of activity.
- [18310898](#) I have frequent conversations about camping.
- [18301254](#) efficient way to say it
- [18300303](#) I dont see myself doing uniform of camping but I often use the term glamping therefore would use this
- [18297893](#) I like the idea of glamping and wouldn't mind trying it myself.
I have also heard it being used as a term to gently mock individuals who cannot face traditional forms of outdoor camping.
- [18296508](#) To describe something
- [18287403](#) Because I would only go camping if it was luxurious
- [18285352](#) The phrase is widespread in its use and can be easily understood, sometimes when going to a luxury camping site some people would not define this as camping so the term allows for leeway.
- [18283976](#) It distinguishes the type of site being referenced.
- [18283728](#) Immediately makes it clear what the word user is referring to.
- [18279839](#) Think it says what it means exactly
- [18279517](#) I'm aware of the process and have used it ironically when other people have gone glamping
- [18276210](#) It makes sense and I think people would understand what it means even if they had never heard of the word before.
- [18229203](#) When trying to convince my sister to come camping with me
- [18229130](#) Word/idea lots of people are familiar with so would use it in conversation.
Also I would consider glamping
- [18227195](#) to talk about a glampsite if i saw one
- [18217229](#) Sounds like what it means
- [18198527](#) I find this an apt descriptive word and would use it when it best fits the situation.
- [18198557](#) I like going to festivals so can see myself in situations where this word would be used
- [18194948](#) It describes a specific thing well
- [18194488](#) Glampsites are becoming more common and it's a place I would like to go
- [18180454](#) Might consider glamping for holidays
- [18176218](#) to describe a fancy campsite
- [18175070](#) I like the concept
- [18167881](#) 'glamping' is well-established and it follows closely from that
- [18165558](#) accustomed to the word
- [18163580](#) To make fun of it
- [18163564](#) Clear meaning; popular activity

- [18163517](#) To describe said accommodation
- [18162979](#) Why not
- [18162266](#) If talking about music festivals to friends.
- [18160983](#) I would probably go glamping myself.
- [18158465](#) It's quirky.
- [18160082](#) Meaning is obvious from "glamping"
- [18159331](#) good description
- [18158750](#) I think it's clever and conveys the meaning.
- [18158413](#) I think it is immediately clear to the speaker what is meant by this when you say this word
- [18157915](#) Have been glamping
- [18298196](#) i would

Appendix 1.18 Absolute values of motives for usage



Appendix 2 Questionnaire study II – Materials and data

Appendix 2.1 Questionnaire study II - sample

Neologism

0%

[Exit Survey](#) ↗

Dear Participant,

Many thanks for participating in this survey on new words. The study investigates new words and their usage amongst native speakers.

This is not about testing your language proficiency, you cannot get it wrong!

The questionnaire lasts about 12 minutes. As a thanks for participating there will be a raffle for several prizes. At the end of the questionnaire you will get a link that will enable you to participate in the raffle.

If you have any questions before, while or after completing the questionnaire, please feel free to contact Sabrina.link@abdn.ac.uk

[Next](#)

Neologism

0%

Questions marked with a * are required

[Exit Survey](#) 

* Before you start, please read the following:

- I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason
- I understand that the information provided by me will be held anonymously so that it is impossible to trace this information back to me individually. In accordance with the Data Protection Act this information may be retained indefinitely.

Next

Neologism

0%

Questions marked with a * are required

[Exit Survey](#) 

* Have you taken part in the previous study about new words?

- Yes
- No

Next

Participant's information

*** Sex**

- Female
- Male
- Non-binary
- I'd prefer not to say

*** Age**

- <25
- 26-45
- 46-65
- >66
- I'd prefer not to say

*** What is your highest level of education?**

- School education
- University degree
- None of the above
- I'd prefer not to say

*** Where do you live?**

- City/suburbs
- Countryside/village
- I'd prefer not to say

*** What is your native language?**

Media usage

* How much time do you spend on the Internet (including social media) on an average day?

- Less than an hour
- 1-3 hours
- 3-5 hours
- More than 5 hours
- Other

Next

Please answer the following questions referring to the respective words in the green box.

GLAMPSITE

* Have you come across this word before?

- Yes
- No

* Do you know the meaning of the word?

- Yes
- No
- I can guess it

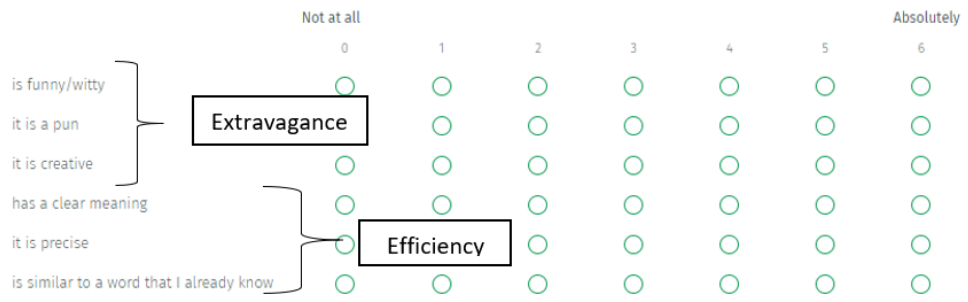
* I like the concept of the word.

	Strongly disagree	Disagree	Somewhat Disagree	Neither Agree nor Disagree	Somewhat agree	Agree	Strongly agree
GLAMPSITE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

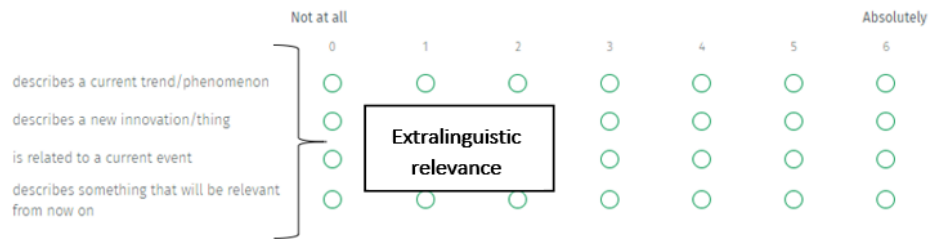
* The word describes 'a place to do glamping (luxurious camping)'. GLAMPSITE is a good way to say this.

	Strongly disagree	Disagree	Somewhat disagree	Neither Agree nor Disagree	Somewhat agree	Agree	Strongly agree
GLAMPSITE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* GLAMPSITE



* GLAMPSITE



*Would you use this word?



* Are any of these words relevant to you?

Drag your choices here to rank them

GLAMPSITE
BELFIE
SMOG
OBLIGATION
CRUFFIN
BRONGERIE
MOTEL
BRUNCH
HONEYTEER
NONE

Appendix 2.2 Complete overview of the demographics of questionnaire study II

	School				University				N/A or other				
	Female	Male	Non-binary	N/A	Female	Male	Non-binary	N/A	Female	Male	Non-binary	N/A	
City	15	4		1	43	22		1					<25
		1			8	11			1				26-45
	1				2	1				1			>46
													N/A
Countryside	4		1		15	5							<25
					3								26-45
					2								>46
													N/A
Prefer not to say													<25
													26-45
													>46
													N/A

Appendix 2.3 Overview of non-significant values regarding demography study II

IN RELATION TO USING A NEOLOGISMS:

AGE	$F(2, 140)=.579, p=.562$
TIME SPENT ON THE INTERNET	$F(3, 139)=1.682, p=.174$
LOCATION/PLACE OF LIVING	$t(141)=-1.466, p=.145$
EDUCATION	$F(2, 140)=.002, p=.998$

IN RELATION TO KNOWING A NEOLOGISMS:

AGE	$F(2, 140)=1.229, p=.296$
TIME SPENT ON THE INTERNET	$F(3, 139)=.207, p=.892$
LOCATION/PLACE OF LIVING	$t(141)=-.588, p=.557$
EDUCATION	$F(2, 140)=-.233, p=.793$
GENDER	$t(138)=-.583, p=.561$

Appendix 2.4 Standardised factor loadings CFA II

	A	B	C	D	E	F	G	H	I	J	K
OVERALL	.92	.89	.87	.68	.93	.82	.94	.74	.79	.75	.75
BELFIE	.91	.88	.89	.66	.82	.63	.43	.72	.80	.65	.68
BRONGERIE	.94	.90	.82	.78	.92	.69	.31	.74	.72	.66	.69
CRUFFIN	.87	.90	.30	.72	.92	.41	.90	.77	.77	.64	.66
GLAMPSITE	.72	.88	.62	.46	.89	.62	.87	.59	.24	.26	.44
HONEYTEER	.94	.85	.85	.63	.84	.73	.92	.77	.70	.69	.69
OBLICATION	.78	.87	.34	.57	.79	.29	.92	.84	.69	.67	.67

Appendix 2.5 Average rating use and appeal per word

NEOLOGISMS	AVERAGE USE	AVERAGE EFFICENCY	AVERAGE EXTRAVAGANCE	AVERAGE EXTRALINGUSTIC RELEVANCE
GLAMPSITE	0.49	0.77	0.66	0.63
OBLICATION	0.42	0.47	0.51	0.41
CRUFFIN	0.28	0.47	0.61	0.45
BELFIE	0.26	0.43	0.46	0.55
BRONGERIE	0.25	0.39	0.49	0.48
HONEYTEER	0.27	0.34	0.41	0.52

Appendix 3 Questionnaire study III - Materials and data

Appendix 3.1 Preliminary word list for the third study

Anglo neologisms	German neologisms
AGE SCAN	BLEISTIFTSTEMMER
BABYFACE	CHEFMASCHE
BABYSHOWER	DIGITALFALLE
BEACHWEAR	FLUGHAFEN-IMKER
BODYBOARDER	FUKUSHIMA-EFFEKT
BOOKCROSSER	GHETTOFAUST
CAKEPOP	INKUSIONSKIND
CHEATDAY	PFANDRING
CLICKBAIT	PLATTENURLAUB
CLOUDRAP	RACHEPORNOGRAPHIE
EGOSURFER	RISIKOGETRÄNK
FOODPORN	SCHWARMSTADT
FOODTRUCK	TOURISTENSTAU
GENDERSTAR	TRINKTOURISMUS
GLAMGIRL	VERBITTERUNGSPOPULISTEN
QUEERBAIR	VERSPÄTUNGSBONUS
UPPERCLASS-KID	WILLKOMMENSKLASSEN
ZERO WASTE	WISSENSCHAFTSREINIGUNG
	ZITTERATTACKE

Appendix 3.2 Definitions

FOODPORN	taking mouth-watering pictures of delicious foods and proliferating them throughout various social media websites as status updates (Urban Dictionary) Leute, die ihr mehr oder weniger leckeres (oder lecker aussehendes) Gericht auf dem Teller fotografieren und im Internet herzeigen (OWID)
FOODTRUCK	Is a licensed, motorized vehicle or mobile food unit that's used for selling food items to the general public (Urban Dictionary)
CLICKBAIT	It's a link which entices you to click on it. When a news article or link has a provocative title in order to get you to click on it (Urban Dictionary) Werbender Webinhalt in Form neugierig machender Überschriften oder reißerischer Texte auf einer Webseite, der den Nutzer dazu verführen soll, auf eine andere Webseite weiterzuklicken (OWID)
PORKDAY	Opposite to veggie day, to ensure that pork stays part of what kids in schools and nurseries get to eat
EGOSURFER	Person who googles himself/herself (Wikipedia)
IMAGEBOOST	Something you or a company does to improve their image
SCHWARMSTADT	Stadt, in die besonders viele junge Menschen ziehen (OWID)
WILLKOMMENS KLASSE	Für Flüchtlingskinder oder Kinder aus zugewanderten Familien ohne ausreichende Deutschkenntnisse eingerichtete Schulklasse, die auf den Besuch regulärer Schulklassen vorbereitet (wortbedeutung.info), Klasse mit Flüchtlingskindern oder Kindern aus zugewanderten Familien ohne Deutschkenntnisse, in der sie auf den Besuch in einer Regelschule vorbereitet werden sollen (OWID)
GHETTOFAUST	Informelle Form der Begrüßung, die oft auch in Filmen oder beim Sport zu sehen ist (Wikipedia), Faust, die zur Begrüßung, Aufmunterung, Anerkennung unter Jugendlichen gegen die Faust des Gegenübers gestoßen wird (OWID)
TRINKTOURISMUS	Tourismus, der lediglich dazu dient, zum Trinken ins Ausland zu fahren (v.a. in Mallorca)
BLEISTIFTSTEMMER	Jemand, der einer Schreibtischtätigkeit nachgeht. der größte physikalische Kraftaufwand, den diese Person zu leisten hat, sei angeblich das Anheben eines Schreibgeräts - Respektive Bleistifts - von der Schreibtischplatte (mundmische.de)
DIGITALFALLE	Versuch der Digitalisierung (eines Unternehmens), die aber nicht erfolgreich verläuft und mehr Probleme schafft als behebt

Neologismen

0%

Sehr geehrte Damen und Herren,

Vielen Dank für Ihre Teilnahme an dieser Studie zu neuen Wörtern im Deutschen. Die Studie untersucht neue Wörter und deren Gebrauch unter Muttersprachlern.

