

Alexander Geyken

Kollokationen im Fußballwortschatz.

Der Beitrag des DWDS-Wortprofils

Abstract

With an estimated 3.5 billion fans football is the world's most popular sport and thus subject to numerous articles and discussions in the media and in everyday life. This paper addresses a specific lexical aspect of football, more precisely it raises the question to what extent syntagmatic structures are covered in current dictionaries. The evaluation is based on two resources for German: the manually compiled Kicktionary, a football specific lexicon and based on framenet, and DWDS-Wortprofil, an automatically built database of statistically salient co-occurrences extracted from large corpora. The main result is that 95% of the lexical units of Kicktionary are contained in the DWDS-Wortprofil which provides evidence for the fact that automatic tools for collocation extraction perform well for this specific vocabulary domain.

1. Einleitung

Gegenstand dieser Arbeit ist die korpusbasierte Beschreibung von Kollokationen in der Fußballsprache. Die Arbeit vergleicht zwei lexikalische Ressourcen, die für das Deutsche aufgebaut wurden: Kicktionary, ein mehrsprachiges Lexikon der Fußballsprache, welches auf der Framesemantik basiert, sowie das DWDS-Wortprofil, eine korpusbasierte Datenbank statistisch salienter Kookkurrenzen. In dieser Arbeit soll der Frage nachgegangen werden, inwieweit das mit automatischen Methoden erstellte Wortprofil die syntagmatischen Strukturen des lexikographisch erstellten Kicktionary abdeckt. Darüber hinaus sollen Unterschiede im Informationsreichtum beider Ressourcen, insbesondere bezüglich der lexikalischen Präferenzen, beschrieben werden. Der Begriff „Kollokation“ wird hier im Sinne von Hausmann ver-

standen. Dieser charakterisiert Kollokationen als „normtypische phraseologische Wortverbindungen, die aus einer Basis und einem Kollokator bestehen. Die Basis ist ein Wort, das ohne Kontext definiert, gelernt und übersetzt werden kann [...]. Der Kollokator ist ein Wort, das beim Formulieren in Abhängigkeit von der Basis gewählt wird und das folglich nicht ohne die Basis definiert, gelernt und übersetzt werden kann“ (Hausmann 2007, 218). Angewandt auf den Fußballwortschatz handelt es sich um Konstruktionen wie *etatmäßiger Libero* (der in der Stammmannschaft auf der Position des *Liberos* spielt), *scharfe Flanke* (für einen fest geschossenen Ball von der Seite), *ins Tor drücken* (für einen Ball, der aus kurzer Entfernung ins Tor geschossen wird) oder *sich abtasten* (für eine wechselseitig vorsichtige Spieltaktik zweier Fußballmannschaften).

Der Bereich des Fußballs wurde für diese Untersuchung aus den folgenden Gründen gewählt:

- Der Fußballwortschatz besteht aus einer begrenzten Menge von Bezeichnungen für Aspekte des Spiels wie z.B. Inventargegenstände, Aktionen und Akteure. Dadurch lässt sich der lexikalische Umfang dieses Wortschatzbereichs vergleichsweise leicht ermitteln.
- Im Gegensatz zu anderen Spezialdiskursen verfügt der Fußballwortschatz über eine Vielzahl von Synonymen, einen hohen Anteil von Phrasemen und typischen Kollokationen sowie eine reichhaltige stilistische Varianz. Diese Reichhaltigkeit an Formulierungsvielfalt ist Ausfluss der gesellschaftlichen Bedeutung des Fußballsports, aber auch der Emotionen, die damit verbunden sind.
- Über Fußball wird in den Medien sehr viel berichtet. Daher ist davon auszugehen, dass der Fußballwortschatz und die damit verbundenen Kollokationen in Zeitungskorpora breit belegt sind und sich somit für statistische und frequenzbasierte Korpusextraktionsmethoden eignen.
- Es liegt bereits eine gut ausgearbeitete lexikographische Beschreibung des Fußballwortschatzes mit dem Kicktionary vor (Schmidt 2010; www.kicktionary.de), welche zu Vergleichszwecken herangezogen werden kann.

Diese Arbeit ist wie folgt aufgebaut. In Abschnitt 2 werden zunächst die Grundlagen für die Untersuchung beschrieben: die Kollokationsdatenbank des DWDS (Digitales Wörterbuch der deutschen Sprache, www.dwds.de) und das DWDS-Wortprofil (vgl. Geyken et al. 2009), welches auf den im DWDS verfügbaren Korpora beruht. In Abschnitt 3 präsentiere ich das Kicktionary, insbesondere den deutschsprachigen Anteil. In Abschnitt 4, dem Hauptteil dieser Arbeit, folgt ein Vergleich zwischen dem DWDS-Wortprofil und dem Kicktionary. Diesem Vergleich liegen quantitative und qualitative

Leitfragen zugrunde: In welchem Umfang wird der im Kicktionary beschriebene Wortschatz vom DWDS-Wortprofil abgedeckt? Wie unterscheiden sich die framebasierten Argumentstellen des Kicktionary und die durch das DWDS-Wortprofil extrahierten syntaktischen Relationen? Abschnitt 5 beschließt die Arbeit mit einer Diskussion der Möglichkeiten und der Grenzen der dargestellten Ressourcen.

2. DWDS-Wortprofil und Sportwortschatz

Das vielleicht wichtigste maschinelle Verfahren zur Unterstützung der lexikografischen Arbeit, welches international in größeren Wörterbuchprojekten eingesetzt wird, stellt die „sketch engine“ dar (Kilgarriff 2004). Die „sketch engine“ enthält neben Korpuserstellungs- und Korpusabfragewerkzeugen auch Kollokationsdatenbanken, sogenannte „word-sketches“. Da maschinelle Verfahren linguistische Kollokationen nicht zuverlässig extrahieren, sondern nur überzufällig häufig miteinander vorkommende Wörter ermitteln können, enthalten „word sketches“ streng genommen keine Kollokationen, sondern nur Kollokationskandidaten. In den „word-sketches“ sind diese tabellenförmig dargestellt und nach syntaktischen Relationen gruppiert bzw. nach statistischer Salienz sortiert. Beispiele für syntaktische Relationen sind die Adjektiv-Nomen-Relation, wie beispielsweise der bereits genannte *etatmäßige Libero* oder die Verb-Objekt-Relation, wie beispielsweise *den Verteidiger umkurven*. Nicht von der „sketch engine“ erfasst werden somit beispielsweise klauselübergreifende Kookkurrenzen, da diese in keiner syntaktisch voneinander abhängigen Relation stehen. Dies gilt in ähnlicher Weise für „triviale“ oder „uninteressante“ Relationen wie Artikel+Nomen (*das Buch*) oder Kardinalzahl+Nomen (*zwei Tische*).

