

Roman Schneider/Christian Lang/Sandra Hansen (Mannheim)
**Das Vokabular von Songtexten im
gesellschaftlichen Kontext –
ein diachron-empirischer Beitrag**

Abstract: Der Beitrag untersucht den Stellenwert gesellschaftlich relevanter Thematiken in deutschsprachigen Songtexten der zurückliegenden fünf Jahrzehnte. Dabei zeigt sich, dass neben individuellen Befindlichkeiten auch politische, sozialkritische oder umweltbezogene Themen signifikant angesprochen werden. Wir kontrastieren Songtexte mit anderen Testsorten und wenden dabei quantitative Methoden auf umfangreiche, breit stratifizierte Datensamples an, um die Phänomenbeschreibungen präzisierbar, generalisierbar und reproduzierbar zu machen. Das longitudinale Korpusdesign bietet Potenzial für diachrone Vergleiche. Im Sinne eines erweiterten „Mixed Methods“-Ansatzes exploriert die Studie zudem ausgewählte Aspekte qualitativ und bettet sie in den zeitlichen Kontext ein.

Das fachwissenschaftliche Potenzial deutschsprachiger Songtexte wird in der germanistischen Forschung gelegentlich kontrovers diskutiert (Achermann/Naschert 2005; von Ammon/von Petersdorff 2019). Obwohl Popmusik konstitutiver Bestandteil vieler Alltagssituationen (Storey 2010; Schneider 2019) und in Anglistik oder Amerikanistik bereits etablierter Forschungsgegenstand ist (vgl. exemplarisch die multidimensionale diachrone Registeranalyse in Werner i. Ersch.), finden sich in der sprachwissenschaftlichen Germanistik bislang erst vereinzelte Untersuchungen, etwa zur Phraseologismendichte (Lenk 2011) oder zu Hiphop als soziolinguistisch informativer Ausdrucksform mit Bezügen zu Jugend-/Migrantensprache (Androutsopoulos 2003; Ross et al. 2018). Politolinguistisch dominant sind Analysen vergleichsweise kleiner Datenmengen, etwa zum Vokabular populärer Subgenres (Psutka/Grassel 2018), zu verbaler Gewalt in Texten ausgewählter Alben (Bloching/Landschoff 2018) oder dem Protestcharakter deutschsprachiger Popsongs (Blühdorn 2003). Korpusbasierte Analysen wiederum ausschließlich englischsprachiger Popsongs erweitern das Spektrum um quantitative Aussagen zu semantischen Begriffsfeldern (Motschenbacher 2016), geschlechtsspezifischen Stereotypen (Kreyer 2015) oder hochfrequenten Inhaltsworten und N-Grammen (Brett/Pinna 2018; Nishina 2017).

Nachfolgend streben wir eine systematische thematische Erfassung des lexikalischen Inventars deutschsprachiger Songtexte seit Anfang der 1970er Jahre an, also seit dem Zeitpunkt, als medial erfolgreiche Veröffentlichungen (auch) poli-

tisch wurden. Diese Politisierung von Popsongs ist zwar ein immer wieder in der medialen Öffentlichkeit sowie der sozialwissenschaftlich orientierten Medienwissenschaft aufgegriffenes Thema (z. B. Kleiner 2016), aber auf breiterer empirischer Basis kaum belegt. Unsere kontrastiv zu anderen Textsorten zu validierende Hypothese lautet, dass Songtexte entgegen zuweilen bestehender Vorurteile nicht durchgehend triviale Themenkreise wie Konsum, Genuss oder Vergnügen behandeln, sondern signifikant auch politische, sozialkritische, umweltbezogene Themen.

Das der Studie zugrunde liegende Songtextkorpus (SK) verfolgt das langfristige Ziel einer Abdeckung wesentlicher Ausprägungen moderner Popmusiktexte. Es ist in thematische und autorenspezifische Archive unterteilt. Enthalten sind die kommerziell erfolgreichsten deutschsprachigen Songs der offiziellen west- bzw. gesamtdeutschen Singlecharts von 1970 bis 2020 sowie der DDR-Jahreshitparaden bis 1990. Damit wird den Kriterien Popularität und zeitliche Streuung Rechnung getragen sowie subjektiver Voreingenommenheit bei der Auswahl („personal bias“) entgegengetreten; die Bandbreite der Subgenres reicht von Schlager über Neue Deutsche Welle bis Hiphop. Weiterhin enthält das Korpus die Werke prominenter Liedermacher und Bands wie Udo Lindenberg, Konstantin Wecker, Stoppok, Ulla Meinecke, Hannes Wader und Element of Crime. Die derzeit ca. 3.500 Songs sind umfassend linguistisch annotiert und mit TEI-konformen Metadaten versehen (Schneider 2020).