Es geht nicht darum, ihr Sprachkönnen zu testen, Sie können keine Fehler machen!

Der Fragebogen dauert circa 10 Minuten. Als Dank für Ihre Teilnahme können Sie am Ende an einem Gewinnspiel teilnehmen.

Falls Sie vor, während oder nach dem Ausfüllen des Fragebogens eine Frage haben, wenden Sie sich bitte an Sabrina.link@abdn.ac.uk

Weiter

Neologismen

0%

- Bevor Sie anfangen, lesen Sie bitte das Folgende:
 - Ich bin mir dessen bewusst, dass meine Teilnahme an der Fragebogenstudie komplett freiwillig ist und ich jederzeit, ohne Angabe eines Grundes, die Teilnahme an der Studie beenden kann.
 - Ich bin mir dessen bewusst, dass die Angaben, die ich mache, anonym sind, so dass es unmöglich ist, diese zu mir zurückzuverfolgen.

Weiter

Teilnehmer Informationen

• Geschlecht

- Weiblich
- Männlich
- Nicht binär
-

• Alter

- <25
- 26-45
- 46-65
- >66
- Keine Angabe

* Welchen Abschluss haben Sie?

- Gesamtschulabschluss
- Realschulabschluss
- Hauptschulabschluss
- Abitur
- Hochschulabschluss
- Keine Angabe
- Anderes

• **In welchem Land leben Sie?**

- Deutschland
- Anderes

• **Wo leben Sie?**

- Stadt
- Land
- Keine Angabe

* **Wieviel Zeit verbringen Sie durchschnittlich am Tag im Internet (inklusive auf sozialen Medien)?**

- Weniger als eine Stunde
- 1-3 Stunden
- 3-5 Stunden
- Mehr als 5 Stunden
- Anderes

* **Haben Sie dieses Wort schon einmal gesehen?**

- Ja
- Nein

* **Wissen Sie, was dieses Wort bedeutet?**

- Ja
- Nein

* Was ist die Bedeutung des Wortes?

* Das Wort beschreibt "einen Imbisswagen ". Das Wort FOODTRUCK eignet sich gut dafür, diese Bedeutung auszudrücken.

	Ich stimme überhaupt nicht zu	Ich stimme nicht zu	Ich stimme zu	Neutral	Ich stimme ein bisschen zu	Ich stimme zu	Ich stimme völlig zu
FOODTRUCK	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* Das Wort FOODTRUCK

	Überhaupt nicht						Absolut
	0	1	2	3	4	5	6
ist lustig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ist einem Wort, das ich schon kenne, ähnlich	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ist präzise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
hat eine klare Bedeutung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ist ein Wortspiel	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
ist kreativ	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* Mir gefällt das Konzept, das hinter dem Wort steht

	Ich stimme überhaupt nicht zu	Ich stimme nicht zu	Ich stimme kaum zu	Neutral	Ich stimme ein bisschen zu	Ich stimme zu	Ich stimme völlig zu
FOODTRUCK	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* Das Wort FOODTRUCK

	Überhaupt nicht						Absolut
	0	1	2	3	4	5	6
beschreibt ein zeitgenössisches Phänomen/einen Trend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
beschreibt eine neue Innovation/Erfindung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
steht in Verbindung mit einem aktuellen Ereignis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
beschreibt etwas, das von nun an relevant sein wird	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

* Würden Sie dieses Wort benutzen?

	Überhaupt nicht						Absolut
	0	1	2	3	4	5	6
FOODTRUCK	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix 3.4 Complete overview of the demographics for questionnaire III

		Gesamtschule			Hauptschule			Realschule			Gymnasium			University			NA					
		female	male	Nonbinary	female	male	Nonbinary	female	male	nonbinary	female	male	Nonbinary	female	male	Nonbinary	female	male	nonbinary			
City	Germany										4	3		2	2					>25		
											1			14	8						26-45	
			1									1			2	2						<45
																						NA
	Abroad														2	1						>25
												1			46	14	1					26-45
									1						6	5						<45
																						NA
Country side	Germany							1			1				1						>25	
					1	1				1		2	2		5	5						26-45
					1							1	1		1	2						<45
																						NA
	Abroad																					>25

Appendix 3.5 Overview of non-significant values regarding demography study
III

IN RELATION TO USAGE OF NEOLOGISMS:

AGE	$F(2, 153)=1.131, p=.326$
LOCATION/PLACE OF LIVING	$t(152)=-1.549, p=.123$
EDUCATION	$F(4, 148)=1.944, p=.106$
TIME SPENT ON THE INTERNET	$F(3,152)=.200, p=.896$
GENDER	$t(152)=-.572, p=.568$

IN RELATION TO KNOWING A NEOLOGISMS:

AGE	$F(2, 153)=2.789, p=.065$
LOCATION/PLACE OF LIVING	$t(152)=1.419, p=.158$
EDUCATION	$F(4, 148)=1.760, p=.140$
TIME SPENT ON THE INTERNET	$F(3,152)=2.255, p=.084$
GENDER	$t(152)=-.585, p=.560$

IN RELATION TO USING ANGLO NEOLOGISMS:

AGE	$F(2, 153)=.232, p=.793$
LOCATION/PLACE OF LIVING	$t(152)=-.589, p=.557$
EDUCATION	$F(4, 148)=.692, p=.599$
COUNTRY OF RESIDCENY	$t(154)=.737, p=.462$

IN RELATION TO KNOWING ANGLO NEOLOGISMS:

AGE	$F(2, 153)=2.582, p=.079$
LOCATION/PLACE OF LIVING	$t(152)=.256, p=.796$
EDUCATION	$F(4, 148)=.616, p=.652$
COUNTRY OF RESIDENCY	$t(154)=-1.523, p=.130$

Appendix 3.6 Model fit CFA I versus CFA II for questionnaire study III

	CHI-SQ (full)	CHI-SQ (reduced)	RMSEA (full)	RMSEA (reduced)	AIC (full)	AIC (reduced)	p-value (full)	p-value (reduced)
OVERALL	76.4	19.9	0.095	0.033	122.357	57.94	0	0.277
FOODTRUCK	71.9	40.2	0.134	0.141	117.941	78.228	0	0.001
WILLKOMMENSKLASSE	52.7	25.3	0.097	0.084	98.727	63.282	0.012	0.089
EGOSURFER	27.7	9.4	0	0	73.693	47.383	0.684	0.927
IMAGEBOOST	64.1	17.6	0.121	0.023	110.122	55.606	0.001	0.414
SCHWARMSTADT	64.5	20.1	0.121	0.052	110.487	58.141	0.001	0.267
FOODPORN	54	25.4	0.090	0.076	100.048	63.386	0.009	0.086
TRINKTOURISMUS	45.1	17.5	0.069	0.019	91.051	55.504	0.063	0.421
GHETTOFAUST	44.6	24	0.068	0.069	90.562	61.977	0.069	0.12
PORKDAY	58.8	37.4	0.099	0.119	104.765	75.415	0.003	0.003
DIGITALFALLE	33.5	12.9	0.024	0	79.529	50.917	0.393	0.742
CLICKBAIT	58	25.8	0.098	0.078	104.024	63.83	.003	0.078
BEISTIFTSTEMMER	36.6	19.9	0.46	0.05	82.572	57.928	0.265	0.278

Appendix 3.7 Standardised factors loadings for CFA I for questionnaire study III

	A	B	C	D	E	F	G	H	I	J	K	L	M
OVERALL	.41	.91	.86	.88	.67	.83	.47	.82	.82	.39	.79	.34	.40
FOODTRUCK	.12	1.15	.83	.81	.46	.73	.34	.66	.73	.76	.26	.35	.79
WILLKOMMENSKLASSE	0	.91	.84	.41	.48	.87	.66	.63	.72	.55	.50	.52	.41
EGOSURFER	.43	1	.87	.66	.61	.97	.63	.82	.47	.65	.56	.32	.43
IMAGEBOOST	.13	.90	1	.70	.71	.80	.64	.71	.59	.68	.39	.28	.49
SCHWARMSTADT	.50	.93	.88	.73	.64	.92	.64	.83	.42	.73	.64	.27	.34
FOODPORN	.39	.89	.71	.69	.37	.77	.27	.79	.67	.57	.74	.48	.51
TRINKTOURISMUS	.13	.74	.97	.86	.62	.91	.53	.80	.57	.72	.02	.35	.56
GHETTOFAUST	.36	1.01	.65	.82	.69	.77	.53	.68	.77	.54	.70	.45	.59
PORKDAY	.16	1	.81	.71	.73	.96	.67	.87	.69	.66	.54	.56	.66
DIGITALFALLE	.20	.88	.91	.66	.79	.97	.74	.74	.48	.58	.52	.27	.31
CLICKBAIT	.31	.91	.91	.69	.76	.91	.78	.82	.55	.26	.51	.28	.28
BEISTIFTSTEMMER	.40	.94	.74	.73	.80	.96	.74	.72	.79	.78	.59	.22	.23

Appendix 3.8 Standardised factor loadings for CFA II for questionnaire study III

	A	B	C	D	E	F	G	H	I	J	K
OVERALL	.41	.91	.86	.88	.67	.83	.47	.82	.82	.39	.79
FOODTRUCK	1.13	.84	.78	.48	.76	.68	.71	.28	.30	.46	.77
WILLKOMMENSKLASSE	.92	.83	.44	.50	.83	.84	.59	.64	.50	.43	.28
EGOSURFER	1.01	.87	.66	.61	.97	.61	.89	.58	.55	.34	.43
IMAGEBOOST	.90	1.00	.69	.71	.81	.55	.90	.53	.39	.44	.43
SCHWARMSTADT	.90	.91	.73	.64	.93	.70	.90	.60	.63	.24	.34
FOODPORN	.63	.69	.66	.39	.82	.56	.83	.23	.68	.46	.46
TRINKTOURISMUS	.64	1.13	.85	.62	.91	.73	.79	.52	-.01	.33	.56
GHETTOFAUST	.99	.67	.83	.69	.75	.61	.68	.51	.73	.45	.53
PORKDAY	.98	.83	.71	.73	.98	.68	.86	.68	.54	.54	.65
DIGITALFALLE	.90	.89	.66	.79	.97	.58	.69	.81	.52	.25	.28
CLICKBAIT	.92	.91	.70	.75	.91	.21	1.15	.56	.50	.18	.28
BEISTIFTSTEMMER	.87	.80	.73	.80	.98	.72	.74	.77	.63	.24	.32

Appendix 4 Follow-up experiment

Appendix 4.1 Preliminary word list for the fourth study

BRADIGAN	HEPEATING
BREATHATARIAN	JANXIETY
BREXCHOSIS	MEGXIT
BREXSICK	RANDONOUT
CORKITECTURE	THERAPET
ENTERTRAINMENT	WALKUMENTARY
FRANKENBEE	ZOODLE

Appendix 4.2 Definitions

BRADIGAN	a bra- and -cardigan mix first worn by Katie Holmes
CORKITECTURE	the use of cork as a building material
JANXIETY	feelings of unhappiness and worry that people often have at the beginning of a new year
MEGXIT	Meghan Markle and Prince Harry to step back from their senior roles in the British royal family.
RANDONOUT	someone who visits a random location generated by a computer bot in the hope of having an unusual, supernatural, or otherwise interesting experience there
WALKUMENTARY	a film or television programme or other event where someone walks around a particular place learning facts and information about the place or someone connected to it

Appendix 4.3 Questionnaire study IV - sample

Neologism

0%

[Exit Survey](#)

Dear Participant,

Many thanks for participating in this short study about new words. The study investigated how you perceive these new words.

It is not about your linguistic proficiency, you cannot get it wrong!

The questionnaire lasts about 10 minutes.

If you have any questions before, while or after completing the questionnaire, please feel free to contact sabirna.link@abdn.ac.uk

Next

Neologism

0%

Questions marked with a * are required

[Exit Survey](#) 

* Before you start, please read the following:

- I understand that participation in this study is entirely voluntary and that I can withdraw from the study at any time without giving a reason
- I understand that the information provided by me will be held anonymously so that it is impossible to trace this information back to me individually. In accordance with the Data Protection Act this information may be retained indefinitely.

Next

Neologism

2%

Questions marked with a * are required

[Exit Survey](#) 

Participant's information

* Sex

- Female
- Male
- Non-binary
- I'd prefer not to say

* Age

- <25
- 26-45
- 46-65
- >66
- I'd prefer not to say

* What is your highest level of education?

- School education
 - University degree
 - None of the above
 - I'd prefer not to say
-

* Where do you live?

- City/suburbs
 - Countryside/village
 - I'd prefer not to say
-

* What is your native language?

Please answer the following questions referring to the respective words in the green box.

MEGXIT

* Have you come across this word before?