Aufgrund der freien Wortstellung und des Kasussynekretismus ist eine direkte Übertragung des Verfahrens, das für das Englische auf Wortarten und darauf basierenden Satzmustern beruht, nicht sinnvoll, da die dadurch definierten regulären Regeln zu ungenau für die Extraktion syntaktischer Relationen sind. Das im Folgenden beschriebene DWDS-Wortprofil beruht daher auf einem allgemeineren Formalismus, der syntaktische Satzfunktionen erkennen und lokale Mehrdeutigkeiten auflösen kann. Das DWDS-Wortprofil wird seit 2007 entwickelt und dient einerseits der Unterstützung der Kollokationsbeschreibung im Rahmen der lexikografischen Arbeiten des Digitalen Wörterbuchs, es ist andererseits als Darstellungselement in einem eigenen Panel auf der Website des DWDS auch für die Nutzer sichtbar.

Die „Berechnung“ des DWDS-Wortprofils erfolgt in drei Etappen: (1) Zunächst werden die zu extrahierenden syntaktischen Relationstypen festgelegt. Im gegenwärtigen Wortprofil werden zwölf Relationstypen verwendet (Didakowski/Geyken 2013). Davon werden in dieser Arbeit die folgenden verwendet: Adjektiv-Nomen (Etikett: ATTR; *klares Feindbild, schöne Beschierung*), Adverb-Verb (VADV; *schallend lachen, freimütig zugeben*), Subjekt (SUBJ: *der Vertrag läuft aus, der Libero zieht sich (etw.) zu*), Objekt-Verb (Etikett: OBJ; *ein Feindbild abbauen, eine Rede halten*), Präpositionalphrase (PREP_P: *die Abwehr um den Libero*), Genitivattribut (GEN-ATTR: *Ausfall des Liberos*), vergleichende Wortgruppe (KOMP: *sich profilieren als Schauspieler*) sowie die Koordination (KOORD: *Brot und Spiele*). Diese syntaktischen Relationen sind nicht symmetrisch, sondern gerichtet. Beispielsweise steht *hat_ATTR* für die Relation zwischen dem Nomen und dem Adjektiv und *ist_ATTR_zu* für die Relation zwischen dem Adjektiv und dem Nomen. Entsprechend gilt dies auch für die anderen Relationen. (2) Sodann erfolgt die Annotation der syntaktischen Relationen mit Hilfe einer Grammatik, die für die Relationsextraktion optimiert ist. (3) In einem dritten Schritt werden statistische Werte ermittelt, die als quantitatives Maß für die Zusammengehörigkeit von Worttupeln (Paaren oder Tripeln) verwendet werden: Je höher der Wert (Salienzwert), desto höher die Assoziationsstärke. Im Rahmen dieser Arbeit wird das für lexikographische Zwecke geeignete logDice-Maß verwendet (Rychly 2008), welches für statistisch signifikante Kollokationen in der Regel Werte zwischen 5 und 12 liefert – der theoretisch erreichbare Maximalwert liegt bei 14.

Ein Beispiel für *Libero* liefert Tabelle 1. In dieser tabellenförmigen Anordnung sind die Kookkurrenzen nach syntaktischen Relationen (Adjektivattribut, Objekt, Präpositionalgruppe etc.) gruppiert und nach ihrer statistischen Salienz (logDice) sortiert. Die maximale Anzahl der Kookkurrenzen, die pro Tabelle angezeigt werden, lässt sich über einen Schieberegler einstellen. Zu den Korpusbelegen gelangt man durch Klick auf einen Kollokanten. Eine ausführlichere Darstellung des DWDS-Wortprofils findet sich in Didakowski/Geyken (2013).

Libero – Substantiv					
Statistik: logDice ▾		Max. Assoziationen: <input type="checkbox"/> 5		Sortierung: Relevanz ▾	
hat Adjektivattribut		ist Akkusativ/Dativ-Objekt		ist in Präpositionalgruppe	
etatmäßige	8.87	beordert	7.24	Abwehr um	9.68
umsichtigen	8.26	spielen	6.09	Abwehr mit	8.22
gesperrten	7.42	aufgeboten	5.96	Deckung um	7.91
aufgerückte	7.08	aufgelöst	5.80	Hintermannschaft um	7.87
kopfbalstarke	6.91	interpretierte	5.57	Fehlen von	7.59
hat Genitivattribut		in Koordination mit		hat Prädikativ	
Helmer	11.02	Manndecker	12.20	fit	6.47
Fußball-	7.70	Kapitän	10.12	Spieler	6.15
Bundesligisten		Torwart	9.94	Mann	2.79
Nationalelf	7.51	Vorstopper	9.72		
FC	7.05	Mittelfeldspieler	9.45		
Bundesligisten	6.90				

Tabelle 1: DWDS-Wortprofil für *Libero*

Die Korpora, die als Datengrundlage des DWDS-Wortprofils dienen, sind grundsätzlich frei wählbar. Die Zusammensetzung und Größe der Korpora spielen für das Wortprofil jedoch eine wichtige Rolle. Die Zusammensetzung der Korpora ist insofern relevant, als die extrahierten syntaktischen Relationen die im Korpus vorkommenden syntaktischen Nachbarn des Wortes widerspiegeln. Daher erhöht ein breit gestreutes, nach Textsorten ausgewogenes Korpus, ein sogenanntes allgemeinsprachliches Referenzkorpus, die Qualität des Wortprofils in Bezug auf die allgemeinsprachliche Aussagekraft. Spezialkorpora oder Zeitungskorpora werden somit andere Wortprofile liefern als Referenzkorpora. Auch die Größe der Korpora ist für die Qualität des Wortprofils von Bedeutung. Hierfür gibt es derzeit noch keine umfassende empirische Evidenz, aber zumindest einige Erfahrungswerte: So berichten mehrere Studien über die Extraktion von Kollokationen aus Korpora, dass nur für Wörter mit einer Vorkommenshäufigkeit von über 1000 statistisch ausreichend abgesicherte Kookkurrenzprofile extrahiert werden können (Kilgarriff et al. 2004, Ivanova 2008, Geyken 2011), die dann wiederum eine ausreichende Grundlage für die Wörterbucharbeit bilden. Für Korpora von einer Milliarde Textwörtern bedeutet dies für etwa 20.000 Stichwörter eine ausreichende Abdeckung (Hvelplund 2013). In der Arbeit von Hvelplund wird anhand einer Zufallsauswahl von 231 niedrigfrequenten Stichwörtern des *Oxford Advanced Learner's Dictionary* (OALD) extrapoliert, dass für

eine ausreichende Beschreibung der Kollokationen des OALD große Korpora im Umfang von wenigstens 10 Milliarden Textwörtern notwendig wären.