Songtexte liegen zwar medial schriftlich vor, weisen aber charakteristische Merkmale sowohl konzeptioneller Schriftlichkeit als auch konzeptioneller Mündlichkeit auf. Unsere Vergleichskorpora repräsentieren deshalb ebenfalls graduell abgestufte Ausprägungen in diesem konzeptuellen Kontinuum und umfassen DEREKO-Archive (Kupietz et al. 2018) mit zudem mutmaßlich unterschiedlich ausgeprägtem gesellschaftlich-politischem Wortschatz: Parlamentsreden (PP), Agenturmeldungen (DPA) und Belletristik des 20./21. Jahrhunderts (LOZ). Als Referenz ziehen wir ein hinsichtlich Region, Register, Medium und Domäne stratifiziertes allgemeinsprachliches Korpus (AK) hinzu. Aus jeder dieser Sammlungen wurde eine zufällige Textstichprobe mit 12 Millionen lemmatisierten und Wortart-annotierten Wörtern extrahiert.

Zur thematischen Beurteilung des lexikalischen Inventars von Songtexten dienen Keywordanalysen (Bondi 2010), also die Feststellung von SK-Begriffsfrequenzen relativ zu den Vergleichskorpora. Als Kriterium erstellen wir eine Keywordliste sämtlicher im Untersuchungszeitraum von der Gesellschaft für deutsche Sprache (GfdS) publizierten „Wörter des Jahres“ (jeweils erster bis dritter Rang) sowie deren paradigmatischen Profile, berechnet unter Zuhilfenahme von DERE-KoVecs-Ähnlichkeitsmaßen (Fankhauser/Kupietz 2019). Um zusätzlich auch einschlägiges Institutionsvokabular – *Parlament, Regierung, Bundespräsident* usw. –, domänenspezifische Eigennamen und Fachvokabular abzubilden, nehmen wir per TextRank-Algorithmus (Mihalcea/Tarau 2004) ermittelte Top-Begriffe aller Artikelüberschriften des Nachrichtenmagazins „Der Spiegel“ seit 1970 in die Liste auf. Jedes Keyword wird einem oder mehreren der nachfolgenden Themenfelder zugeordnet (unabhängige manuelle Annotation mit Feststellung der Interrater-Reliabilität):

- **Bildung/Wissenschaft:** *Abitur, Ausbildung, Labor, Lehrer, Schule, Universität, ...*
- **Gesundheit/Drogen:** *Alkohol, Arznei, Dealer, Krankenhaus, Nikotin, Pandemie, Suizid, ...*
- **Gewalt/Krieg/Kriminalität:** *Atombombe, Auschwitz, Folter, Gangster, Geiselnahme, Mord, ...*
- **Liebe/Partnerschaft:** *Ehe, Erotik, Freier, Geliebte, Porno, Romanze, Tinder, Viagra, ...*
- **Kultur/Medien:** *Abendblatt, Berlinale, DSDS, Facebook, Fernsehen, Schlagzeile, ...*
- **Politik/Justiz/Religion:** *CDU, Fundamentalist, Merkel, Vatikan, Verfassungsgericht, ...*
- **Sport:** *Beckenbauer, Bundesliga, Nationalspieler, Stadion, Training, Turnschuh, Yoga, ...*
- **Umwelt/Naturschutz:** *Atomkraft, Endlager, Fukushima, GAU, Klimaschutz, Wackersdorf, ...*
- **Wirtschaft/Soziales:** *Arbeitsplatz, BASF, Dienstwagen, Euro, Fabrik, Privatisierung, ...*
- **Reisen/Geografie:** *Bahnhof, Bergsteiger, Lufthansa, Mallorca, Titicacasee, Urlaub, ...*

Die finale Liste umfasst ausschließlich Nomina und Eigennamen (NN/NE). Nicht aufgenommen werden mehrfach ambige und dadurch eingeschränkt spezifische Begriffe wie *Abstieg, Absturz, Abseits, Abwehr, Apparat, Anschluss* usw.