- Yes
 - No
-

* Do you know the meaning of the word?

- Yes
- No

The word MEGXIT describes 'Meghan Markle and Prince Harry to step back from their senior roles in the British royal family.'

• The word MEGXIT is

	Not at all						Absolutely
	0	1	2	3	4	5	6
is funny/witty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
it is a pun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
it is creative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
has a clear meaning	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
it is precise	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
describes a current trend/phenomenon	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
is related to a current event	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
describes something that will be relevant from now on	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

• Would you use this word?

	Strongly disagree	Disagree	Somewhat disagree	Neither Agree nor Disagree	Somewhat agree	Agree	Strongly agree
GLAMPSITE	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Appendix 4.4 Complete overview of the demographics for questionnaire IV

	School				University				N/A or other				
	Female	Male	Non-binary	N/A	Female	Male	Non-binary	N/A	Female	Male	Non-binary	N/A	
City	1	2			7	2	1						<25
	1	1			3	3							26-45
													>46
													N/A
Countryside													<25
													26-45
									1				>46
													N/A
Prefer not to say													<25
													26-45
													>46
													N/A

Appendix 4.5 Efficiency, extravagance and extralinguistic relevance rating - study IV

NEOLOGISM	EFFICIENCY	EXTRAVAGANCE	EXTRALINGUISTIC RELEVANCE	OVERALL APPEAL
MEGXIT	0.67	0.56	0.77	0.67
JANXIETY	0.67	0.55	0.61	0.61
WALKUMENTARY	0.75	0.54	0.52	0.60
CORKITECTURE	0.65	0.55	0.44	0.55
BRADIGAN	0.52	0.48	0.43	0.48
RANDONOUT	0.38	0.46	0.48	0.44

Appendix 5 Forces affecting usage, conventionalisation and entrenchment taken from Schmid, 2020: 300

Forces affecting usage	Forces affecting conventionalization	Forces affecting entrenchment
conventionalization, entrenchment, and frequency of repetition	co-semiosis and co-adaptation	frequency of repetition self-priming
production circumstances	interpersonal activities and forces affecting them	similarity and analogy
cognitive economy	subjectivity and intersubjectivity	embodiment and other types of basic experience
communicative efficiency	identity and social order	salience
extravagance and expressivity	prestige and stigma	iconicity
foregrounding and salience	mobility	conventionalization and forces affecting it
politeness and impoliteness	multilingualism	
solidarity and distance	language contact	
power	entrenchment and forces affecting it	
	frequency of repetition	

Bibliography

About Words - A blog from Cambridge Dictionary. <https://dictionaryblog.cambridge.org/>. 30.09.2020.

Collins English Dictionary. <https://www.collinsdictionary.com>. 19.05.2020.

COSMAS I/II (Corpus Search, Management and Analysis System) © 1991-2020 Leibniz-Institut für Deutsche Sprache, Mannheim. <http://www.ids-mannheim.de/cosmas2/>. 16.08.2020.

MacMillian Dictionary Blog. <https://www.macmillandictionaryblog.com/>. 19.10.2020.

Merriam Webster Online Dictionary. <https://www.merriam-webster.com>. 19.05.2020.

New Words by Merriam Webster Online Dictionary. <https://www.merriam-webster.com/words-at-play/new-words-in-the-dictionary>. 28.10.2020.

Online Etymological Dictionary. <https://www.etymonline.com>. 19.05.2020.

Oxford Learner's Dictionary. <https://www.oxfordlearnersdictionaries.com/>. 20.05.2020.

Urban Dictionary. <https://www.urbandictionary.com/>. 28.10.2020.

Victoria & Albert Museum, Dundee: Now Accepting Contactless 2021.
<https://www.vam.ac.uk/dundee/exhibitions/now-accepting-contactless>. 23.11.2020

WordSpy: Dictionary of New Words. <https://wordspy.com/>. 28.10.2020.

Wortwarte. <https://wortwarte.de/>. 30.09.2020.

Oxford Reference. <https://www.oxfordreference.com/>. 08.12.2020.

Abbot-Smith, Kirsten & Michael Tomasello. 2006. "Exemplar-Learning and Schematization in a Usage-Based Account of Syntactic Acquisition". *The Linguistic Review*. 23 (3), 275-90.

- Abu-Algasim Mohamed, Entesar Alsir. 2020. "Newspapers as Carriers of New Terms Namely (Neologism) and Their role in Enhancing Students' Vocabulary: Guardian Newspaper as a Model". *International Journal of Linguistics, Literature and Translation (IJLLT)*. 3 (2), 180-96.
- Adamo, Giovanni & Valeria Della Valle. 2005. *2006 Parole Nuove*. 2nd. Milan: Sperling & Kupfer Editori.
- Adams, Valerie. 1973. *An Introduction to Modern English Word-Formation*. London: Longman.
- Adelman, James S., Gordon D. Brown & José F. Quesada. 2006. "Contextual Diversity, Not Word Frequency, Determines Word-Naming and Lexical Decision Times". *Psychological Science*. 17 (9), 814-23.
- Agha, Asif. 2003. "The Social Life of Cultural Value". *Language & Communication*. 23 (3), 231-73.
- Ahrens, R. 1977. "Wortfindungsstörungen für zusammengesetzte Worte (Nomina Composita) bei Aphasien". *Archiv für Psychiatrie und Nervenkrankheiten* 224, 73-87.
- Aitchison, Jean. 1991. *Language Change: Progress or Decay?* 2nd. Douglas, Isle of Man: Fontana.
- Aitchison, Jean. 1994. *Words in the Mind: An Introduction to the Mental Lexicon*. 2nd. Malden, Oxford, Victoria: Blackwell Publishing.
- Aitchison, Jean. 2001. *Language Change: Process or Decay?* 3rd. Cambridge: Cambridge University Press.
- Alegre, Maria & Peter Gordon. 1999. "Frequency Effects and the Representational Status of Regular Inflections". *Journal of Memory and Language*. 40, 41-61.
- Algeo, John. 1977. "Blends, a structural and systemic view". *American Speech*. 52, 47-64.
- Algeo, John. 1991. *Fifty Years Among the New Words: A Dictionary of Neologisms 1941-1991*. Cambridge: Cambridge University Press.
- Algeo, John. 1993. "Desuetude among New English Words". *International Journal of Lexicography*. 6 (4), 281-93.

- Ambridge, Ben & Elena Lieven. 2015. "A Constructivist Account of Child Language Acquisition". In: Brian MacWhinney & William O'Grady (eds.). *The Handbook of Language Emergence*. Chichester: John Wiley & Sons. 478-510.
- Andrews, Sally. 1986. "Morphological influences on lexical access: Lexical or nonlexical effects?". *Journal of Memory and Language*. 25 (7), 726-40.
- Arndt-Lappe, Sabine & Ingo Plag. 2013. "The role of prosodic structure in the formation of English blends". *English Language & Linguistics*. 17 (3), 537-63.
- Aronoff, Mark. 1976. *Word formation in generative grammar*. Cambridge, Mass: MIT Press.
- Aronoff, Mark & Frank Anshen. 2001. "Morphology and the lexicon: Lexicalization and productivity". In: Andrew Spencer & Arnold M. Zwicky (eds.). *The handbook of morphology (Blackwell Handbooks in Linguistics)*. Oxford: Basil Blackwell. 237-247.
- Aronoff, Mark & Kirsten Fudeman. 2005. *What is Morphology*. Oxford: Blackwell.
- Aronoff, Mark & Nanna Fuhrhop. 2002. "Restricting suffix combination in German and English: Closing suffixes and the monosuffix constraint". *Natural Language & Linguistic Theory*. 20, 451-90.
- Asif, Muhammed, Deng Zhiyong, Anila Iram & Maria Nisar. 2020. "Linguistic Analysis of Neologism Related to Coronavirus (COVID-19)". *Social Sciences & Humanities Open*.
- Auer, Peter & Frans Hinskens. 1996. "The convergence and divergence of dialects in Europe. New and not so new developments in an old area". *Sociolinguistica*. 10, 1-30.
- Auer, Peter & Frans Hinskens. 2005. "The role of interpersonal accommodation in a theory of language change". In: Peter Auer, Frans Hinskens & Paul Kerswill (eds.). *Dialect Change*. Cambridge: Cambridge University Press. 335-357.
- Augst, Gerhard. 1975. *Untersuchungen zum Morpheminventar der deutschen Gegenwartssprache (Forschungsberichte des Instituts für Deutsche Sprache, Mannheim 25)*. Tübingen: Narr.
- Ayto, John. 1990. *The Longman register of new words. Vol. II*. London: Longman.

- Ayto, John. 2007. *A Century of New Words*. Oxford: Oxford University Press.
- Baayen, Harald. 2010. "Demythologizing the Word Frequency Effect. A Discriminative Learning Perspective". *The Mental Lexicon*. 5 (3), 436-61.
- Baayen, Harald, Willem Levelt & Alette Haveman. 1993. *Research reported in annual report of the Max-Planck Institut für Psycholinguistik*. 6. Technical report. Nijmegen: Max-Planck-Institut.
- Baayen, Harald, Rochelle Lieber & Robert Schreuder. 1997. "The morphological complexity of simplex nouns". *Linguistics*. 35 (5), 861-78.
- Baayen, Harald, Cyrus Shaoul, Jon Willits & Michael Ramscar. 2016. "Comprehension without Segmentation: A Proof of Concept with Naive Discriminative Learning". *Language, Cognition and Neuroscience*. 31 (1), 106-28.
- Baayen, Harald, Viviane Van Casteren & Ton Dijkstra. 1993. *Research reported in annual report of the Max-Planck Institut für Psycholinguistik*, 6. Technical Report. Nijmegen: Max-Planck-Institut.
- Baddeley, Alan. 2000. "The Episodic Buffer: A New Component of Working Memory?". *Trends in Cognitive Sciences*. 4 (11), 417-23.
- Baddeley, Alan. 2010. "Working Memory". *Current Biology*. 20 (4), R136-40.
- Baddeley, Alan D. & Gregory Hitch. 1974. "Working Memory". In: Gordon H. Bower (ed.). *The Psychology of Learning and Motivation: Advances in Research and Theory*. New York: Academic Press. 47-89.
- Bailey, Guy, Tom Wike, Jan Tillery & Lori Sand. 1993. "Some Patterns of Linguistic Diffusion". *Language Variation and Change*. 5 (3), 359-90.
- Bakken, Kristin. 2006. "Lexicalization". In: Keith Brown (ed.). *Encyclopaedia of language and linguistics*. Vol VII. 2nd Oxford: Elsevier. 106-108.
- Barkema, Henk. 1996. "Idiomaticity and terminology: a multi-dimensional descriptive model". *Studia Linguistica*. 50 (2), 125-60.

- Barlow, Michael. 2013. "Individual Differences and Usage-Based Grammar". *International Journal of Corpus Linguistics*. 18 (4), 443-78.
- Barto, Andrew, Marco Mirolli & Gianluca Baldassarre. 2013. "Novelty or Surprise?". *Frontiers in Psychology*. 4 (907).
- Bat-El, Outi. 1996. "Selecting the best of the worse: The grammar of Hebrew blends". *Phonology*. 13, 283-328.
- Bat-El, Outi. 2006. "Blend". In: Keith Brown (ed.). *Encyclopaedia of language and linguistics*. 2nd ed. Amsterdam: Elsevier. 66-70.
- Bat-El, Outi & Evan-Gary Cohen. 2012. "Stress in English blends: A constraint-based approach". In: Vincent Renner, François Maniez & Pierre Arnaud (eds.). *Cross-disciplinary perspectives on lexical blending*. Berlin: de Gruyter. 193-212.
- Bauer, Laurie. 1983. *English Word-Formation*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 1988. *Introducing Linguistic Morphology*. Edinburgh: Edinburgh University Press.
- Bauer, Laurie. 1994. *Watching English change: An Introduction to the Study of Linguistic Change in Standard Englishes in the Twentieth Century*. Harlow: Longman.
- Bauer, Laurie. 1998. "Is there a class of neoclassical compounds, and if so is it productive?". *Linguistics*. 36 (3), 403-22.
- Bauer, Laurie. 2001. *Morphological Productivity*. Cambridge: Cambridge University Press.
- Bauer, Laurie. 2012a. "Blends: Core and periphery". In: Vincent Renner, François Maniez & Pierre Arnaud (eds.). *Cross-disciplinary perspectives on lexical blending*. Berlin: de Gruyter. 11-22.
- Bauer, Laurie. 2012b. *Introducing linguistic morphology*. Edinburgh: Edinburgh University Press.
- Bauer, Laurie. 2017. *Compounds and Compounding*. Cambridge: Cambridge University Press.