Basis des DWDS-Wortprofils (Stand 2013) ist ein Teilkorpus des DWDS im Umfang von 2,3 Milliarden Textwörtern (siehe Tabelle 2), welches aus dem nach Textsorten und für den Zeitraum des 20. Jh. ausgewogenen DWDS-Kernkorpus sowie einer Auswahl von wichtigen Tageszeitungen aus dem Zeitraum von 1993-2005¹ sowie der Wochenzeitung DIE ZEIT (1946-2010) besteht. Tabelle 2 enthält darüber hinaus die Anzahl der Dokumente, die in den Korpora dem Ressort „Sport“ zugeordnet werden können. Der Anteil der Sportkorpora bei den Tageszeitungen liegt durchschnittlich bei etwa 10% der Gesamtzahl aller Dokumente des Teilkorpus. Davon weichen lediglich die *Bild-Zeitung* (20,5%), welche traditionell über einen gegenüber den anderen Ressorts größeren Sportteil verfügt, und der Berliner *Tagesspiegel* ab. Beim *Tagesspiegel* erklärt sich die große Abweichung (5 %) dadurch, dass in allen Artikeln vor dem 26. April 2002 in den vom Verlag übermittelten XML-Quellen kein einziger Artikel mit dem Ressort „Sport“ ausgezeichnet wurde. Legt man nur den Zeitraum Mai 2002 bis 2005 zugrunde, liegt der Anteil der Sportartikel bei 10,5% (165.792 Artikel von Mai 2002 bis Mai 2005, davon Sport 17.501 Artikel) und damit im Bereich des Durchschnitts der anderen Zeitungen.

Darüber hinaus enthält auch das Kernkorpus keine Angaben über den Sportanteil. Dies liegt daran, dass die Metadaten des Kernkorpus nur mit den vier Textsorten Belletristik, Gebrauchsliteratur, Wissenschaft und Zeitungen annotiert wurden und eine feinere Unterteilung in Zeitungsressorts nicht vorliegt. Schließlich verfügt die Wochenzeitung DIE ZEIT traditionell über einen gegenüber den anderen Ressorts kaum vorhandenen Sportanteil (0,4 %).

KORPUS	WÖRTER	SÄTZE	DOKU- MENTE	DOKU- MENTE SPORT	ANTEIL SPORT
Kernkorpus	125.990.080	7.046.937	79.312	---	---
Berliner Zeitung	242.046.373	15.951.701	869.023	91.488	0,105
Tagesspiegel	184.202.717	10.392.257	394.465	17.765	0,045

¹ Vgl. www.dwds.de/textbasis.

Süddeutsche Zeitung	453.945.194	29.125.790	1.099.920	104.063	0,094
DIE WELT	238.403.711	15.787.624	600.007	54.211	0,090
Bild	121.520.037	12.629.828	548.181	112.864	0,205
DIE ZEIT	417.422.714	23.631.230	499.520	2.266	0,004
FAZ	557.876.936	30.914.994	1.129.338	153.394	0,135
insgesamt	2.341.407.762	145.480.361	5.219.766	536.051	0,678

Tabelle 2: Korpusgrundlage des DWDS-Wortprofils

Aus dem in Tabelle 2 dargestellten Korpus von 2,3 Milliarden Textwörtern wurde mit den oben beschriebenen Verfahren eine Datenbank mit 104.704 Lemma/Part-of-Speech-Paaren² und 11.980.910 verschiedenen syntaktischen Kookkurrenzen erstellt. Im Schnitt verfügt ein Lemma somit über 120 verschiedene Kookkurrenzen.

3. Fußballwortschatz: das Kicktionary

Das Kicktionary (www.kicktionary.de) ist ein dreisprachiges Wörterbuch und lexikalisches Informationssystem zur Fußballsprache (Deutsch, Englisch, Französisch), welches über das Internet abfragbar ist. Das Kicktionary beruht auf der Framesemantik (Fillmore 1982) und nutzt die in WordNet definierten Relationen für die relationalen Bezüge zwischen Lemmata bzw. zwischen einzelnen Bedeutungspositionen von Lemmata (Fellbaum 1998). Basis der Erstellung des deutschen Teils des Kicktionary sind zwei Spezialkorpora der Fußballsprache, darunter ein etwa eine Million Textwörter umfassendes Korpus geschriebener Sprache und ein deutlich kleineres, etwa 15.000 Textwörter umfassendes Korpus gesprochener Sprache.

² Der Grund, weshalb man Lemma/Part-of-Speech-Paare und nicht einfach Lemmata verwendet, besteht darin, dass Lemmata bezüglich ihrer Wortart im STTS-Tagset ambig sein können und somit in den Wortprofilen vermischt würden. Beispielsweise ist (*grau*, Adjektiv) Attribut eines Nomens, wohingegen (*grau*, Modifizierer) Attribut eines Adjektivs ist.

Das Kicktionary besteht konzeptionell aus vier Teilbereichen: Szenen, Frames, Frame-Elementen und lexikalischen Einheiten (Lemmata und Mehrworteinheiten, abgekürzt LE). Ein Lexikoneintrag für eine LE besteht aus folgenden Informationen: Angabe zur Sprache (obligatorisch), zur lexikalischen Kategorie (optional), zur Definition (optional), einer Reihe von Korpusbeispielen (obligatorisch). Jede LE ist genau einem Frame zugeordnet. Frames enthalten z.B. die an ihnen beteiligten Aktanten und Gegenstände, sogenannte Frame-Elemente. Die lexikalischen Relationen (Hyponymie, Hyperonymie, Holonymie, Meronymie) werden vornehmlich für die im Kicktionary enthaltenen Nomen verwendet. Durch Verkettung dieser Relationen entstehen Konzepthierarchien. Szenen bündeln mehrere Frames und beschreiben auf allgemeinerer Ebene als Frames prototypische Handlungsabläufe. Jeder Frame ist genau einer Szene zugeordnet. Die Szenen und Frames dienen als Mittel zur Makrostrukturierung des Wörterbuchs mit dem Ziel der Bildung ‚nützlicher‘ Frames. Als ‚nützlich‘ werden Frames dann angesehen, wenn die Anzahl der darin enthaltenen lexikalischen Einheiten (LE) weder zu klein noch zu groß ist (Schmidt 2011).

Aus den Spezialkorpora des Kicktionary wurden für die drei Sprachen Deutsch, Englisch und Französisch insgesamt 1874 lexikalische Einheiten gewonnen und dann Frame-Elementen, Frames oder Szenen zugeordnet. Der deutsche Teil von Kicktionary umfasst 792 LE, darunter 451 Nomen, 305 Verben und 46 anders kategorisierte Einheiten. Diese sind 104 Frames zugeordnet, die wiederum in 16 Szenen zusammengefasst sind. Da jeder Frame genau einer Szene zugeordnet ist, enthält jede Szene durchschnittlich 6,5 Frames.

Beispiele für Szenen sind „der Schuss“, die „Eins-gegen-Eins-Situation“, „der Wechsel“ oder „der Spielstand“. Diese lassen sich wiederum in semantisch „feinere“ Frames unterteilen. So gehören beispielsweise zur „Eins-gegen-Eins-Situation“ die Frames CHALLENGE, BEAT, DENY, die wiederum aus folgenden LE bestehen:

CHALLENGE: *angreifen, attackieren, bedrängen*

BEAT: *ausspielen, umdribbeln, tunneln, vernaschen*

DENY: *abgrätschen, abjagen, blocken, stoppen*

Eine LE kann mehreren Frames zugeordnet sein. So gehört beispielsweise das Verb *angreifen* zu den beiden Frames CHALLENGE und MOVE. Im Frame CHALLENGE ist *angreifen* über die Frame-Elemente *ballführender Spieler, angreifender Spieler* und der *Ort, an dem angegriffen wird*, mit den dazugehörigen Akteuren bzw. Gegenständen verknüpft. Die Frame-Elemente wiederum werden mit konkreten LE bzw. Eigennamen gefüllt. Ein typisches Beispiel für den Frame CHALLENGE mit der LE *angreifen* wäre: *Max wur-*

de in guter Schussposition vor dem Strafraum nicht angegriffen. Wichtig ist hier festzustellen, dass nicht alle Frame-Elemente in den Korpora auch realisiert werden. Im Frame MOVE wiederum hat *angreifen* die Frame-Elemente *Team*, *Führungsspieler* und *Ort*. Ein typisches Beispiel hierfür ist: *Real greift über die Flügel an*. Den Frame-Elementen sind somit wiederum LE zugeordnet, und sie sind ihrerseits über lexikalische Relationen miteinander vernetzt. Beispielsweise steht die LE *Manndecker* in der Hyperonymrelation mit der LE *Verteidiger*, welche wiederum über das Holonym *Verteidigung* verfügt. Die LE *Mannschaft* verfügt beispielsweise über die Meronyme *Sturm*, *Mittelfeld* und *Abwehr*.