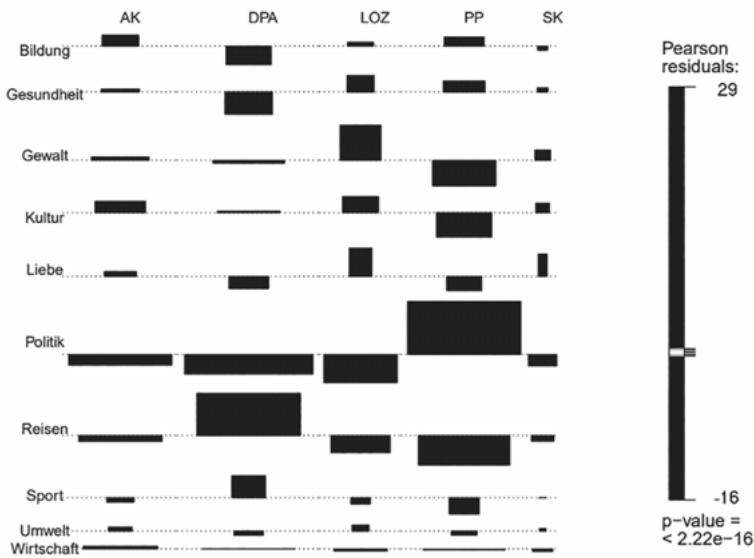


Abb. 1: Assoziationsplot für Themenfelder und Korpora

Abbildung 1 visualisiert standardisierte Abweichungen zwischen beobachteten und erwarteten Häufigkeiten (Pearson-Residuen) der themenspezifischen Keywords als Tokens oder Lemmata in den Korpus-Stichproben. Die Balkenbreite korrespondiert jeweils mit den erwarteten Häufigkeiten, die Balkenhöhen mit dem Grad der Abweichung nach oben bzw. unten. Es wird deutlich, dass im Songtextkorpus nicht nur der Themenbereich Liebe/Partnerschaft (erwartbar) im Fokus steht, sondern beispielsweise auch Gewalt/Krieg/Kriminalität, Gesundheit/Drogen oder Kultur/Medien. Insgesamt deutet diese erste Übersicht auf bemerkenswerte Ähnlichkeiten in der thematischen Gewichtung zwischen Songtextkorpus und belletristischer Textsammlung, aber auch dem alltagspraktischen Korpus hin.

Um systematisch zu überprüfen, inwieweit sich die thematische Bandbreite des lexikalischen Inventars in den verschiedenen Korpora unterscheidet, führen wir binär logistische Regressionsanalysen durch (vgl. Jaeger 2008) und ermitteln Wahrscheinlichkeiten für das Vorkommen der Themenfelder gemäß folgender Modellspezifikation¹:

$$P(\text{THEMA} = 1) = \text{logit}^{-1} (\beta_0 + \text{korpusdpa} * \beta_1 + \text{korpusloz} * \beta_2 + \text{korpuspp} * \beta_3 + \text{korpusk} * \beta_4)$$

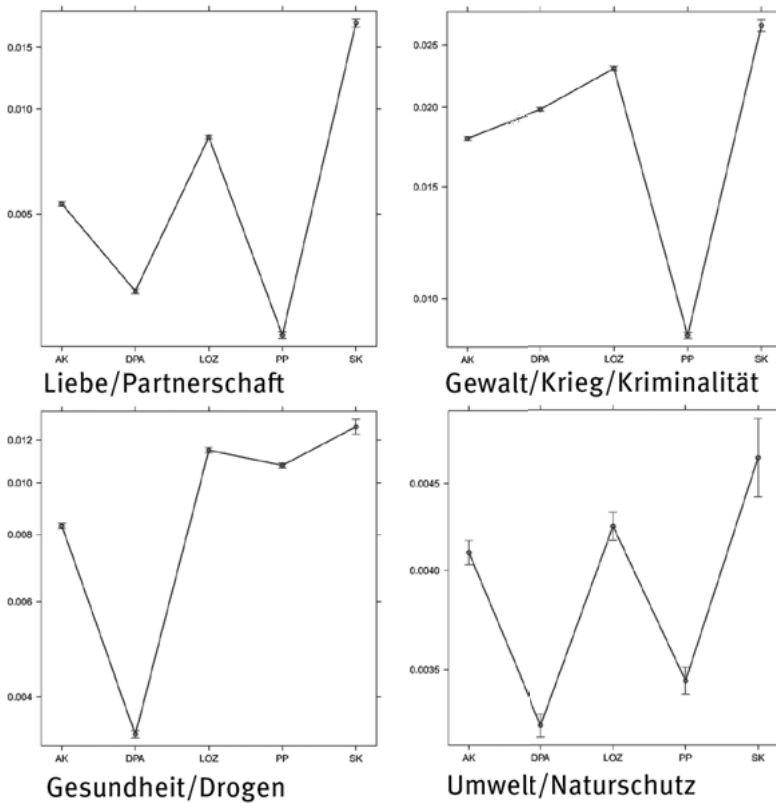


Abb. 2: Regressionsanalysen für ausgewählte Themenfelder

¹ THEMA steht hier als Platzhalter für die jeweiligen Themenfelder. Für jedes Themenfeld (= Kriteriumsvariable) wurde eine Regressionsanalyse berechnet.