- Bauer, Tina. *Für die perfekte Reichweite: Das sind die aktuellen Top Hashtags auf Instagram*. <https://onlinemarketing.de/news/instagram-hacks-top-hashtags>. 09.06.2020.
- BBC News Asia. *Coronavirus disease named Covid-19*. <https://www.bbc.com/news/world-asia-china-51466362>. 24.08.2020.
- Becker, Thomas. 1992. "Compounding in German". *Rivista di Linguistica*. 4, 5-36.
- Bhera, Bhagavan & Priyadarshani Mishra. 2013. "The Burgeoning Usage of Neologisms in Contemporary English". *Journal Of Humanities And Social Science*. 18 (3), 25-35.
- Behrens, Heike. 2017. "The Role of Analogy in Language Processing and Acquisition". In: Marianne Hundt, Sandra Mollin & Simone Pfenninger (eds.). *The Changing English Language: Psycholinguistic Perspectives*. Cambridge: Cambridge University Press. 215-239.
- Bell, Allan. 1984. "Language Style as Audience Design". *Language in Society*. 13 (2), 145-204.
- Berger, Peter & Thomas Luckmann. 1966. *The Social Construction of Reality. A Treatise in the Sociology of Knowledge*. Harmondsworth: Penguin.
- Bertram, Raymond, Harald Baayen & Robert Schreuder. 2000. "Effects of family size for complex words". *Journal of Memory and Language*. 42, 390-405.
- Blank, Andreas. 2001. "Pathways of lexicalization". In: Martin Haspelmath, Ekkehard König, Wulf Oesterreicher & Wolfgang Raible (eds.). *Language Typology and Language Univerisals/La Typologie des Langues et les Universeaux Linguistiques/ Sprachtypologie und Sprachliches Universalien. An International Handbook/ Manuel International/ Ein Internationales Handbuch. Vol II*. Berlin: de Gruyter. 1596-1608.
- Bloomfield, Leonard. 1933. *Language*. Chicago: University of Chicago Press.
- Blumenthal-Dramé, Alice. 2012. *Entrenchment in usage-based theories: What corpus data do and do not reveal about the mind*. Berlin: de Gruyter.

- Blumenthal-Dramé, Alice. 2017. "Entrenchment from a Psycholinguistic and Neurolinguistic Perspective". In: Hans-Jörg Schmid (ed.). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistics Knowledge*. Berlin, New York: de Gruyter. 129-152.
- Blythe, Richard & William A. Croft. 2009. "The speech community in evolutionary language dynamics". *Language Learning*. 59, 269-304.
- Boulanger, Jean-Claude. 2010. "Sur l'existence des concepts de "néologie" et de "néologisme". Propos sur un paradoxe lexical et historique". In: M. T. Cabré, Ona Domènech Bagaria, Rosa Estopà, Judit Freixa Aymerich & Mercè Lorente Casafont (eds.). *Actes del I Congrés Internacional de Neologia de les Llengües Romàniques*. Barcelona: Universitat Pompeu Fabra. 31-73.
- Bourdieu, Pierre. 1991. *Language and Symbolic Power*. Cambridge, MA: Harvard University Press.
- Bowen, Natasha & Shenyang Guo. 2012. *Structural Equation Modeling*. Oxford: Oxford University Press.
- Boyd, Jeremy K. & Adele E. Goldberg. 2011. "Learning What Not to Say: The Role of Statistical Preemption and Categorization in a-Adjective Production". *Language*. 87 (1), 55-83.
- Brace, Ian. 2004. *Questionnaire Design: How to plan, structure and write survey material for effective market research*. London: Kogan Page.
- Braungart, Georg, Harald Fricke, Klaus Grubmüller, Jan-Dirk Müller, Friedrich Vollhardt & Klaus Weimer. 2007. *Reallexikon der deutschen Literaturwissenschaft: Neubearbeitung des Reallexikons der deutschen Literaturgeschichte. Band I. A-G*. Berlin, New York: de Gruyter.
- Brinton, Laurel J. & Elizabeth Traugott. 2005. *Lexicalization and Language Change*. Cambridge: Cambridge University Press.
- Britain, David. 2010. "Contact and Dialectology". In: Raymond Hickey (ed.). *The Handbook of Language Contact*. Chichester: John Wiley & Sons. 209-229.
- Brown, Timothy. 2015. *Confirmatory Factor Analysis for Applied Research, Second Edition*. New York: Guildford Publications.

- Bryant, Fred B. & Albert Satorra. 2012. "Principles and Practice of Scaled Difference Chi-Square Testing". *Structural Equation Modeling: A Multidisciplinary Journal*. 19 (3), 372-98.
- Bucholtz, Mary. 1999. "'Why be Normal?': Language and Identity Practices in a Community of Nerd Girls". *Language in Society*. 28 (2), 203-23.
- Buchsbaum, Bradley R. 2016. "Chapter 69—Working Memory and Language". In: Gregory Hickok & Steven L. Small (eds.). *Neurobiology of Language*. San Diego, CA: Academic Press. 863-875.
- Burkhardt, Armin. 2013. "Die „Anglizismen-Frage“ aus der Sicht der GfdS". *Sprachreport*. 29, 38-42.
- Bussmann, Hadumod. 1996. *Routledge Dictionary of Language and Linguistics*. London and New York: Routledge.
- Butterworth, Brian. 1983. "Lexical representation". In: Brian Butterworth (ed.). *Language production (Vol. II): Development, Writing and Other language Processes*. London: Academic Press. 257-294.
- Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam and Philadelphia, PA: Benjamins.
- Bybee, Joan. 1995. "Regular Morphology and the Lexicon". *Language and Cognitive Processes*. 10 (5), 425-55.
- Bybee, Joan. 2006. "From usage to grammar: The mind's response to repetition". *Language*. 82, 711-33.
- Bybee, Joan. 2007. *Frequency of Use and the Organization of Language*. Oxford: Oxford University Press.
- Bybee, Joan. 2010. *Language, Usage and Cognition*. Cambridge: Cambridge University Press.
- Bybee, Joan & Jay McClelland. 2005. "Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition". *The Linguistic Review*. 22, 381-410.
- Cabré, M. Teresa. 1993. *La terminología. Teoría, metodología, aplicaciones*. Barcelona: Empúries.
- Cabré, M. Teresa. 2015. "La neologia: un nou camp a la cerca de la seva consolidació científica". *Caplletra*. 59, 125-36.

- Cabré, M. Teresa, Ona Domènech & Rosa Estopà. 2014. *Mots nous en català / New words in Catalan. Una panoràmica geolectal / A diatopic view*. Amsterdam, Philadelphia: Benjamins.
- Cameron, Deborah, & Don Kulick. 2003. *Language and Sexuality*. Cambridge: Cambridge University Press.
- Cannon, Garland. 1986. "Blends in English word formation". *Linguistics*. 24, 725-53.
- Caramazza, Alfonso, Albert Costa, Yanchao Bi & Michele Miozzo. 2001. "The Specific-Word Frequency Effect: Implications for the Representation of Homophones in Speech Production". *Journal of Experimental Psychology: Learning, Memory, and Cognition*. 27 (6), 1430-1450.
- Caramazza, Alfonso, Alessadro Laudanna & Cristina Romani. 1988. "Lexical Access and Inflectional Morphology". *Cognition*. 28. 297-332.
- Caramazza, Alfonso, Gabriele Miceli, M. Caterina Silveri & Alessandro Laudanna. 1985. "Reading Mechanisms and the Organisation of the Lexicon: Evidence from Acquired Dyslexia". *Cognitive Neuropsychology*. 2 (1), 81-114.
- Cartier, Emmanuel. 2017. "Neoveille, a Web Platform for Neologism Tracking". In: Proceedings of the EACL 2017 Software Demonstrations. Valencia, Spain. 95-98.
- Casasanto, Daniel & Gary Lupyan. 2015. "All Concepts Are Ad Hoc Concepts". In: Eric Margolis & Stephen Laurence (eds.). *The Conceptual Mind: New Directions in the Study of Concepts*. Cambridge, MA: MIT Press. 543-566.
- Chaffin, Roger, Robin Morris & Rachel Seely. 2001. "Learning New Word Meanings from Context: A Study of Eye Movements". *Journal of Experimental Psychology Learning Memory and Cognition*. 27 (1), 225-35.
- Chambers, Jack K. 2002. "Patterns of Variation Including Change ". In: Jack K. Chambers, Peter Trudgill & Natalie Schilling-Estes (eds.). *The Handbook of Language Variation and Change*. Oxford: Blackwell. 349-372.
- Chambers, Jack K. & Peter Trudgill. 1998. *Dialectology*. 2nd. Cambridge: Cambridge University Press.

- Chen, Jenn. *Social media demographics to inform your brand's strategy in 2020*. <https://sproutsocial.com/insights/new-social-media-demographics/>. 04.06.2020.
- Chialant, Doriana & Alfonso Caramazza. 1995. "Where is morphology and how is it processed? The case of written word recognition". In: Laurie Feldman (ed.). *Morphological aspects of language processing*. Lawrence: Erlbaum Associates. 55-76.
- Christiansen, Morten H. & Nick Chater. 2016b. "The Now-or-Never Bottleneck: A Fundamental Constraint on Language". 39, 1-72.
- Christiansen, Morton H. & Nick Chater. 2016a. *Creating Language. Integrating Evolution, Acquisition, and Processing*. Cambridge, MA, and London: MIT Press.
- Clark, Andy. 2013. "Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science". *Behavioral Brain Science*. 36 (3), 181-204.
- Clark, Eve. 1993. *The Lexicon in Acquisition*. Cambridge: Cambridge University Press.
- Clark, Herbert. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Clauzure, Émilie. 2003. "Le term *neologism* dans le monde anglo-saxon: origine et évolution". *Sablayrolles*. 207-23.
- Coseriu, Eugenio. 1958/1973. *Sincronía, diachronía, e historia. El problema del cambio lingüístico*. 2nd. Madrid: Biblioteca Romantica Hispanica.
- Coseriu, Eugenio. 1967. "Das Phänomen der Sprache und das Daseinsverständnis des heutigen Menschen". *Pädagogische Provinz*. 1, 11-28.
- Coupland, Nikolas. 1984. "Accommodation at work: some phonological data and their implications". *International Journal of the Sociology of Language*. 46, 49-70.
- Cowan, Nelson. 2001. "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity". *The Behavioral and Brain Sciences*. 24 (1), 87-114.

- Creswell, John & Vicki Plano Clark. 2010. *Designing and conducting mixed methods research*. 2nd. Thousand Oaks, CA: SAGE.
- Croft, William A. 2000. *Explaining Language Change*. Harlow: Longman.
- Croft, William A. & Alan D. Cruse. 2004. *Cognitive linguistics*. Cambridge University Press: Cambridge.
- Crystal, David. 1995. *The Cambridge encyclopaedia of the English language*. Cambridge: Cambridge University Press.
- Crystal, David. 1998. *Language Play*. London: Penguin.
- Crystal, David. 2002 [1995]. *The Cambridge Encyclopaedia of the English Language*. Cambridge: Cambridge University Press.
- Crystal, David. 2003. *English as a global language*. 2nd. Cambridge: Cambridge University Press.
- Cukor-Avila, Patricia & Guy Bailey. 2001. "The Effects of the Race of the Interviewer on Sociolinguistic Fieldwork". *Journal of Sociolinguistics*. 5 (2), 252-70.
- Culpeper, Jonathan & Minna Nevala. 2012. "Sociocultural Processes and the History of English". In: Terttu Nevalainen & Elizabeth Traugott (eds.). *The Oxford Handbook of the History of English*. Oxford: Oxford University Press. 365-391.
- Czerwenka, Christine. 2007. *Substantiv + Substantiv im Spanischen: Wortbildung und Grammatik*. Frankfurt: Peter Lang.
- de Saussure, Ferdinand. 1916. *Cours de linguistique générale*. Lausanne; Paris: Payot.
- De Smet, Herbert. 2016. "How gradual change progresses: The interaction between convention and innovation". *Language Variation and Change*. 28, 83-102.
- De Smet, Herbert & Hanne Cuyckens. 2007. "Diachronic aspects of complementation: Constructions, entrenchment and the matching-problem". In: Christopher Cain & Geoffrey Russom (eds.). *Studies in*

- the history of the English language III: Managing chaos: Strategies for identifying change in English.*
Berlin, New York: de Gruyter. 1-37.
- de Vaan, Laura, Robert Schreuder & Harald Baayen. 2007. "Regular morphologically complex neologisms leave detectable traces in the mental lexicon". *The Mental Lexicon*. 2 (1), 1-24.
- Demske, Ulrike. 2006. "Kommentar. Über Wolfgang Motsch: Deutsche Wortbildung in Grundzügen. Berlin/New York: de Gruyter, 1999.". *Zeitschrift für germanistische Linguistik*. 29 (1), 75-84.
- Dewaele, Jean-Marc. 2018. "Online Questionnaires". In: Aek Phakiti, Peter De Costa, Luke Plonsky & Sue Starfield (eds.). *The Palgrave Handbook of Applied Linguistics Research Methodology*. 269-286.
- Di Sciullo, Anna-Maria & Edwin Williams. 1987. *On the Definition of Word*. Cambridge Mass.: MIT Press.
- Diessel, Holger. 2007. "Frequency Effects in Language Acquisition, Language Use, and Diachronic Change". *New Ideas in Psychology*. 25 (2), 108-27.
- Donalies, Elke. 1996. "Da keuchgrinste sie süßsauerlich. Über kopulative Verb- und Adjektivkomposita". *Zeitschrift für germanistische Linguistik*. 24, 273-86.
- Donalies, Elke. 2005. *Die Wortbildung des Deutschen. Ein Überblick*. Tübingen: Narr.
- Dörnyei, Zoltán. 2007. *Research methods in applied linguistics*. Oxford: Oxford University Press.
- Dörnyei, Zoltán & Tatsuya Taguchi. 2010. *Questionnaires in second language research: Construction, administration, and processing*. 2nd. London: Routledge.
- Downing, Pamela. 1977. "On the Creation and Use of English Compound Nouns". *Language*. 53, 810-42.
- Drieghe, Denis, Alexander Pollatsek, Barbara Juhasz & Keith Rayner. 2010. "Parafoveal processing during reading is reduced across a morphological boundary". *Cognition*. 116, 136-42.
- Du Bois, John. 2014. "Towards a Dialogic Syntax". *Cognitive Linguistics*. 25 (3), 359-410.
- Dunning, Ted. 1993. "Accurate Methods for the Statistics of Surprise and Coincidence". *Computational Linguistics*. 19 (1), 61-74.