Zur Einordnung der Methode der Frames für die Beschreibung des Fußballwortschatzes kann mit Schmidt (2011) festgestellt werden, dass die Szenen- und Frames-Analysen eher für den Verb- als für den Nomenbereich geeignet sind. Insbesondere sind Szenen und Frames deshalb nützlich zur Strukturierung des Wortschatzes, weil „sich über sie das (teilweise nicht-sprachliche) Wissen über prototypische Handlungsabläufe in Bezug zu sprachlichen Mitteln setzen lässt, mit denen dieses Wissen ausgedrückt werden kann“ (Schmidt 2011, 59). Für die Elemente der Fußballsprache, die keine Handlungsabläufe bezeichnen, wie beispielsweise die Bezeichnungen für Spielerpositionen (*Libero*, *Verteidiger*, *Stürmer* etc.) oder die Bezeichnungen für Gegenstände auf dem Spielfeld (*Tor*, *Strafraum*, *Mittellinie* etc.), ist die Frame-Analyse nicht das Beschreibungsmittel der ersten Wahl. Hier sind semantische Relationen deutlich geeigneter zur Wörterbuchstrukturierung.

4. Wortprofil und Kicktionary im Vergleich

Wie im vorigen Abschnitt beschrieben, handelt es sich bei dem Kicktionary um eine umfangreiche lexikografische Beschreibung des Fußballwortschatzes. Vor diesem Hintergrund stellt sich die Frage, wie groß der Grad der Abdeckung dieses Wortschatzes durch das DWDS-Wortprofil ist.

Der Vergleich beider Ressourcen soll dabei quantitativ und qualitativ erfolgen. Auf quantitativer Ebene müssen die den Ressourcen zugrundeliegenden Korpora verglichen werden. Der Aufbau von Kicktionary basiert auf zwei kleinen Spezialkorpora der Fußballsprache von etwas über 1 Million Token. Das DWDS-Wortprofil basiert hingegen auf einem 2,3 Mrd. Textwörtern großen Korpus überwiegend moderner Zeitungssprache, wovon – wie oben erläutert – etwa 10 % aus dem Themenbereich „Sport“ stammen. Da die Metadaten der DWDS-Korpora nur die allgemeine Oberkategorie

„Sport“, nicht aber „Fußball“ umfassen, lässt sich der genaue Anteil von Fußballtexten in den Korpora nicht automatisch ermitteln. Aber selbst wenn man konservativ annimmt, dass nur jeder zehnte Sportartikel das Thema Fußball adressiert, so wäre das daraus resultierende Teilkorpus des Wortprofils immer noch mindestens zehnmals größer als das Kicktionary-Korpus. Das Beispiel zeigt aber auch, dass es wünschenswert wäre, die Metadaten der DWDS-Korpora im Hinblick auf Themen, Sachgebiete, Texttypen usw. zu verfeinern.

Vergleicht man die lexikalischen Einheiten (LE) des Kicktionary mit den Lemmata des DWDS-Wortprofils, so ergibt sich folgendes Bild (siehe Tabelle 3): Für 675 der 792 bzw. 706 (siehe unten) der deutschen LE des Kicktionary gibt es auch DWDS-Wortprofile.

	Gesamt	Nomina	Verben	Rest
Kicktionary (deutsch)	792	451	305	46
Kicktionary (bereinigt)	706	431	264	13
davon im Wortprofil	675	411	253	13

Tabelle 3: Kicktionary und DWDS-Wortprofil

Diese Zahlen kommen wie folgt zustande: Zunächst reduziert sich die Kandidatenliste von 792 auf 706 (bereinigte Liste), da die LE im Kicktionary Mehrwortausdrücke enthalten, z.B. *auf dem falschen Fuß erwischen* oder *Vorteil gelten lassen*, die per se im Wortprofil nicht enthalten sind, da diese außerhalb der derzeitigen Musterextraktion (syntaktische Relationen) des Wortprofils stehen. Dies sind insgesamt 32 Fälle. Darüber hinaus enthält das Kicktionary 54 Duplikate, d.h. LE, die mehreren Frames zugeordnet sind, z.B. das bereits erwähnte *angreifen* (Frame CHALLENGE und MOVE), *abfangen* (Frame DENY oder INTERCEPT), *Angriff* (MOVE oder TEAM), *Duell* (ONE-ON-ONE SITUATION, MATCH) oder *geben* (PASS, REFEREE_DECISION, AWARD_GOAL).

Von den verbleibenden 706 Kandidaten des Kicktionary gibt es für 675 LE ein Wortprofil: das sind 95,6 %. Für 31 LE (4,4 %) gibt es kein Wortprofil. Dies sind *Anstoßpunkt*, *antäuschen*, *Bananenflanke*, *durchwechselln*, *einnetzen*, *Eins-gegen-Eins-Situation*, *ersprinten*, *freilaufen*, *Halbvolley*, *Halbzeit2*, *Hechkopfball*, *herauseilen*, *herunterstoppen*, *Kopfball-Torpedo*, *Linksflanke*, *Maßflanke*, *Mittelfeldgeplänkel*, *nachstoßen*, *Pass-Stafette*, *Pressball*, *Pressschlag*, *Rechtsaußen*, *Revanchefoul*, *Scherenschlag*, *Schiedsrichter-assistent*, *Sechzehner*, *Sechzehnmeterraum*, *überlupfen*, *umgrätschen*, *um-*

rempeln, *Volleykracher*. Hierbei fällt auf, dass einige dieser LE durchaus häufig in den DWDS-Korpora vorkommen. Beispielsweise kommen *antäuschen* 73 Mal, *Bananenflanke* 104 Mal, *Hechkopfball* 73 Mal in den DWDS-Korpora vor. Jedoch sind die syntagmatischen Kontexte nicht regelmäßig genug (für den Syntaxparser des Wortprofils), um daraus Kollokationskandidaten zu extrahieren.

Für diejenigen LE aus Kicktionary, für die es ein Wortprofil gibt, ergibt sich die Häufigkeitsverteilung, wie in Tabelle 4 aufgeführt. So bedeutet etwa Zeile Nr. 1, dass für 93 LE aus Kicktionary ein Wortprofil mit einer Zahl von ein bis zehn Kollokationskandidaten existiert (LogDice Schwellwert = 0; Minimalfrequenz = 5), usw.