Abbildung 2 visualisiert die geschätzten Werte ausgewählter statistischer Modelle als Effektplots; die Koeffizienten der Regressionsanalysen finden sich in Tabelle 1. Sowohl Liebe/Partnerschaft, Gewalt/Krieg/Kriminalität, Umwelt/Naturschutz als auch Gesundheit/Drogen weisen im SK signifikant höhere Wahrscheinlichkeiten als in den anderen Textsorten – inkl. dem AK – auf, wobei sich notabene der erstgenannte Themenbereich keineswegs auf „Herzschmerz“-Begriffe reduzieren lässt.

Tab. 1: Koeffizienten der Regressionsanalysen (Themenfeld-Bezeichnungen abgekürzt)

	z value Liebe	Pr(> z) Liebe	z value Gewalt	Pr(> z) Gewalt	z value Umwelt	Pr(> z) Umwelt	z value Gesund- heit	Pr(> z) Gesund- heit
(inter- cept)	-712.67	<2e-16 ***	- 992.06	<2e-16 ***	-655.98	<2e-16 ***	-811.70	<2e-16 ***
DPA	-51.70	<2e-16 ***	20.30	<2e-16 ***	-19.95	<2e-16 ***	-90.76	<2e-16 ***
LOZ	44.01	<2e-16 ***	44.23	<2e-16 ***	2.82	0.005 **	39.35	<2e-16 ***
PP	-64.08	<2e-16 ***	-102.20	<2e-16 ***	-13.83	<2e-16 ***	33.13	<2e-16 ***
SK	76.16	<2e-16 ***	34.44	<2e-16 ***	4.55	5.33e- 06 ***	24.56	<2e-16 ***

Ein genauerer Blick auf die Daten deutet unterschiedliche Ausschöpfungen der Keyword-Vokabulare an: Während etwa das Themengebiet Gesundheit/Drogen in Songtexten in der Mehrzahl der Fälle durch Aufputzmittel abgedeckt wird (*Bier*, *Kokain* etc.), sind in den anderen Korpora eher Begriffe aus dem Gesundheitsbetrieb dominant (*Medikamente*, *Krankenhaus*, *Pflege* etc.). Unregelmäßige Häufungen („Clumpiness“) von Keywords lassen sich anhand von Dispersionsmaßen nachweisen: Selbst ein vergleichsweise frequenter Begriff wie *Krieg* tritt mit einem DP-Wert (Gries 2008) von 0,939 nur in vergleichsweise wenigen Songtexten auf.