- Eckert, Penelope. 2000. *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Malden, MA: Blackwell Publishers.
- Eckert, Penelope. 2012. "Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation". *Annual Review of Anthropology*. 41 87-100.
- Eckert, Penelope & Sally McConnell-Ginet. 1992. "Communities of Practice: Where Language, Gender, and Power All Live". In: Kira Hall, Mary Bucholtz & Birch Moonwomon (eds.). *Locating Power. Proceedings of the 1992 Berkeley Women and Language Conference*. Berkeley, CA: Berkeley Women and Language Group, University of California. 89-99.
- Eddington, Davin. 2004. "Issues in modeling language processing analogically". *Lingua*. 114 (7), 849-71.
- Edwards, John. 2009. *Language and Identity: An introduction*. Cambridge: Cambridge University Press.
- Eitelmann, Matthias & Dagmar Haumann. 2019. "Extravagant Morphology. Workshop Proposal SLE 52". London.
- Elsner, Sylvia & Walter Huber. 1995. "Die Verarbeitung von N-N-Komposita bei Aphasie". *Neurolinguistik*. 9, 59-80.
- Ettinger, Stefan. 1974. *Diminutiv- und Augmentativbildung: Regeln und Restriktionen*. Tübingen: Tübinger Beiträge zur Linguistik.
- Fairclough, Norman. 2001. *Language and Power*. 2nd. London: London.
- Fairon, Cedrik, Kevin Mace & Hubert Naets. 2008. "GlossaNet 2: a linguistic search engine for RSS-based corpora". In: Stefan Evert, Adam Kilgarriff & Serge Sharoff (eds.). *Proceedings of the 4th Web as Corpus Workshop (WAC-4) - Can we beat Google?* Marrakech: cental.fltr.ucl.ac.be/team/~ced/papers/2008-wac4-glossanet.pdf.
- Falk, Ingrid, Delphine Bernhard and Christophe Gérard. 2014. "From Non Word to New Word: Automatically Identifying Neologisms in French Newspapers". In: Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). 4337-4344.

- Felder, Ekkehard. *Willkür im Duden*. <https://www.spektrum.de/kolumne/willkuer-im-duden/1494781>. 21.10.2020.
- Ferlazzo, Larry & Katie Hull Sypniewski. 2012. *The ESL / ELL Teacher's Survival Guide: Ready-to-Use Strategies, Tools, and Activities for Teaching English Language Learners on all Levels*. San Francisco: John Wiley & Sons.
- Fernández Fernández, Maximiliano. 2004. "La incorporación de neologismos al español actual. Extranjerismos y nuevas acepciones". *Lección inaugural en la solemne apertura del curso académico 2004/2005*. Ávila: Universidad Católica de Ávila.
- Fernández-Domínguez, Jesus. 2010. "Productivity vs lexicalisation: frequency-based hypotheses on word-formation". *Poznań Studies in Contemporary Linguistics*. 46 (2), 193-219.
- Fiez, Julie A. 2016. "Chapter 68—Neural Basis of Phonological Short-Term Memory". In: Gregory Hickok & Steven L. Small (eds.). *Neurobiology of Language*. San Diego, CA: Academic Press. 855-862.
- Fiorentino, Robert & David Poeppel. 2007. "Compound words and structure in the lexicon". *Language and Cognitive Processes*. 22, 953-1000.
- Fischer, Roswitha. 1998. *Lexical Change in Present-Day English. A Corpus Based Study of Motivation, Institutionalisation, and Productivity of Creative Neologisms*. Tübingen: Narr.
- Fleischer, Wolfgang & Irmhild Barz. 1995. *Wortbildung der deutschen Gegenwartssprache*. Tübingen: Niemeyer.
- Fletcher, William. 2007. "Concordancing the web: promise and problems, tools and techniques". In: Marianne Hundt, Nadja Nesselhauf & Carolin Biewer (eds.). *Corpus Linguistics and the Web*. 25-45.
- Forster, Kenneth I. & Susan M. Chambers. 1973. "Lexical Access and Naming Time". *Journal of Verbal Learning and Verbal Behavior*. 12 (6), 627-35.
- Fowler, Henry W. 2015. *Fowler's Concise Dictionary of Modern English Usage*. 3rd. Oxford: Oxford University Press.

- Fowler, Henry W. & Francis G. Fowler. 1954. *The King's English. First published edition 1906*. Oxford: Clarendon Press.
- Friston, Karl. 2010. "The Free-Energy Principle: A Unified Brain Theory?". *Nature Reviews. Neuroscience*. 11 (2), 127-38.
- Frost, Ram, Blair C. Armstrong, Noam Siegelman & Morten H. Christiansen. 2015. "Domain Generality vs. Modality Specificity: The Paradox of Statistical Learning". *Trends in Cognitive Sciences*. 19 (3), 117-25.
- Fuhrhop, Nanna. 1996. "Fugenelemente". In: Ewald Lang & Gisela Zifonun (eds.). *Deutsch typologisch. Jahrbuch 1995 des IDS. Berlin*. New York: de Gruyter. 525-550.
- Fuhrhop, Nanna. 1998. *Grenzfälle morphologischer Einheiten*. Tübingen: Stauffenburg.
- Fuhrhop, Nanna. 2000. "Zeigen Fugenelemente die Morphologisierung von Komposita an?". In: Rolf Thieroff, Nanna Fuhrhop & Matthias Tamrat (eds.). *Deutsche Grammatik in Theorie und Praxis*. Tübingen: Niemeyer. 201-213.
- Gaskell, M. Gareth & Nicolas Dumay. 2003. "Lexical competition and the acquisition of novel words". *Cognition*. 89 (2), 105-32.
- Geeraerts, Dirk, Dirk Grondelaers & Peter Bakema. 1994. *The Structure of Lexical Variation. A Descriptive Framework for Cognitive Lexicology*. Berlin: de Gruyter.
- Giles, Howard, Nikolas Coupland & Justine Coupland. 1991. "Accommodation theory. Communication, context, and consequences". In: Howard Giles, Nikolas Coupland & Justine Coupland (eds.). *Contexts of Accommodation*. Cambridge, Paris: Cambridge University Press, Editions de la maison des sciences de l'homme. 1-68.
- Giles, Howard & Tania Ogay. 2007. "Communication Accommodation Theory". In: Bryan B. Whaley & Wendy Samter (eds.). *Explaining Communication: Contemporary Theories and Exemplars*. Mahwah, NJ: Lawrence Erlbaum. 293-310.

- Givon, Talmy. 1971. "Historical Syntax & Synchronic Morphology: An Archaeologist's Field Trip". *Chicago Linguistics Society*. 7 (1), 394-415.
- Gleick, James. 2006. "Cyber-Neologolieferation. The Oxford English Dictionary meets cyberspace". *New York Times Magazine* 5.
- Gleitman, Henry & Lila Gleitman. 1979. "Language use and language judgment". In: Charles Fillmore, Daniel Kempler & William Wang (eds.). *Individual Differences in Language Ability and Language Behavior*. New York: Academic. 103-126.
- Gobet, Fernand, Peter C. Lane, Steve Croker, Peter C. -H Cheng, Gary Jones, Iain Oliver & Julian M. Pine. 2001. "Chunking Mechanisms in Human Learning". *Trends in Cognitive Sciences*. 5 (6), 236-43.
- Goldberg, Adele. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.
- Goldberg, Adele. 2019. *Explain Me This: Creativity, Competition, and the Partial Productivity of Constructions*. Princeton, NJ: Princeton University Press.
- Golinkoff, Roberta. 1975. "A Comparison of Reading Comprehension Processes in Good and Poor Comprehenders". *Reading Research Quarterly*. 11 (4), 623-59.
- Grainger, Jonathan, J., Kevin O'Regan, Arthur M. Jacobs & Juan Segui. 1989. "On the Role of Competing Word Units in Visual Word Recognition: The Neighborhood Frequency Effect". *Perception & Psychophysics*. 45 (3), 189-95.
- Greenberg, Joseph. 1966. *Language Universals*. Berlin, New York: de Gruyter.
- Gries, Stefan. 2005. "Syntactic Priming: A Corpus-Based Approach". *Journal of Psycholinguistic Research*. 34 (4), 365-99.
- Gries, Stefan. 2006. "Cognitive determinants of subtractive word-formation processes: A corpus-based perspective". *Cognitive Linguistics*. 17 (4), 535-58.
- Grimm, Jacob. 1826. *Deutsche Grammatik II. Nachdruck der Ausgabe 1878, Hildesheim 1967*.

- Günther, Franziska. 2016. *Constructions in Cognitive Contexts. Why Individuals Matter in Linguistic Relativity Research*. Berlin and Boston, MA: de Gruyter.
- Günther, Franziska, Thomas Geyer & Hermann Müller. 2017. "Salience, Attention, and Perception". In: Hans-Jörg Schmid (ed.). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Boston, MA, and Berlin: APA and de Gruyter. 289-311.
- Habermas, Jürgen. 1979. *Communication and Evolution of Society*. Toronto: Beacon Press.
- Haentjes, Willie. *Gibt es Wörter, die niemand braucht?* <https://www.bild.de/news/inland/duden/nimmt-5000-neue-woerter-auf-52790202.bild.html>. 21.10.2020.
- Hägerstrand, Torsten. 1952. *The Propagation of Innovation Waves*. London: Royal University of Lund, Dept of Geography.
- Häikiö, Tuomo, Raymond Bertram & Jukka Hyönä. 2010. "Development of parafoveal processing within and across words in reading: Evidence from the boundary paradigm". *The Quarterly Journal of Experimental Psychology*. 63, 1982-1998.
- Halle, Morris. 1973. "Prolegomena to a theory of word formation". *Linguistic Inquiry*. 4 (1), 3-16.
- Hamblin, Robert L., Jerry L. Miller & D. Eugene Saxton. 1979. "Modeling Use Diffusion". *Social Forces*. 57, 799-811.
- Hanisch, Horst. 2018. *Ü65- und Senioren-Knigge 2100: Die jungen Alten und die alten Jungen. Kommunikation und Verständnis zwischen den Generationen. Einsamkeit und technischer Fortschritt*. 3rd. Norderstedt: Books of Demand.
- Hartsuiker, Robert & Agnes Moors. 2017. "On the automaticity of language processing". In: Hans-Jörg Schmid (ed.). *Language and the human lifespan series. Entrenchment and the psychology of language learning: How we reorganize and adapt linguistic knowledge*. Oxford: Oxford University Press. 201-225.
- Haspelmath, Martin. 1999. "Why is Grammaticalization irreversible?". *Linguistics*. 37, 1043-68.

- Hayes, Andrew F. 2018. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. New York, London: Guildford Press.
- Hempel, Heinrich. 1980. *Bedeutungslehre und allgemeine Sprachwissenschaft: Sprachtheoretisch-linguistisch Arbeiten 1952-1973*. Tübingen: Narr.
- Henrich, Joseph & Francisco J. Gil-White. 2001. "The Evolution of Prestige: Freely Conferred Deference as a Mechanism for Enhancing the Benefits of Cultural Transmission". *Evolution and Human Behavior*. 22 (3), 165-96.
- Henzen, Walter. 1965. *Deutsche Wortbildung*. Tübingen: Niemeyer Verlag.
- Herberg, Dieter. 2002. "Neologismen in der deutschen Gegenwartssprache. Probleme ihrer Erfassung und Beschreibung". *Deutsch als Fremdsprache*. 39 (4), 195-200.
- Herberg, Dieter, Michael Kinne, Doris Steffens, Elke Tellenbach & Doris Al-Wadi. 2004. *Neuer Wortschatz: Neologismen der 90er Jahre im Deutschen*. Berlin: de Gruyter.
- Herbermann, Clemens-Peter. 1981. *Wort, Basis, Lexem und die Grenze zwischen Lexikon und Grammatik*. München: Fink.
- Herbst, Thomas. 2010. *English Linguistics: A Coursebook for Students of English*. Berlin: de Gruyter.
- Hoffmann, Sebastian. 2004. "Are low-frequency complex prepositions grammaticalized? On the limits of corpus data - and the importance of intuition". In: Hans Lindquist & Christian Mair (eds.). *Corpus Approaches to Grammaticalization in English*. Amsterdam, Philadelphia: Benjamins. 171-210.
- Hohenhaus, Peter. 1996. *Ad-hoc Wortbildung. Terminologie, Typologie und Theorie kreativer Wortbildung Im Englischen*. Frankfurt am Main: Peter Lang.
- Hohenhaus, Peter. 2005. "Lexicalisation and Institutionalisation". In: Pavol Štekauer & Rochelle Lieber (eds.). *Handbook of Word-Formation*. Dordrecht: Springer. 17-27.