	Anzahl LE	Frequenz
0	31	0 mal
1	93	1-10 mal
2	143	11-50
3	132	51-250
4	128	251-1000
5	159	1001-10.000
6	20	< 19.414
Gesamt	706	914.745

Tabelle 4: Häufigkeitsverteilung des Kicktionary im Wortprofil nach Kollokationen

Tabelle 5 enthält die Häufigkeitsverteilung bezüglich der Anzahl der verschiedenen syntaktischen Relationen zu einer LE des Kicktionary:

	Anzahl LE	Anzahl Syntaktische Relationen
0	32	0
1	59	1-2
2	89	3-5
3	277	6-10

4	249	11-12
Gesamt	706	5.660

Tabelle 5: Häufigkeitsverteilung des Kicktionary im Wortprofil nach syntaktischen Relationen

Der rein quantitative Vergleich hat jedoch in seiner Aussagekraft Grenzen, da die Kollokationen zu einem Lemma in Kicktionary entweder gar nicht aus dem Fußballbereich stammen (z.B. *auflaufen* mit der Kollokation *auf einer Sandbank*), ambig zwischen Fußball und einer anderen Domäne sein können, wie z.B. *steile Flanke* (neben Fußball auch im Alpinismuskontext), *jdn. vernaschen* (neben Fußball auch Boxsport oder Sexualität), oder unterschiedlichen Sportkontexten zugeordnet sein können, z.B. *das Sprintduell verlieren* (auch Leichtathletik).

Aufschlussreich ist daher ein qualitativer Vergleich zwischen beiden Ressourcen, der mit folgenden Beispielen veranschaulicht werden soll. Im Kicktionary wird die Szene *Actors* in folgende Frames unterteilt: TEAM, PLAYER, COACH, REFEREE, BODY_PARTS, EQUIPMENT, BALL, SPECTATORS. Der Frame PLAYER wiederum umfasst die folgenden deutschsprachigen LE: *Abräumer, Abwehrspieler, Akteur, Angreifer, Außenstürmer, Außenverteidiger, Ersatzmann, Ersatzspieler, Feldspieler, Flügelstürmer, Gegenspieler, Goalgetter, Innenverteidiger, Joker, Kapitän, Keeper, Leistungsträger, Libero, Linksaußen, Linksverteidiger, Mitspieler, Mittelfeldmotor, Mittelfeldspieler, Mittelstürmer, Rechtsaußen, Rechtsverteidiger, Regisseur, Reservist, Routinier, Schlüsselspieler, Schlussmann, Spieler, Spielführer, Spielgestalter, Spielmacher, Stammspieler, Stürmer, Teamkollege, Torhüter, Torjäger, Torwart, Verteidiger*.

Diese LE sind Bestandteil der folgenden Frames: PLAYER, TEAM, POSITION_SPECIFICATION. Im Kicktionary gehören z.B. *Abwehrspieler* oder *Libero* nur zum Frame PLAYER, andere wie *Angreifer* oder *Außenstürmer* gehören zusätzlich zum Frame TEAM (wie in *Angreifer des FC Köln*). Diese Zuordnung ist jedoch nicht vollständig, sondern entspricht den Belegsätzen aus den Korpora des Kicktionary. Beispielsweise könnten grundsätzlich alle LE aus dem Frame PLAYER einem TEAM zugeordnet werden, d.h. ein Satz der Bauart *X des FC Köln* ist für jedes X aus dem Frame Player möglich. Spezifisch für einzelne LE hingegen ist der Frame POSITION_SPECIFICATION. Dieser trifft nur selten zu, beispielsweise für die LE *Abräumer*, wie in *Abräumer vor der Abwehr*.

Die Frames PLAYER und TEAM werden nicht vom DWDS-Wortprofil erfasst, da es sich dabei um paradigmatische Kontexte der o.g. LE handelt. Die syntagmatischen Kontexte durch die Frames TEAM und POSITION-

_SPECIFICATION werden jedoch vom Wortprofil durch die beiden syntaktischen Relationen Genitivattribut (z.B. *Abwehr des FC Köln*) und Präpositionalgruppe (z.B. *Abräumer vor der Abwehr*) abgedeckt. Auf der anderen Seite enthalten die Wortprofile zu den o.g. LE wichtige weitere Informationen zu typischen Verbindungen. Im Folgenden wird das für die LE *Abwehrspieler*, *Verteidiger*, *Libero* und *Manndecker* beschrieben. In Klammern hinter den einzelnen LE ist die Anzahl der Kookkurrenzen vermerkt. Der Schwellwert für die logDice-Salienz ist 5. Es werden nur die Top-5 Kookkurrenzen aufgeführt (wenn vorhanden).

Abwehrspieler (288):

- hat_ATTR: *kopfballstarke, torgefährliche, gegnerische, ausgeliehene, aufgerückte A.*,
- ist_SUBJ_zu: *der A. zieht sich etwas zu, laboriert an etwas, unterschreibt, köpft, fällt aus, pausiert*
- hat/ist_PREP_P_zu: *der A. vom Fußball-Zweitligisten, vom Fußball-Meister, vom Bundesligisten; Ablöse für den A., Kopfball von A.*
- ist_PREP_P_zu: *umfunktioniert zum, abgefälscht von, Handspiel vom A.*
- ist_GEN-ATTR_zu: *Kopfball, Transfer, Verpflichtung, Ausfall, Bein, Wechsel des A.*
- ist_OBJ_zu: *den A. umkurven, verpflichten, ausleihen, auswechseln, nominieren*
- ist_KOMP_zu: *einsetzen als A.*

Verteidiger (1550):

- hat_ATTR: *eisenharte, gegnerische, linke, rechte, verletzte, stürmende V.*
- ist_SUBJ_zu: *der V. plädiert, beantragt, lädt ein, argumentiert, zieht sich zu*
- ist_PREP_P_zu (mit Nomen): *Eigentor, Kopfball von; Kreuzverhör von; Platzverweis für den V.*
- ist_PREP_P_zu (mit Verb): *umschulen, aufschwingen, umfunktionieren zum V.;*
- ist_GEN-ATTR_zu: *Plädoyer, Antrag, Befragen, Befangenheitsantrag, Beweisantrag; Eigentor des V.*
- ist_OBJ_zu: *den V. verpflichten, zustellen, aufbieten, einwechseln, umkurven, narren*

- ist_KOMP_zu: *sich aufspielen als, aushelfen als; sich hervortun als, sich profilieren als*

Libero (310):

- hat_ATTR: *etatmäßige, umsichtige, gesperrte, aufgerückte, kopfballstarke L.,*
- ist_SUBJ_zu: *L. fällt aus, zieht (sich) etw. zu, läuft aus, verletzt (sich)*
- ist_PREPP_zu: *Abwehr um/mit den/dem Libero, L. hinter den Mann-deckern; L. vor der Abwehr; Deckung um den L.; umfunktioniert zum L.;*
- ist_GEN-ATTR_zu: *Part des, Ausfall des, Erfinder des, Posten des L.,*
- ist_OBJ_zu: *den L. beordern, spielen, aufbieten, interpretieren, abschaffen*
- ist_KOMP_zu: *aufbieten als, auflaufen als, aushelfen als L.*