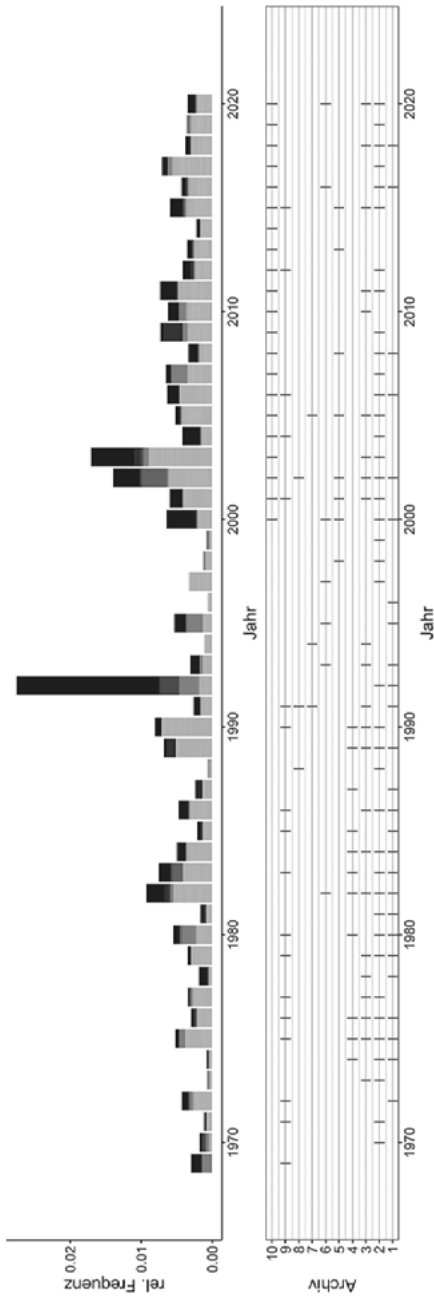


Abb. 3: Diachrone Frequenzen ausgewählter Keywords zur Thematik Gewalt/Krieg/Kriminalität

Am Beispiel des Themengebiets Gewalt/Krieg/Kriminalität lässt sich unter Zuhilfenahme diachroner Analysen nachvollziehen, dass prägende Ereignisse aus der realen Welt mutmaßlich ihre Spuren in Songtexten hinterlassen. Abbildung 3 visualisiert eine Teilmenge der einschlägigen Keywords. Der Spitzenwert im Jahr 1992 legt eine Verbindung zum zweiten Golfkrieg, der Anstieg Anfang der 2000er-Jahre Bezüge zum „Krieg gegen den Terror“ nahe. Ein Blick in die Verteilung über die SK-Archive belegt, dass hier neben den Liedermachern speziell das Charts-Archiv als prominenter Datengeber auftritt.

Zusammenfassend bestätigt die Studie unsere eingangs formulierte Hypothese. Es lässt sich zeigen, dass in Songtexten nicht allein der Themenbereich Liebe/Partnerschaft dominiert, sondern weitere Thematiken mit gesellschaftlich-politischer Relevanz nicht nur sporadisch aufgegriffen werden, zum Teil sogar deutlich häufiger als in allgemeinsprachlichen Texten. Dabei bleiben Texte, in denen eine solche Relevanz gar nicht durch Keywordanalysen fassbar ist, zwangsläufig außen vor. Diese Beschränkung des Ansatzes könnten ergänzende Sentimentanalysen oder Topic Modelling kompensieren, ebenso wären Untersuchungen zur Verteilung von Modalverben (Ermittlung dominanter Modalitäten) oder Satztypen (Quantifizierung von Frage-/Aufforderungscharakter) zur Erhöhung der Deskriptivität denkbar. Die im Rahmen der Studie generierten Keywordlisten und Ergebnisse werden unter <https://songkorpus.de> zur Verfügung gestellt, sodass zukünftige übergreifende Bestandsaufnahmen für den betrachteten Sprachausschnitt nachhaltig auf empirisch abgesicherte Daten zurückgreifen können.

Literatur

- Achermann, Eric/Naschert, Guido (2005): *Songs, Besonderheit von Liedern und Lyrik.* (= Mitteilungen des Deutschen Germanistenverbandes 2). Bielefeld: Aisthesis.
- Ammon, Frieder von/Petersdorff, Dirk von (Hg.) (2019): *Lyrik / Lyrics. Songtexte als Gegenstand der Literaturwissenschaft.* Göttingen: Wallstein.
- Androutsopoulos, Jannis (Hg.) (2003): *HipHop: Globale Kultur – lokale Praktiken.* Bielefeld: transcript.
- Bloching, Sven/Landschoff, Jöran (2018): *Diffamierungen, Humor und Männlichkeitskonstruktion. Eine linguistische Perspektive auf Farid Bangs und Kollegahs Album JBG3.* In: *SPRACHREPORT 4/2018*, S. 14–27.
- Brett, David/Pinna, Antonio (2018). *Words (don't come easy): The automatic retrieval and analysis of popular song lyrics.* In: Suhr, Carla/Nevalainen, Terttu/Taavitsainen, Irma (Hg.) (2019): *From data to evidence in english language research.* (= *Language and Computers 83*). Amsterdam/Leiden: Brill, S. 307–325.