- Hohenhaus, Peter. 2006. "*Bouncebackability*. A Web-as-corpus-based case study of a new formation, its interpretation, generalization/spread and subsequent decline". *SKASE Journal of Theoretical Linguistics*. 3 (2), 17-27.
- Hohenhaus, Peter. 2007. "How to do (even more) things with nonce words (other than naming)". In: Judith Munat (ed.). *Lexical Creativity, Texts and Contexts*. Amsterdam: Benjamins. 15-38.
- Holmes, Janet. 2008. *An Introduction to Sociolinguistics*. 3rd. London: Pearson Longman.
- Hooper, Daire, Joseph Coughlan & Michael Mullen. 2008. "Structural equation modelling: Guidelines for determining model fit". *Journal of Business Research Methods*. 6, 53-60.
- Hopper, Paul & Elizabeth Traugott. 2003. *Grammaticalization*. 2nd. Cambridge: Cambridge University Press.
- Huang, Yanping & Rajesh Rao. 2011. "Predictive Coding", *Wiley Interdisciplinary Reviews*. 2 (5), 580-93.
- Inhoff, Albrecht, Deborah Briihl & Jaqueline Schwartz. 1996. "Compound word effects differ in reading, on-line naming, and delayed naming tasks". *Memory & Cognition*. 24, 466-76.
- Inhoff, Albrecht & Keith Rayner. 1986. "Parafoveal word processing during eye fixations in reading: Effects of word frequency". *Perception & Psychophysics*. 40 (5), 431-9.
- Jackendoff, Ray. 1997. *The Architecture of the Language Faculty*. 2nd. Massachusetts: Linguistic enquiry monographs.
- Jaeger, Jeri. 1986. "On the acquisition of abstract representations for English vowels". *Phonology Yearbook*. 3, 71-97.
- Jesenšek, Vida. 1995. "Medienwirksame Neologismen in der deutschen Gegenwartssprache". *Linguistica*. 35 (2), 171-207.
- Johnson, James. *Social media usage in the United Kingdom (UK) - Statistics & Facts*.
<https://www.statista.com/topics/3236/social-media-usage-in-the-uk/#:~:text=At%20the%20beginning%20of%202017,percent%20and%2062%20percent%20respecti>

[vely.&text=Women%20in%20the%20UK%20were,social%20media%20profile%20than%20men.](#)

04.06.2020.

Johnson, Rebecca, Sarah Rose Slate, Allison Teevan & Barbara Juhasz. 2019. "The processing of blend words in naming and sentence reading". *Quarterly Journal of Experimental Psychology*. 72 (4), 847-57.

Jones, Charles. 1993. *Historical Linguistics: Problems and Perspectives*. New York: Routledge.

Juhasz, Barbara, Rebecca Johnson & Jennifer Brewer. 2017. "An investigation into the processing of lexicalized blend words: Evidence from lexical decisions and eye movements during reading". *Journal of Psycholinguistic Research*. 46, 281-94.

Juhasz, Barbara, Matthew Starr, Albrecht Inhoff & Lars Placke. 2003. "The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions, and eye fixations". *British Journal of Psychology*. 94, 223-44.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kastovsky, Dieter. 1982. *Wortbildung und Semantik (Studienreihe Englisch 14)*. Düsseldorf, Bern, München: Bagel/Francke.

Katamba, Francis. 1994. *English Words*. London, New York: Routledge.

Keller, Rudi. 1994. *On Language Change: The Invisible Hand in Language*. New York: Routledge.

Keller, Rudi. 1995. *Zeichentheorie*. Tübingen: Francke.

Keller, Rudi. 2014. *Sprachwandel. Von der unsichtbaren Hand in der Sprache*. 4th. Tübingen: Francke.

Kelly, Michael. 1998. "To "brunch" or "brench": Some aspects of blend structure". *Linguistics*. 36, 579-90.

Kemmer, Suzanne. 2003. "Schemas and lexical blends". In: Hubert Cuyckens, Thomas Berg, René Dirven & Klaus-Uwe Panther (eds.). *Motivation in Language: Studies in honor of Günter Radden*. 69-97.

- Kerremans, Daphne. 2015. *A Web of New Words: A Corpus-Based Study of the Conventionalization of Process of English Neologisms*. Frankfurt am Main: Peter Lang Verlag.
- Kerremans, Daphné, Susanne Stegmayr & Hans-Jörg Schmid. 2011. "The NeoCrawler: Identifying and Retrieving Neologisms from the Internet and Monitoring Ongoing Change". In: Kathryn Allan & Justyna A. Robinson (eds.). *Current Methods in Historical Semantics*. Berlin: de Gruyter. 59-96.
- Kerremans, Daphné, Jelena Prokić, Quirin Würschinger & Hans-Jörg Schmid. 2019. "Using data-mining to identify and study patterns in lexical innovation on the web: The *NeoCrawler*". *Pragmatics & Cognition* 25 (1), 174-200
- Kerswill, Paul & Ann Williams. 1999. "Mobility versus Social Class in Dialect Levelling: Evidence from New and Old Towns in England". *Cuadernos de Filología Inglesa*. 8, 47-57.
- Ketabi, Saeed, Mir Emad Aghili & M. R. Abou Alhasani. 2010. "Gender and its Relation to Acceptability and transparency of neologisms: A Case Study of General Word Approved by Academy of Persian Language and Literature". *Women in Culture and Art*. 1 (3), 99-108.
- Kiefer, Markus & Friedemann Pulvermüller. 2012. "Conceptual Representations in Mind and Brain: Theoretical Developments, Current Evidence and Future Directions". *Cortex*. 48 (7), 805-25.
- Kiesling, Scott F. 2011. *Linguistic Variation and Change*. Edinburgh: Edinburgh University Press.
- Kilgarriff, Adam. "Linguistic Search Engine". In: Proceedings of the Shallow Processing of Large Corpora Workshop (SProLaC 2003) Corpus Linguistics 2003. Lancaster University.
- Kinne, Michael. 1996. "Neologismus und Neologismenlexikographie im Deutschen. Zur Forschungsgeschichte und zur Terminologie, über Vorbilder und Aufgaben". *Deutsche Sprache*. 24, 327-58.
- Kjellmer, Göran. 2000. "Potential Words". *Word*. 51 (2), 205-28.
- Kroczyk, Leon & Thomas C. Gunter. 2017. "Communicative Predictions Can Overrule Linguistic Priors". *Scientific Reports*. 7 (1), 175-81.

- Krott, Andrea. 2009. "The role of analogy for compound words". In: James P. Blevins & Juliette Blevins (eds.). *Analogy in Grammar: Form and Acquisition*. Oxford: Oxford University Press. 118-136.
- Krott, Andrea, Robert Schreuder, R. Harald Baayen & Wolfgang U. Dressler. 2007. "Analogical effects on linking elements in German compound words". *Language and cognitive processes*. 22 (1), 25-57.
- Kumar, Satyendra. 2020. "Corona Virus Outbreak: Keep Physical Distancing, Not Social Distancing". *Social Science Research Network*.
- Kuperman, Victor, Robert Schreuder, Raymond Bertram & Harald Baayen. 2009. "Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing". *Journal of Experimental Psychology: Human Perception and Performance*. 35, 876-95.
- Labov, William. 1964. "Phonological Correlates of Social Stratification". *American Anthropologist, New Series*. 33 (3), 164-76.
- Labov, William. 1966. *The social stratification of English in New York City*. Washington DC: Center for Applied Linguistics.
- Labov, William. 1972. *Language in the Inner City: Studies in Black English Vernacular*. Philadelphia: University of Philadelphia Press.
- Labov, William. 1980. "The social origins of sound change". In: William Labov (ed.). *Location Language in Time and Space*. New York: Academic Press. 251-266.
- Labov, William. 1990. "The intersection of sex and social class in the course of linguistic change". *Language Variation and Change*. 2, 205-54.
- Labov, William. 1994. *Principles of Linguistic Change. Vol I. Internal Factors*. Oxford: Basil Blackwell.
- Labov, William. 2001. *Principles of Linguistic Change. Vol II. Social Factors*. Oxford: Blackwell.
- Labov, William. 2003. "Pursuing the Cascade Model". In: David Britain & Jenny Cheshire (eds.). *Social Dialectology: In Honour of Peter Trudgill*. Amsterdam, Philadelphia: Benjamins. 9-22.

- Lalić -Krstin, Gordana & Nadežda Silaški. 2018. "From Brexit to Bregret". *English Today*. 34 (2), 3-8.
- Langacker, Ronald. 2000. *Grammar and Conceptualization*. Berlin and New York: de Gruyter.
- Langacker, Ronald. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.
- Langacker, Ronald. 2017. "Entrenchment in Cognitive Grammar". In: Hans-Jörg Schmid (ed.). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Berlin, New York: de Gruyter. 39-56.
- Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar. Vol. 1: Theoretical prerequisites*. Stanford: Stanford University Press.
- Langlotz, Andreas. 2006. *Idiomatic Creativity*. Amsterdam: Benjamins.
- Lave, Jean & Étienne Wenger. 1991. *Situated Learning: Legitimate Peripheral Participation*. Cambridge: Cambridge University Press.
- Lawson, Robert. *Coronavirus has led to an explosion of new words and phrases – and that helps us cope*. <https://theconversation.com/coronavirus-has-led-to-an-explosion-of-new-words-and-phrases-and-that-helps-us-cope-136909>. 2020.
- Le Page, Robert. 1954. *Linguistic Survey of the British Caribbean: Questionnaire A*. Jamaica: University College of the West Indies.
- Le Page, Robert & Andrée Tabouret-Keller. 1985. *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: Cambridge University Press.
- Ledgeway, Adam & Martin Maiden. 2016. *The Oxford Guide to the Romance Languages*. Oxford: Oxford University Press.
- Leech, Geoffrey. 1974. *Semantics. The study of Meaning*. Harmondsworth: Penguin.
- Leech, Geoffrey. 1981. *Semantics*. 2nd. Harmondsworth: Penguin.

- Lees, Robert. 1960; 1968. *The Grammar of English Nominalizations*. 5th. Bloomington, The Hague: Indiana University Press, Mouton.
- Lehmann, Christian. 2002. "New reflections on grammaticalization and lexicalization". In: Ilse Wischer & Gabriele Diewald (eds.). *New Reflections on Grammaticalization*. Amsterdam, Philadelphia: Benjamins. 1-18.
- Lehrer, Adrienne. 1996. "Identifying and interpreting blends: an experimental approach". *Cognitive Linguistics*. 7, 359-90.
- Lehrer, Adrienne. 1998. "Scapes, Holics, and Thons: The Semantics of English Combining Forms". *American Speech*. 73 (1), 3-28.
- Lehrer, Adrienne. 2003. "Understanding trendy neologisms". *Rivista di Linguistica*. 15, 371-84.
- Lehrer, Adrienne. 2006. "Neologisms". In: Keith Brown (ed.). *Encyclopedia of Language and Linguistics*. 2nd Oxford: Elsevier. 590-593.
- Lehrer, Adrienne. 2007. "Blendalicious". In: Judith Munat (ed.). *Lexical Creativity, Texts, and Contexts*. Amsterdam, Philadelphia: Benjamins. 115-133.
- Leisi, Ernst. 1955; 1985. *Das heutige Englisch*. Heidelberg: Winter.
- Lemnitzer, Lothar. 2011. "Making sense of nonce words". In: Margrethe Heidemann Andersen & Jörgen N. Jensen (eds.). *Sprognaevets Konferenceseerie 1*. Nye Ord. Copenhagen: 7-18.
- Levelt, Willem, Ardi Roelofs & Antje S. Meyer. 1999. "A Theory of Lexical Access in Speech Production". *The Behavioral and Brain Sciences*. 22 (1), 1-38.
- Lewis, David. 1969. *Convention*. Cambridge: Cambridge University Press.
- Lichtenberger, Elizabeth O. & Alan S. Kaufman. 2012. *Essentials of WAIS-IV Assessment*. 2nd. New Jersey: John Wiley & Sons.