Manndecker (80):

- hat_ATTR: *eisenharte, aufgerückte, etatmäßige, rustikale M.*
- ist_SUBJ_zu: *M. fällt aus, (alles andere mit Salienz < 4)*
- ist_PREP_P_zu: *Libero hinter dem M., Abwehr mit dem M.; umfunktioniert zum M.; Kopfball von dem M.*
- ist_GEN-ATTR_zu: *Position des M. (alles andere mit Salienz < 5)*
- ist_OBJ_zu: *den M. aufbieten (alles andere mit Salienz < 5)*
- ist_KOMP_zu: *ranmüssen, auflaufen, nachrücken, ersetzen als M.*

Ein weiteres Beispiel stellen die nahen Synonyme *Platzherren* (15) vs. *Heimmannschaft* (12) dar. Beide entstammen der Szene MATCH und dem Frame Heimspiel (HOMEGAME). Sie umfassen deutlich weniger saliente Verbindungen als das häufige *Libero*, daher wurde als Schwelle ein Salienz-wert > 0,5 gewählt.

Bei *Platzherren* (nur Pl.) finden sich folgende saliente Verbindungen:

- ist_SUBJ_zu: *die P. agieren, bestimmen*
- ist_PREP_P_zu: *erzielte für die P.*
- ist_GEN-ATTR_zu: *Tor, Führung der P.*
- ist_OBJ_zu: *den P. fehlte*
- ist_KOMP_zu: –

Bei *Heimmannschaft* finden sich:

- hat_ATTR: *überlegene, schwächste, schlechteste H.* (*beste* hat einen Salienzwert von -0.08)
- ist_SUBJ_zu: *die H. gewinnt, spielt*
- ist_GEN-ATTR_zu: *Fans, Spieler, Tor, Anhänger, Sieg der H.*
- ist_KOMP_zu: *auftrat wie die H.*

Als Zwischenfazit lässt sich festhalten: Es gibt Unterschiede im Grad des Anteils der Fußballkollokationen für die einzelnen Lemmata. Bei *Verteidiger* beziehen sich die meisten Kollokationen auf eine andere Lesart („Strafverteidiger“), bei *Abwehrspieler*, *Platzherren*, *Heimmannschaft* und *Manndecker* auf den Sportkontext, bei *Libero* sogar eindeutig auf den Fußballkontext. Aber auch die Kollokanten unterscheiden sich stark voneinander, selbst bei nahen Synonymen. Beispielsweise selektieren *Verteidiger* und *Abwehrspieler* ganz andere typische Kollokationen: Während bei *Verteidiger* Härte (*eisenhart*) und formale Position bei der Aufstellung (*linke, rechte*) im Fokus sind, sind dies bei *Abwehrspieler* dessen spielerische Flexibilität: *kopfballstark, stürmend, aufgerückt*. Bei *Platzherren* werden gar keine Adjektive verwendet, wohingegen diese bei dem Quasi-Synonym *Heimmannschaft* geläufig sind. Bei den Subjektpositionen steht bei *Platzherren* die Dominanz (des Spiels) im Vordergrund, während bei *Heimmannschaft* das spielerische Vermögen thematisiert wird. Bei den Genitiv-Attributen steht bei *Platzherren* die Aktion im Vordergrund, bei *Heimmannschaft* werden demgegenüber vor allem die Akteure benannt.

Im Zentrum der framebasierten Analyse stehen die Verben. Als Grundlage für den Vergleich von Kicktionary und DWDS-Wortprofil soll hier exemplarisch die Szene Eins-gegen-Eins (One_On_One) und dort der Frame ‚Ausspielen‘ (BEAT) dienen. Die Liste der LE umfasst: *abschütteln, ausdribbeln, ausspielen, austanzen, austricksen, düpieren, entwischen, narren, sich_durchsetzen, sich_lösen, stehen_lassen, tunneln, überlaufen, überspringen, umdribbeln, umkurven, umspielen, verladen, vernaschen, versetzen*. Der Frame BEAT hat die Frame-Elemente ballführender Spieler (PLAYER), angreifender Spieler (OPPONENT PLAYER), Spielfeld (AREA) und Art und Weise (ACTION). Die meisten Verben enthalten nur die ersten drei Frame-Elemente. Nur *austricksen, narren* und *sich_lösen* beziehen das Element ACTION ein, z.B. *A trickst B mit einer Drehung aus*. Andere Verben wie *tunneln* oder *umkurven* inkorporieren die Semantik der Art und Weise des Ausspielens. Aus Platzgründen soll hier für den Vergleich der Informationstiefe des Kicktionary mit dem Wortprofil nur auf folgende Verben Bezug genommen werden: *abschütteln, ausdribbeln, tunneln, ausspielen, umkurven, vernaschen* und *umdribbeln*.

Das Verb *abschütteln* ist ein relativ hochfrequentes Verb (118 Kollokationen). Unter den 20 salientesten Kollokationen finden sich jedoch nur zwei fußballtypische Kollokationen: *Bewacher* und *Gegenspieler*. Alle anderen stammen aus politischen oder sozialen Kontexten oder beschreiben die literale Bedeutung des Verbs *abschütteln*, z.B. *Joch*, *Verfolger*, *Fremdherrschaft*, *Staub*. Für dieses Verb wird somit nur auf den Frame OPPO-NENT_PLAYER Bezug genommen, die anderen Frames werden vom Wortprofil nicht extrahiert.

Genau entgegengesetzt ist die Situation bei dem Verb *ausdribbeln*, welches nur über zwei Kollokationen verfügt, die beide dem Fußballkontext zuzuordnen sind: *er* (Subjekt) sowie die Relation AKK/DAT-OBJ = *Gegenspieler*. Ebenfalls selten ist das Verb *tunneln*, welches im Wortprofil vornehmlich die physikalische Lesart (*Elektronen t.*) umfasst. Nur das Objekt findet bei *tunneln* im Wortprofil typische Wortverbindungen, das Subjekt wird nur über Eigennamen oder Pronomen erwähnt.

tunneln (6):

- hat_VADV: *hindurch* (nur mit Subj. *Elektronen*)
- hat_PREP_P: *durch Barriere t.*
- hat_SUBJ: *Elektronen*
- hat_AKK/DAT-OBJ: *Gegenspieler, Gegner*

Allen Verben aus der Gruppe des BEAT-Frames ist gemein, dass das Wortprofil nur den Frame OPPONENT_PLAYER in Form der Relation AKK/DAT-OBJ kodiert. Den Frame PLAYER hingegen enthält das Wortprofil nicht, da dieser typischerweise in Form eines Eigennamens kodiert ist, z.B. *X tunnelt seinen Gegenspieler...* Eigennamen werden jedoch vom Wortprofil gefiltert, da sie zwar pragmatisch, nicht aber linguistisch interessante Informationen liefern. Auch über die beiden Frames AREA und ACTION sagen die Wortprofile beider Verben nichts aus, da keine Adverbiale vom Wortprofil ermittelt werden konnten.