- Blühdorn, Annette (2003): Pop and poetry – pleasure and protest: Udo Lindenberg, Konstantin Wecker and the Tradition of German Cabaret. (= German Linguistic and Cultural Studies 13). Oxford/Bern/Berlin: Lang.
- Bondi, Marina (2010): Perspectives on keywords and keyness. An introduction. In: Bondi, Marina/Scott, Mike (Hg.): Keyness in texts. Amsterdam/Philadelphia: Benjamins, S. 1–18.
- Fankhauser, Peter/Kupietz, Marc (2019): Analyzing domain specific word embeddings for a large corpus of contemporary German. In: International Corpus Linguistics Conference. Cardiff, 2019. DOI: www.doi.org/10.14618/ids-pub-9117 (Stand: 3.8.2021).
- Gries, Stefan Th. (2008) Dispersions and adjusted frequencies in corpora. In: International Journal of Corpus Linguistics 13, 4, S. 403-437. DOI: www.doi.org/10.1075/ijcl.13.4.02gri (Stand: 3.8.2021).
- Jaeger, Florian T. (2008): Categorical data analysis: Away from ANOVAs (Transformation or not) and towards logit mixed models. In: Journal of Memory and Language 59, 4. Special Issue: Emerging Data Analysis. S. 434–446. DOI: www.doi.org/10.1016/j.jml.2007.11.007 (Stand: 3.8.2021).
- Kleiner, Marcus S. (2016): Flüchtlinge und deutsche Popmusik. In: POP. Kultur und Kritik 5, 2, S. 40–44. DOI: www.doi.org/10.25969/mediarep/1538 (Stand: 3.8.2021).
- Kreyer, Rolf (2015): „Funky fresh dressed to impress“: A corpus-linguistic view on gender roles in pop songs. In: International Journal of Corpus Linguistics 20, 2, S. 174–204.
- Kupietz, Marc/Lüngen, Harald/Kamocki, Pawel/Witt, Andreas (2018): The german reference corpus DEReKo: New developments – new opportunities. In: Proceedings of the 11th International Conference on Language Resources and Evaluation (LREC 2018), Miyazaki. Paris: ELRA, S. 4353–4360.
- Lenk, Hartmut E. H. (2011): Phraseologismen im Austropop-Klamauk. Ihre spielerische Verwendung in den Texten der ERSTEN ALLGEMEINEN VERUNSICHERUNG. In: Lenk, Hartmut E. H./Stein, Stephan (Hg.) (2011): Phraseologismen in Textsorten. (= Germanistische Linguistik 211/212). Hildesheim/Zürich/New York: Olms. S. 255–280.
- Mihalcea, Rada/Tarau, Paul (2004): TextRank: bringing order into text. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2004), Barcelona. ACL. S. 404–411. Internet: www.aclanthology.org/W04-3252.pdf (Stand: 3.8.2021).
- Motschenbacher, Heiko (2016): A corpus linguistic study of the situatedness of English pop song lyrics. In: Corpora 11, 1, S. 1–28.
- Nishina, Yasunori (2017): A study of pop songs based on the Billboard Corpus. In: International Journal of Language and Linguistics 4, 2, S. 125–134.
- Psutka, Christine/Grassel, Marco (2018): Porno-Rap. Möglichkeiten sprachlich-subversiver Strategien der Umwertung verletzender Sprachhandlungen im deutschsprachigen Rap. In: SPRACHREPORT 4/2018, S. 28–39.
- Ross, Andrew/Rivers, Damian (Hg.) (2018): The sociolinguistics of hip-hop as critical conscience. Dissatisfaction and dissent. Palgrave Macmillan. Internet: www.digitale-objekte.hbz-nrw.de/storage2/2018/09/05/file_12/8148562.pdf (Stand: 3.8.2021).
- Schneider, Roman (2020): A corpus linguistic perspective on contemporary german pop lyrics with the multi-layer annotated „Songkorpus“ In: Proceedings of The 12th Language Resources and Evaluation Conference (LREC 2020), Marseille. Paris: ELRA, S. 835–841. Internet: www.lrec-conf.org/proceedings/lrec2020/pdf/2020.lrec-1.105.pdf (Stand: 3.8.2021).

Schneider, Roman (2019): „Konservenglück in Tiefkühl-Town“ — Das Songkorpus als empirische Ressource interdisziplinärer Erforschung deutschsprachiger Poptexte. In: Proceedings of the 15th Conference on Natural Language Processing (KONVENS 2019), Erlangen. GSCL. S. 229–236.

Internet: www.corpora.linguistik.uni-erlangen.de/data/konvens/proceedings/papers/KONVENS2019_paper_15.pdf (Stand: 3.8.2021).

Storey, John (2010): Music. In: Storey, John: Cultural Studies and the Study of Popular Culture. Edinburgh: Edinburgh University Press, S. 118-137.

Werner, Valentin (i. Ersch.): Catchy and conversational? A register analysis of pop lyrics. In: Corpora 16, 2.