- Lieber, Rochelle & Pavol Štekauer. 2009. *The Oxford Handbook of Compounding*. Oxford: Oxford University Press.
- Lipka, Leonhard. 1972. *Semantic Structure and Word-formation. Verb-particle Constructions in Contemporary English*. München: Fink.
- Lipka, Leonhard. 1975. "Re-Discovery Procedures and the Lexicon". *Lingua*. 37 (2-3), 197-224.
- Lipka, Leonhard. 1977. "Lexikalisierung, Idiomatisierung und Hypostasierung als Probleme einer synchronischen Wortbildungslehre". In: Herbert Brekle & Dieter Kastovsky (eds.). *Perspektiven der Wortbildungsforschung. Beiträge zum Wuppertaler Wortbildungskolloquium vom 9.-10. Juli 1976 anlässlich des 70. Geburtstags von Hans Marchand am 1. Oktober 1977*. Bonn: Bouvier Verlag. 155-164.
- Lipka, Leonhard. 1981. "Zur Lexikalisierung im Deutschen und Englischen". In: Leonhard Lipka & Hartmut Günther (eds.). *Wortbildung*. Darmstadt: (Wege der Forschung 564). 119-132.
- Lipka, Leonhard. 1992. *An Outline of English Lexicology*. 2nd. Tübingen: Niemeyer Verlag.
- Lipka, Leonhard. 2002. *English Lexicology. Lexical structure, word semantics, and word-formation*. Tübingen: Narr.
- Lipka, Leonhard. 2010. "Observational linguistics, neologisms, entrenchment, and the Tea Party Movement". *Brno Studies in English*. 36 (1), 95-101.
- Lipka, Leonhard, Susanne Handl & Wolfgang Falkner. 2004. "Lexicalization and Institutionalization. The State of the Art in 2004". *SKASE Journal of Theoretical Linguistics*. 1, 2-19.
- Little, William, Charles Onions, George W. S. Friedrichsen, Henry W. Fowler & Jessie Coulson. 1973. *Shorter Oxford English Dictionary*. 3rd. Oxford: Oxford University Press.
- Lohde, Michael. 2006. *Wortbildung des modernen Deutschen: ein Lehr- und Übungsbuch*. Tübingen: Narr Verlag.

- López Fernández, Susana. 2005. *Novas palabras galegas. Repertorio de creacións léxicas rexistradas na prensa e Internet*. Vigo: Servizo de Publicacións da Universidade de Vigo.
- López Rúa, Paula. 2004. "The categorical continuum of English blends". *English Studies*. 85, 63-76.
- Lüdeling, Anke, Stefan Evert & Marco Baroni. 2007. "Using Web Data for Linguistic Purposes". *Corpus Linguistics and the Web*. 59, 7-24.
- Lutz, Angelika. 2013. "Language Contact and Prestige". *Anglia*. 131 (4), 562-90.
- Lyons, John. 1977. *Semantics*. Cambridge: Cambridge University Press.
- Lyons, John. 1982. "Deixis and Subjectivity: Loquor, ergo Sum?". In: Robert J. Jarvella & Wolfgang Klein (eds.). *Speech, Place, and Action: Studies in Deixis and Related Topics*. New York: Wiley. 101-124.
- MacKinnon, David Peter. 2008. *Introduction to Statistical Mediation Analysis*. New York: Erlbaum.
- Mahdawi, Arwa. *From Covidiot to doomscrolling: how coronavirus is changing our language*. <https://www.theguardian.com/commentisfree/2020/apr/15/from-covidiot-to-doomscrolling-how-coronavirus-is-changing-our-language>. 2020.
- Mair, Christian. 2006. *Twentieth-Century English: History, Variation and Standardization*. Cambridge: Cambridge University Press.
- Marchand, Hans. 1960. *The Categories and Types of Present-Day English Word-Formation*. Wiesbaden: Otto Harrassowitz.
- Marchand, Hans. 1969. *The Categories and Types of Present-Day English Word-Formation*. 2nd revised. München: Beck.
- Marslen-Wilson, William, Mike Ford, Lianne Older & Zhou Xiaolin. 1996. "The Combinatorial Lexicon: Priming Derivational Affixes". In: Garrison W. Cottrell (ed.). *Proceedings of the 18th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum. 223-227.

- Maybaum, Rebecca. 2013. "Language Change as a Social Process: Diffusion Patterns of Lexical Innovations in Twitter". *Berkeley Linguistics Society*. 152-66.
- McDonald, Scott & Richard Shillcock. 2001. "Rethinking the Word Frequency Effect: The Neglected Role of Distributional Information in Lexical Processing". *Language and Speech*. 44 (3), 295-323.
- McIntosh, Angus. 1952. *An Introduction to a Survey of Scottish Dialects*. Edinburgh: Nelson.
- Metcalf, Allan. 2002. *Predicting New Words*. Boston: HMC0.
- Michael, Lev. 2014. "Social Dimensions of Language Change". In: Claire Bower & Bethwyn Evans (eds.). *The Routledge Handbook of Historical Linguistics*. London: Routledge. 484-502.
- Milroy, James. 1992. *Linguistic Variation and Change*. Oxford: Oxford University Press.
- Milroy, James & Matthew Gordon. 2003. *Sociolinguistics. Method and Interpretation*. Malden; Oxford: Blackwell Publishing.
- Milroy, James & Lesley Milroy. 1985. "Linguistic Change, Social Network and Speaker Innovation". *Journal of Linguistics*. 21 (2), 339-84.
- Milroy, Lesley & James Milroy. 1992. "Social Network and Social Class: Toward an Integrated Sociolinguistic Model". *Language in Society*. 21 (1), 1-26.
- Moreno Cabrera, Juan C. 1998. "On the relationships between grammaticalization and lexicalization". In: Anna G. Ramat & Paul J. Hopper (eds.). *The Limits of Grammaticalization*. Amsterdam: Benjamins.
- Morton, John. 1969. "Interaction of Information in Word Recognition". *Psychological Review*. 76 (2), 165-78.
- Moseley, Tom. *The rise of the word Brexit*. <https://www.bbc.co.uk/news/uk-politics-37896977>. 17.04.2020.
- Motsch, Wolfgang. 2004. *Deutsche Wortbildung in Grundzügen*. Berlin, New York: de Gruyter.

- Murray, Ian. *Bring on the 'Brixit': EU withdrawal would bring benefits for both Britain and the US*. <https://www.dailymail.co.uk/debate/article-2163244/Bring-Brixit-EU-withdrawal-bring-benefits-Britain-US.html>. 17.04.2020.
- Murray, Thomas, Timothy Frazer & Beth Lee Simon. 1996. "Need + Past Participle in American English". *American Speech*. 71, 255-71.
- Murray, Thomas & Beth Lee Simon. 1999. "Want + past participle in American English". *American Speech*. 74, 140-64.
- Nagy, William, Richard Anderson, Marlene Schommer, Judith Ann Scott & Anne Stallman. 1989. "Morphological Families in the Internal Lexicon". *Reading Research Quarterly*. 24 (3), 262-82.
- Narrog, Heiko & Bernd Heine. 2011. *The Oxford Handbook of Grammaticalization*. Oxford: Oxford University Press.
- Neely, James H. 1977. "Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention". *Journal of Experimental Psychology: General*. 106 (3), 226-54.
- Nübling, Damaris & Renata Szczepaniak. 2011. "Merkmal(s?)analyse, Seminar(s?)arbeit und Essen(s?)ausgabe: Zweifelsfälle der Verfungung als Indikatoren für Sprachwandel". *Zeitschrift für Sprachwissenschaft*. 30, 45-73.
- Observatori de Neologia. 1998. *Diccionari de paraules noves. Neologismes recollits a la premsa*. Barcelona: Enciclopèdia Catalana.
- Ochs, Elinor. 1992. "Indexing Gender". In: Alessandro Duranti & Charles Goodwin (eds.). *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge: Cambridge University Press. 335-358.
- Oltra-Massuet, Isabel. 2017. "Towards a morphosyntactic analysis of -ish". *Word Structure*. 10 (1), 54-78.
- Oppenheim, Abraham N. 1992. *Questionnaire Design, Interviewing and Attitude Measurement*. London: Continuum.

- Ortiz-Ospina, Esteban. *The rise of social media*. <https://ourworldindata.org/rise-of-social-media>.
04/06.2020.
- Ortner, Lorelies & Elgin Müller-Bollhagen. 1991. *Deutsche Wortbildung. Typen und Tendenzen in der Gegenwart. Vierter Hauptteil: Substantivkomposita*. Berlin, New York: de Gruyter.
- Paap, Kenneth, Sandra Newsome, James McDonald & Roger Schvaneveldt. 1982. "An Activation-Verification Model for Letter and Word Recognition: The Word-Superiority Effect". *Psychological Review*. 89 (5), 573-94.
- Paolillo, John C. 2001. "Language Variation on Internet Relay Chat: A Social Network Approach". *Journal of Sociolinguistics*. 5 (2), 180-213.
- Paradowski, Michal & Lukasz Jonak. 2012. "Diffusion of linguistic innovation as social coordination". *Psychology of Language and Communication*. 16, 53-64.
- Paul, Hermann. 1920. *Prinzipien der Sprachgeschichte*. 5th. Halle: Niemeyer.
- Petré, Peter. 2017. "The extravagant progressive: An experimental corpus study on the history of emphatic [be Ving]". *English Language and Linguistics*. 21 (02), 227-50.
- Pickering, Martin & Victor Ferreira. 2008. "Structural Priming: A Critical Review". *Psychological Bulletin*. 134 (3), 427-59.
- Pickering, Martin & Simon Garrod. 2004. "Toward a Mechanistic Psychology of Dialogue". *Behavioral and Brain Sciences*. 27 (2), 169-90.
- Pidd, Helen. *Mind your language: German linguists oppose influx of English words*.
<https://www.theguardian.com/world/2011/mar/14/german-language-anglicisms-challenge>.
15.06.2020.
- Pinker, Steven. 1998. "Words and Rules". *Lingua*. 106 (1), 219-42.
- Pinker, Steven & Alan Prince. 1988. "On Language and Connectionism: Analysis of a Parallel Distributed Processing Model of Language Acquisition". *Cognition*. 28 (1-2), 73-193.

- Pinker, Steven & Michael T. Ullman. 2002. "The past and future of the past tense". *Trends in cognitive science*. 6 (11), 456-63.
- Plag, Ingo. 2003. *Word Formation in English*. Cambridge: Cambridge University Press.
- Pollatsek, Alexander, Jukka Hyönä & Raymond Bertram. 2000. "The role of morphological constituents in reading Finnish compound words". *Journal of Experimental Psychology: Human Perception and Performance*. 26, 820-33.
- Polyudova, Elena. 2014. *Acquiring Lingua Franca of the Modern Time: Current Issues and Strategies in ESL Studies*. Cambridge: Cambridge Scholars Publishing.
- Prasada, Sandeep & Stephen Pinker. 1993. "Generalisation of Regular and Irregular Morphological Patterns". *Language and Cognitive Processes*. 8 (1), 1-56.
- Pyles, Thomas. 1952. *Words and Ways of American English*. New York: Random House.
- Quirk, Randolph & Sidney Greenbaum. 1973. *A University Grammar of English*. London: Longman.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985, 2006. *A Comprehensive Grammar of the English Language*. London, New York: Longman.
- Ramat, Paolo. 1992. "Thoughts on degrammaticalization". *Linguistics*. 30, 549-60.
- Ramscar, Michael & Robert Port. 2015. "Categorization (without Categories)". In: Ewa Dąbrowska & Dagmar Divjak (eds.). *Handbook of Cognitive Linguistics*. Berlin and Boston, MA: de Gruyter. 75-99.
- Rauss, Karsten & Gilles Pourtois. 2013. "What Is Bottom-Up and What Is Top-Down in Predictive Coding?". *Frontiers in Psychology*. 4 (276).
- Rayner, Keith & Susan Duffy. 1986. "Lexical Complexity and Fixation Times in Reading: Effects of Word Frequency, Verb Complexity, and Lexical Ambiguity". *Memory & Cognition*. 14 (3), 191-201.
- RDUES (Research and Development Unit for English Studies, Birmingham City University). *Webcorp*. <http://webcorp.org.uk/live/>. 24.04.2020.