Komplex ist das Wortprofil des Verbs *ausspielen*, welches entweder in Nicht-Sport-Kontexten (*Marktmacht, als Trumpfkarte, im Stadtverkehr*) vorkommt, oder aber in Sportkontexten, die nicht zum Fußball oder nicht in den BEAT-Frame gehören (z.B. *in Sechsergruppe, in Play offs*), sondern einem Frame, wie z.B. *Qualification_And_Elimination* zugehörig ist. Eindeutig dem Fußballkontext zuzurechnen ist nur die LE *Torwart*. Im Fußballkontext (aber nicht nur dort) finden sich ferner im Frame ACTION *geschickt, gekonnt, Überlegenheit* und *Schnelligkeit*. Mehrfach ambig ist z.B. auch die Kollokation *Stärken ausspielen*. Diese bezieht sich primär auf politische,

wirtschaftliche oder technische Kontexte, kann sich aber auch auf den Sportkontext beziehen (z.B. „... Badstuber, der seine Stärken besonders gegen sich in der eigenen Hälfte verschanzende Gegner ausspielen kann.“ (DIE ZEIT, 06.09.2010).

ausspielen (522):

- hat_VADV: *gegeneinander, voll, geschickt, genüsslich, gnadenlos, gekonnt*
- hat_PREP_P: *nach Modus, zum Stich, in Sechsergruppe, in Play-offs, im Stadtverkehr a.*
- hat_SUBJ: *Hinterhand, Vorhand, Stärken a.*
- hat_OBJ: *Trumpf, Stärke, Überlegenheit, Routine, Schnelligkeit, Vorteile, Karte, Torwart, Marktmacht*
- hat_KOMP: *als Trumpfkarte, im Stadtverkehr*

Aufgrund der Kollokationen ausschließlich auf den Fußballkontext fokussiert ist das Verb *umkurven*. Hier ist nur eine Kollokation nicht dem Fußballkontext zuzuordnen (*Schlaglöcher*). *Umkurven* umfasst in Kicktionary die Frameelemente PLAYER, OPPONENT_PLAYER, ACTION und AREA, die sich allesamt auch im DWDS-Wortprofil finden:

umkurven (38)

- hat_VADV: *weiträumig, elegant, geschickt, souverän,*
- hat_PREP_P: *nach Steilpass, Traumpass, Zuspiel, Pass u.; im Strafraum u.*
- hat_SUBJ: *Stürmer*
- hat_OBJ: *Torwart, Schlussmann, Abwehrspieler, Keeper, Schlaglöcher, Gegenspieler,*
- hat_KOMP: *wie Slalomstangen u.*
- KOORD: *einschob, schoss, traf, erzielte*

Ebenso wie bei den Nomen ist es auch bei den Verben so, dass nahe Synonyme zum Teil über ganz unterschiedliche Wortprofile verfügen können. Beispielsweise sind die Verben *vernaschen* oder *ausdribbeln* im Fußballkontext synonym, da beide die Kollokationen *Gegenspieler* und *Gegner* besitzen. In anderen Kontexten werden die Verben allerdings ganz unterschiedlich verwendet, z.B. schwingt bei *vernaschen* der ironische Kontext (*zum Frühstück*) oder der sexuelle Kontext (*Mädchen, Männer*) mit.

5. Möglichkeiten und Grenzen des Wortprofils zur Analyse des Sportwortschatzes

Der Vergleich des DWDS-Wortprofils mit dem Kicktionary lässt die Aussage zu, dass das Wortprofil den Fußballwortschatz bereits erstaunlich gut abdeckt. Dies lässt sich nicht nur an der quantitativen Abdeckung ablesen – für mehr als 95% aller lexikalischen Einheiten des Kicktionary gibt es ein Wortprofil – sondern auch am qualitativen Vergleich. Hier zeigt sich, dass die typischen Verbindungen, die das Wortprofil aufführt, oft sogar informationsreicher sind als die Frame-basierten Informationen des Kicktionary. Dies gilt im Nomenbereich noch stärker als für den Bereich der Verben, weil die in den Beispielen betrachteten Wortprofile überwiegend deutlich kleiner waren als diejenigen der Nomen. Größere Korpora könnten hier in der Zukunft Abhilfe schaffen. Dies legt auch die Untersuchung von Hvelplund (2013) nahe, in der nachgewiesen wurde, dass Korpora unterhalb einer Größenordnung von 10 Milliarden Textwörtern zu klein für die Aktualisierung größerer einsprachiger Wörterbücher sind. Mit anderen Worten: Erst bei einer Größe von etwa 10 Milliarden Textwörtern erhalten auch Wörter im mittleren bis unteren Frequenzbereich ein ausreichend großes Wortprofil.

Ein Problem bei der Auswertung der Wortprofile zur Beschreibung des Wortschatzes stellt die Polysemie der betrachteten Wörter dar. Zunächst einmal kann von der Anzahl der Kollokationen eines Wortes nicht auf die Bedeutung des Wortes im Fußballbereich geschlossen werden, da die Häufigkeitsverteilung von Lesarten polysemer Wörter nicht vorhersehbar ist. Beispielsweise sind unter den 50 salientesten Kollokationen von *Stürmer* 48 aus dem Sportkontext. Für das Holonym *Sturm* stammt hingegen nur eine einzige Kollokation (*Mittelfeld*) aus dem Bereich des Fußballs. In den mittleren und unteren Salienzbereichen der insgesamt 1382 Kookkurrenzen des Wortprofils zu *Sturm* finden sich jedoch weitere Fußballkollokationen, die zum Teil aber auch ambig mit der meteorologischen Lesart sein können: *harmloser/gefährlicher Sturm*, *Sturm der Liga/Bundesliga*, *Sturm und Mittelfeld*; *der Sturm ist harmlos*, *den Sturm stoppen* („wer stoppt den gefährlichsten Sturm der Liga?“).

Abschließend soll auf einen anderen Aspekt eingegangen werden, nämlich auf die Abdeckung von phraseologischen Wendungen des Fußballbereichs in den beiden Ressourcen. Das WDR hat kurz vor der Fußball-WM 2014 eine fünfminütige Satire ausgestrahlt mit dem Titel „Fußballfloskeln wörtlich genommen“ (WDR vom 22.6.2014). Diese erlangte einen großen Bekannt-

heitsgrad und wurde beispielsweise alleine auf YouTube über 2,5 Millionen Mal aufgerufen³. In dem Film wurden viele bekannte Phraseme aus dem Fußballbereich bzw. Lemmata mit einer Spezialbedeutung im Fußballjargon wörtlich genommen und nachgespielt. Im Anhang sind alle in dieser Sendung erwähnten 41 Phraseme aus dem Fußballwortschatz aufgeführt. Der Abgleich mit Kicktionary und DWDS-Wortprofil zeigt zunächst, dass Kicktionary in diesem Bereich der Fußballphrasen noch sehr lückenhaft ist: Die Abdeckung beträgt nur 6 von 41. Zwar ist auch das Wortprofil nicht vollständig, enthält aber immerhin 30 der 41 Phrasen. Hierfür sind drei Gründe zu nennen: 1) Manche Phraseme kommen in den Korpora des DWDS nicht oder zu selten vor und sind somit nicht im Wortprofil enthalten. Beispiele hierfür sind: (*Verteidiger*) *klebt am Angreifer* und *die/eine Abwehrkette bilden*; 2) Es gibt Phraseme, die nicht dem verwendeten Hausmann'schen Kollokationen-Begriff entsprechen und somit nicht im Wortprofil enthalten sind. Das sind Fälle wie *der (Torwart/Spieler) sah bei dieser Aktion nicht gut aus* oder *die Mannschaft kommt nicht aus dem (eigenen) Strafraum heraus*; 3) Teilweise finden sich im Wortprofil Alternativen, die in der Regel in der Allgemeinsprache auch geläufiger sind. Beispielhaft seien hier die beiden im Film genannten *die Mauer dirigieren* oder *die rechte Seite beackern*, die im Wortprofil nicht aufgeführt sind, erwähnt. Im Gegensatz dazu sind *die Abwehr dirigieren* bzw. *die rechte Außenbahn beackern* aber im Wortprofil dokumentiert.

Zusammenfassend lässt sich zum Vergleich der beiden Ressourcen sagen, dass die DWDS-Wortprofile Aussagen über die tatsächlich realisierten Präferenzen der Verwendungsweisen in den Textkorpora liefern, wohingegen die framebasierte Analyse des Kicktionary eher den Möglichkeitsraum über die beteiligten Aktanten bzw. Gegenstände aufspannt, ohne aber deren Gebrauchlichkeit in den Textkorpora näher zu spezifizieren.

6. Literatur

Didakowski, J./Geyken, A. (2012): From DWDS corpora to a German Word Profile – methodological problems and solutions. In: OPAL – Online publizierte Arbeiten zur Linguistik 2, 43-52.

Fellbaum, Ch. (Hg.)(1998): WordNet – An Electronic Lexical Database. Boston.

³ https://www.youtube.com/watch?v=mf72t9B_CEI (zuletzt geprüft 28.09.2015).

- Fillmore, Ch. (1982): Frame semantics. In: *Linguistics in the Morning Calm. Selected Papers from SICOL-1981*. Edited by the Linguistic Society of Korea. Seoul, 111-137.
- Geyken, A. (2011): Statistische Wortprofile zur schnellen Analyse der Syntagmatik in Textkorpora. In: Abel, A./ Zanin, R. (Hg.): *Korpora in Lehre und Forschung*. Bozen, 115-137.
- Geyken A./Didakowski, J./Siebert A. (2009): Generation of word profiles for large German corpora. In: Kawaguchi, Y./Minegishi, M./Durand, J. (eds.): *Corpus Analysis and Variation in Linguistics*. Tokyo, 141-157.
- Hausmann, F.J. (2007): Die Kollokationen im Rahmen der Phraseologie – Systematische und historische Darstellung. In: *Zeitschrift für Anglistik und Amerikanistik* 55, 217-234.
- Hvelplund, H./Kilgarriff, A./Lannoy, V./White, P. (2013): Augmenting online dictionary entries with corpus data for Search Engine Optimisation. In: *Proceedings of eLex 2013*. Tallinn, 66-75. <http://eki.ee/elex2013/conf-proceedings/> (zuletzt geprüft am 28.09.2015).
- Ivanova, K./Heid, U./Schulte im Walde, S./Kilgarriff, A./Pomikálek, J. (2008): Evaluating a German Sketch Grammar: A Case Study on Noun Phrase Case. *Proceedings of the 6th Conference on Language Resources and Evaluation*. Marrakech, 2101-2107.
- Kilgarriff, A./Rychly, P./Smrz, P./Tugwell, D. (2004): The Sketch Engine. In: *Proceedings of the Euralex 2004*. Lorient, 105-116.
- Rychly, P. (2008): A lexicographer-friendly association score. In: Sojka, P./Horák, A. (eds.): *Proceedings of the Second Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2008*, Brno, 6-9.
- Schmidt, Th. (2008): The Kicktionary Revisited. In: Storrer, A./Geyken, A./Siebert, A./Würzner, K.-M. (Hg.): *Text Resources and Lexical Knowledge*. Berlin, 239-252.
- Schmidt, Th. (2009): The Kicktionary – A Multilingual Lexical Resource of Football Language. In: Boas, Hans C. (ed.): *Multilingual Framenets in Computational Lexicography*. New York, 101-134.
- Schmidt, Th. (2010): Der Fußballwortschatz im Kicktionary. In: *Der Deutschunterricht* 3, 17-25.
- Schmidt, Th. (2011): Datenmodelle und Datenformate für die Modellierung des Fußballwortschatzes im Kicktionary. In: *OPAL – Online publizierte Arbeiten zur Linguistik* 2/2011, 53-60.
- Seelbach, D. (2001). Das kleine multilinguale Fußball-Lexikon. In: Bisang, W./ Schmidt, G. (Hg.): *Philologica et Linguistica. Historia, Pluralitas*, Universitas. Trier.

7. Anhang: „Fußballfloskeln wörtlich genommen“

Phrase/Lexem	Kicktionary	Wort-profil
Staubsauger (vor der Abwehr)	0	1
(Mannschaften) tasten sich ab	0	1
Räume im Mittelfeld eng machen	1	1
sich auf den Füßen stehen	0	0
sich das (runde) Leder schnappen	0	1
(runde) Leder	0	1
Notbremse ziehen	0	1
die Mauer dirigieren	0	1
Ball freigeben	0	1
Kiste sauber halten	0	1
Ecke treten	1	1
Ecke verwandeln	1	1
Ball aus dem Netz fischen	0	1
Abwehr schläft	0	1
in der Schaltzentrale (die Bälle verteilen)	1	1
(den Verteidiger) austanzen	1	1
(dem Verteidiger) Knoten in die Beine spielen	0	1
(Verteidiger) klebt am Angreifer	0	0
es klingelt im Gehäuse	0	1
(Torwart/Spieler) sah bei dieser Aktion nicht gut aus	0	0
Erleichterung steht (Trainer) ins Gesicht geschrieben	0	0
jdm. einen Pass schicken	0	0
Pass annehmen	1	1
Hand im Spiel sein	0	1
blinder Schiedsrichter	0	1
Tomaten auf den Augen haben	0	1
den Ball ins Tor zaubern	0	1
hinter sich greifen müssen (von Torwart)	0	1
(Trainer) wäscht (Spielern) den Kopf (i.d. Pause)	0	1
eine Schippe drauflegen	0	1
Abwehrkette (verschieben)	0	0
eine Lücke in der Abwehrkette entdecken	0	0
ein Bein stehen lassen	0	0
das Ding reinhauen	0	1
Spieler/Mannschaft kommt nicht aus dem (eigenen) Strafraum heraus	0	0
alles nach vorne werfen	0	0
(die linke/rechte) Seite beackern	0	1
ein flacher Ball	0	1
Chaos im Strafraum	0	0
den Ball (trocken) abstauben	0	1
(unter die Torlatte) nageln	0	1
Summe:	7	28