- Reindl, Eva & Maria Thurmair. 1992. "Der Fürstbischof im Hosenrock. Eine Studie zu den nominalen Kopulativkomposita des Deutschen". *Deutsche Sprache*. 20, 32-61.
- Renouf, Antoinette, Andrew Kehoe and Jay Banerjee. 2005. "The WebCorp Search Engine A holistic approach to web text search ". In: Electronic Proceedings of CL2005. University of Birmingham.
- Rescorla, Robert A. & Allan R. Wagner. 1972. "A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement". In: Abraham H. Black & William F. Prokasy (eds.). *Classical Conditioning II: Current Research and Theory*. New York: Appleton-Century-Crofts. 64-99.
- Rey, Alain. 1976. "Le néologisme: un pseudo-concept?". *Cahiers de Lexicologie*. 28, 3-17.
- Rickford, John R. 1993. "Concord and Conflict in the Speech Community. Revised Version of "Concord and Conflict in the Character of the Speech Community", 1986". *Sheffield Working Papers in Language and Linguistics* 3. <http://www.johnrickford.com/portals/45/documents/papers/Rickford-1986f-1993-Concord-and-Conflict-in-the-Speech-Community.pdf>, accessed 24.09.2020: 87-119.
- Rogers, Everett M. 1995, 2003. *Diffusion of Innovations*. New York, London: Free Press.
- Ronneberger-Sibold, Elke. 2008. "Word Creation: Definition – Function – Typology". In: Franz Rainer, Wolfgang Dressler, Dieter Kastovsky, Hans Luschützky (eds.): *Variation and change in morphology: selected papers from the 13th International Morphology Meeting, Vienna, February 2008*. Amsterdam: Benjamins, 2010. - S. 201-216.
- Rumelhart, David & James McClelland. 1986. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. Cambridge, MA: MIT Press.
- Ryder, Mary Ellen. 1994. *Ordered Chaos: the interpretation of English noun–noun compounds*. Berkeley, CA: University of California Press.
- Ryder, Mary Ellen. 1999. "Bankers and blue-chippers: An account of -er formations in Present-day English". *English Language and Linguistics*. 3, 269-97.

- Salis, Amanda. *The science of 'hangry': Why some people get grumpy when they are hungry*.
<https://edition.cnn.com/2015/07/20/health/science-behind-being-hangry/index.html>. 17.04.2020.
- Sandra, Dominiek. 1990. "On the representation and processing of compound words: Automatic access to constituent morphemes does not occur". *Quarterly Journal of Experimental Psychology A*. 42, 529-67.
- Sandra, Dominiek. 1994. *Morphology in the Reader's Mental Lexicon*. Frankfurt am Main: Lang.
- Sankoff, Gillian & H el ene Blondeau. 2007. "Language Change across the Lifespan: /r/ in Montreal French". 83 (3), 560-88.
- Sauer-Egner, Anja & Birgit Reker. 2007. *St orungen der Verarbeitung von Nomina-Komposita Sprachliche Schnittstellen*. Hofheim: Natverlag.
- Scalise, Sergio & Emiliano Guevara. 2005. "The lexicalist approach to word-formation". In: Pavol  stekauer & Rochelle Lieber (eds.). *Handbook of Word-Formation*. Dordrecht: Springer. 147-187.
- Schiffer, Stephen. 1972. *Meaning*. Oxford: Clarendon Press.
- Schl ucker, Barbara. "Die deutsche Kompositionsfreudigkeit  bersicht und Einf uhrung". In: Livio Gaeta & Barabara Schl ucker (eds.). *Das Deutsche als kompositionsfreudige Sprache. Strukturelle Eigenschaften und systembezogene Aspekte*. Berlin, New York: de Gruyter. 1-25.
- Schmid, Hans-J org. 2007. "Entrenchment, Saliency and Basic Levels". In: Dirk Geeraerts & Hubert Cuyckens (eds.). *The Oxford Handbook of Cognitive Linguistics*. Oxford: Oxford University Press. 117-137.
- Schmid, Hans-J org. 2008. "New words in the mind: concept-formation and entrenchment in neologisms". *Anglia. Zeitschrift f ur Englische Philologie*. 126 (1), 1-36.
- Schmid, Hans-J org. 2010. "Does frequency in text really instantiate entrenchment in the cognitive system?". In: Dylan Glynn & Kerstin Fischer (eds.). *Quantitative methods in cognitive semantics: Corpus-driven approaches*. Berlin, New York: de Gruyter. 101-133.

- Schmid, Hans-Jörg. 2011. *English Morphology and Word-Formation*. 2nd revised and translated. Berlin: Erich Schmidt Verlag.
- Schmid, Hans-Jörg. 2014. "Lexico-grammatical patterns, pragmatic associations and discourse frequency". In: Thomas Herbst, Hans-Jörg Schmid & Susen Faulhaber (eds.). *Constructions - collocations - patterns*. Berlin, New York: de Gruyter. 239-293.
- Schmid, Hans-Jörg. 2020. *The Dynamics of the Linguistic System: Usage, Conventionalization, and Entrenchment*. Oxford: Oxford University Press.
- Schmid, Hans-Jörg & Franziska Günther. 2016. "Toward a Unified Socio-Cognitive Framework for Salience in Language". *Frontiers in Psychology*. 7 (1110).
- Schmidt, Johannes. 1872. *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Hermann Böhlau.
- Schneider, Edgar W. 2011. *English around the World: An Introduction*. Cambridge: Cambridge University Press.
- Selting, Margret. 1983. "Institutionelle Kommunikation: Stilwechsel als Mittel strategischer Interaktion". *Linguistische Berichte*. 86. 29-48.
- Seyboth, Margret. 2014. *Ein Ganzes oder zwei Teile? Untersuchungen zur Verarbeitung von Komposita durch Sprachgesunde, Aphasiker und Synästhetiker*. Digitale Bibliothek Thüringen 2014.
- Simonini, Ross. 1966. "Word-making in present-day English". *The English Journal*. 55 (6), 752-7.
- Soanes, Catherine & Angus Stevenson. 2008. *Concise Oxford English Dictionary*. 11th. Oxford: Oxford University Press.
- Sólyom, Réka. 2014. "Semantic Features of Hungarian Neologisms with the Prefix Be: Analysis of Some Frequently Used Verbs (e.g., bevállal)". *Hungarian Cultural Studies. e-Journal of the American Hungarian Educators Association*. 7.

- Stefanowitsch, Anatol. 2008. "Negative Entrenchment: A Usage-Based Approach to Negative Evidence". *Cognitive Linguistics*. 19 (3), 513-31.
- Stefanowitsch, Anatol & Susanne Flach. 2017. "The Corpus-Based Perspective on Entrenchment". In: Hans-Jörg Schmid (ed.). *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge*. Berlin, New York: de Gruyter. 101-127.
- Steinkrauss, Rasmus & Monika Schmid. 2017. "Entrenchment and Language Attrition". In: Hans-Jörg Schmid (ed.). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Boston, MA, and Berlin: APA and de Gruyter. 367-383.
- Štekauer, Pavol. 2002. "On the Theory of Neologisms and Nonce-formations". *Australian Journal of Linguistics*. 22 (1), 97-112.
- Strang, Barbara. 1970. *A History of English*. London: Methuen.
- Svanlund, Jan. 2018. "Metalinguistic Comments and Signals: What Can They Tell Us About the Conventionalization of Neologies?". *Pragmatics & Cognition*. 25 (1), 121-40.
- Szmrecsanyi, Benedikt. 2005. "Language Users as Creatures of Habit: A Corpus-Based Analysis of Persistence in Spoken English". *Corpus Linguistics and Linguistic Theory*. 1 (1), 113-50.
- Szmrecsanyi, Benedikt. 2006. *Morphosyntactic Persistence in Spoken English. A Corpus Study at the Intersection of Variationist Sociolinguistics, Psycholinguistics, and Discourse Analysis*. Berlin and Boston, MA: de Gruyter.
- Taft, Marcus & Kenneth Forster. 1975. "Lexical storage and retrieval of prefixed words". *Journal of Verbal Learning & Verbal Behavior*. 14, 683-47.
- Takashima, Atsuko & Iske Bakker. 2017. "Memory Consolidation". In: Hans-Jörg Schmid (ed.). *Entrenchment and the Psychology of Language Learning: How We Reorganize and Adapt Linguistic Knowledge*. Boston, MA, and Berlin: APA and de Gruyter. 177-200.

- Tamminen, Jakke & Gareth Gaskell. 2013. "Novel Word Integration in the Mental Lexicon: Evidence from Unmasked and Masked Semantic Priming". *The Quarterly Journal of Experimental Psychology*. 66 (5), 1001-25.
- Tangermann, Victor. *It's Officially Time to Stop Using The Phrase 'Social Distancing'*. <https://www.sciencealert.com/who-is-no-longer-using-the-phrase-social-distancing>. 2020.
- Tomasello, Michael. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.
- Tottie, Gunnel. 1991. "Lexical diffusion in syntactic change: Frequency as a determinant of linguistic conservatism in the development of negation in English". In: Dieter Kastovsky (ed.). *Historical English syntax*. Berlin: de Gruyter .439-467.
- Tournier, Jean. 1985. *Introduction descriptive à la lexicogénétique de l'anglais contemporain*. Paris: Champion-Slatkine.
- Traugott, Elizabeth. 2010. "(Inter)Subjectivity and (Inter)Subjectification: A Reassessment". In: Kristin Davidse, Lieven Vandelanotte & Hubert Cuyckens (eds.). *Subjectification, Intersubjectification and Grammaticalization*. Berlin and Boston, MA: de Gruyter. 29-74.
- Traugott, Elizabeth & Richard Dasher. 2002. *Regularity in Semantic Change*. Cambridge: Cambridge University Press.
- Traugott, Elizabeth & Graeme Trousdale. 2013. *Constructionalization and Constructional Changes*. Oxford: Oxford University Press.
- Trudgill, Peter. 1972. "Sex, Covert Prestige and Linguistic Change in the Urban British English of Norwich". *Language in Society*. 1, 179-95.
- Trudgill, Peter. 1974a. "Linguistic Change and Diffusion: Description and Explanation in Sociolinguistic Dialect Geography". *Language in Society*. 3 (2), 215-46.

- Trudgill, Peter. 1974b. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.
- Trudgill, Peter. 1986. *Dialects in contact*. Oxford, New York: Basil Blackwell.
- Trudgill, Peter. 2008. "Colonial dialect contact in the history of European languages: On the irrelevance of identity in new-dialect formation". *Language in Society*. 37, 241-80.
- Tulloch, Sara. 1993. *The Oxford Dictionary of New Words: A Popular Guide to Words in the News (Oxford Quick Reference)*. Oxford: Oxford University Press.
- Ullmann, Stephen. 1962. *Semantics. An Introduction to the Science of Meaning*. Oxford: Basil Blackwell.
- Ungerer, Friedrich & Hans-Jörg Schmid. 1998. "Englische Komposita und Kategorisierung". *Rostocker Beiträge zur Sprachwissenschaft*. 5, 77-98.
- Van der Ven, Frauke, Atsuko Takashima, Eliane Segers & Ludo Verhoeven. 2015. "Learning Word Meanings: Overnight Integration and Study Modality Affects". *PLOS ONE*. 10 (5), e0124926.
- Weber, Andrea & Odette Scharenborg. 2012. "Models of Spoken-Word Recognition". *Wiley Interdisciplinary Reviews: Cognitive Science*. 3 (3), 387-401.
- Weinreich, Uriel, William Labov & Marvin Herzog. 1968. "Empirical Foundations for a Theory of Language Change". In: Winfred Lehmann & Yakov Malkiel (eds.). *Directions for Historical Linguistics. A Symposium*. Austin, TX: University of Texas Press. 95-195.
- Williams, Edwin. 1981. "On the notions 'lexically related' and the 'head of the word'". *Linguistic Inquiry*. 12, 245-74.
- Wooldridge, Adrian. *A Brexit looms*. <https://www.economist.com/bagehots-notebook/2012/06/21/a-brixit-looms>. 17.04.2020.
- Würschinger, Quirin. "Social network analysis of lexical diffusion. Investigating the spread of new words on Twitter". *ICAME 2019, Neuchâtel Workshop 'Corpus approaches to social media'*.

- Würschinger, Quirin, Mohammad Fazleh Elahi, Desislava Zhekova and Hans-Jörg Schmid. 2016. "Using the Web and Social Media as Corpora for Monitoring the Spread of Neologisms. The case of *rapefugee*, *rapeugee*, and *rapugee*". In: Proceedings of the 10th Web as Corpus Workshop (WAC-X) and the EmpiriST Shared Task, August 7-12, 2016, Association for Computational Linguistics. Berlin. 35-43.
- Ziem, Alexander & Alexander Lasch. 2013. *Konstruktionsgrammatik: Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin: de Gruyter.
- Zimmer, Karl E. 1981. "Some general observations about nominal compounds". In: Leonhard Lipka & Hartmut Günther (eds.). *Wortbildung*. Darmstadt: WBG (Wissenschaftliche Buchgesellschaft). 223-257.
- Zipf, George. 1949. *Human Behavior and the Principle of Least Effort. An Introduction to Human Ecology*. Cambridge, MA: Addison-Wesley Press.
- Zwitzerlood, Pienie. 1989. "The Locus of the Effects of Sentential-Semantic Context in Spoken-Word Processing". *Cognition*. 32 (1), 25-64.
- Zwitzerlood, Pienie. 1993. "The role of semantic transparency in the processing and representation of Dutch compounds". *Language and Cognitive Processes*. 9 (3), 341-368