

Anglicisms and German near-synonyms.

What lexical co-occurrence reveals about their meanings

Heike Baeskow, Bergische Universität Wuppertal & Jürgen Rolshoven, Universität zu Köln

In this article, subtle semantic and connotative contrasts between the well-established anglicisms Dealer, Drink, Song, Skyline and shoppen and German near-synonyms will be identified for the first time on the basis of their distribution in large quantities of text. Co-occurrence matrices automatically generated from three journalistic corpora and two web-based corpora will show that despite a certain degree of semantic overlap, the anglicisms convey specific meaning components or connotative nuances which make them suitable for semantic and/or connotative differentiation. Semantic information retrieved from co-occurrence profiles will be mapped onto qualia structures, which are part of Pustejovsky's (1996) Generative Lexicon and help to model the generic knowledge associated with the referents of lexical items. Connotations, which are not part of the semantic representation, are supposed to comprise stylistic, emotional and communicative-pragmatic information as well as the Nebensinn ('by-sense') of a lexical item.

Keywords: *anglicisms, contrastive analyses, lexical co-occurrence, corpus linguistics, qualia structures, connotations*

1. Introduction

It is a well-known fact that borrowing is driven not only by naming need or the need to fill lexical gaps in the receptor language. Quite frequently, a loan word displays a meaning component or a connotative nuance that is not inherent to a semantically similar native word and thus allows for lexical differentiation. Anglicisms in German are a case in point. For example, while lexical items like *Quad*, *beamen* or *smart* were required to denote concepts for which German lacked adequate expressions, loan words such as *Drink*, *Song* or *shoppen* semantically correspond with *Getränk*, *Lied*, and *einkaufen*. Nevertheless, these pairs are not completely synonymous because there are very subtle contrasts.¹

Although there is a vast amount of literature on anglicisms in German, contrastive analyses are restricted to the studies of Yang (1990), Kettemann (2006) and Onysko & Winter-Froemel (2011) so far, and apart from Yang's attempt to describe semantic contrasts in terms of word-field theory, a theoretical framework is missing (cf. Baeskow (2018) for an overview of these studies). Moreover, computer-based corpus analyses are still at a very early stage in this field. While English loan words and the contexts in which they occur had to be manually collected and saved on filing cards by the time Yang wrote his thesis on anglicisms in the German newsmagazine *Der Spiegel*, Kettemann's analyses benefit from COSMAS II, a linguistic corpus provided by the IDS Mannheim, whose concordances allow him to describe the distribution of *cool*, *shoppen*, *Event* and their native near-equivalents *kühl*, *einkaufen* and *Ereignis* in German newspapers. The analyses performed by Onysko & Winter-Froemel (2011),

¹ As pointed out for example by Weisgerber (1962: 167), complete synonymy is generally avoided.

which aim at a pragmatic classification of anglicisms in German, are based on *Spiegel 2000* – an electronic resource comprising *Spiegel* editions from 1994 up to 2000.² While these studies – like the present article – focus on anglicisms and German near-synonyms, Sosnizka (2014) compares collocational patterns of the source and target language in electronically accessible German and American business and news magazines by using *WordSmith Tools*. Her analyses follow the tradition of British contextualism (e.g. Firth 1957, Jones & Sinclair 1974).

The aim of this study is to present results of innovative contrastive analyses performed on the basis of lexical co-occurrence in large quantities of text. The database for this study, which will be described in detail in section 2, comprises three journalistic corpora (the *Spiegel*-, *Focus*- and *Zeit-Archiv*) and two web-based corpora (Twitter and the German version of *WaCky*). Proceeding from co-occurrence matrices generated from these corpora for selected anglicisms and potential German equivalents, it will be argued that semantically relevant lexical items which are among the 30 most frequent co-occurrences³ of a key word in at least two corpora used for this study reflect some generic knowledge we associate with the concept the key word denotes and thus help to identify lexical contrasts as well as semantic overlap. Furthermore, it will be shown that despite a certain degree of semantic overlap, the anglicisms convey specific meaning components or shades of meaning and thus allow for semantic and/or connotative differentiation. The search for semantically relevant information in the co-occurrence matrices was guided by the notion of *qualia structure* (henceforth abbreviated as QS), which is an essential component of Pustejovsky's (1996) Generative Lexicon.

The article is structured as follows: Section 2 provides an overview of the Generative Lexicon and the connotative framework. In section 3, the corpora as well as the methods developed for the contrastive analyses will be introduced, and it will be shown that qualia structures are a suitable device for information retrieval. In section 4, representative contrastive analyses of *Dealer vs. Händler*, *Drink vs. Getränk*, *Song vs. Lied*, *Chanson, Arie, Skyline vs. Stadtsilhouette* and *shoppen vs. einkaufen* will be presented. The article ends with a conclusion in section 6.

2. The theoretical framework

Referring to the Aristotelian philosophy, Moravcsik (1975) emphasizes that the ideas we have about objects, processes etc. are based on understanding, which goes beyond mathematical propositions and generic knowledge. Understanding a concept involves understanding that it is part of the extra-linguistic reality, where it interacts with other concepts. Moravcsik's (1981) account of understanding is intensional and suggests that

² This reference corpus is also used by Onysko (2007) for a comprehensive analysis of anglicisms in German.

³ This notion will be applied here to any of the 30 lexical items that occur most frequently in the context of a given key word, i.e. of an anglicism or its German near-synonym. Given this rather broad conception of the linguistic context, the much and controversially discussed term 'collocation' (cf. Sosnizka 2014, chap. 2 for a historical overview) will be restricted to fixed collocations, i.e. to unpredictable combinations of words which have to be learned and memorized as units and which affect the levels of both syntax and word-formation, e.g. *Er raucht stark* "He smokes heavily", *ein starker Raucher* "a heavy smoker" (Lipka 2002: 182f).

the capacity to identify extensions is not part of the speakers' linguistic competence. For example, although speakers generally know the concept DISEASE, they are not necessarily able to identify instances of it without consulting a physician. While the extension remains constant, the intensional representation of concepts is incomplete, or, more precisely, it allows for variation. In particular, it is enriched from childhood to adolescence, and further change may be due to new scientific insights. Just like an adult's intensional representations of concepts differ from those of a child, the knowledge of an expert is more specific than that of the layperson. Variation as to intensional knowledge is possible even among different generations of scientists. Nevertheless, conceptual representations that differ in specificity are not mutually exclusive, but complementary, and incompleteness does not necessarily constitute an obstacle to communication.

2.1 *From Aristotelian aitiaie to qualia structures*

In the Aristotelian philosophy, understanding a concept means knowing its *aitiaie*, i.e. the four generative factors which define its distinctive properties, constituency, function, and coming into existence. These factors provide a common scheme for the representation of intensional knowledge – independently of its specificity. The essence of the aitiational framework is that semantic competence is measured not by the ability to distinguish an entity *x* from all other entities in the world, but by the degree to which an individual is able to distinguish the entity *x* from some family of elements *y, z, w* – i.e. from elements of the same semantic field. Thus, semantic competence involves for example the ability to distinguish bees from wasps, hornets etc. or to differentiate between verbs of motion such as *run, walk, or gallop* (Moravcsik 1981: 22–23). This assumption makes the aitiational approach a solid basis for contrastive analyses.

In Pustejovsky's *Generative Lexicon* (henceforth abbreviated as GL), the ensemble of generative factors, which are referred to as 'modes of explanation' (1996:76), makes up the qualia structure of a concept.⁴

- FORMAL: that which distinguishes it in a larger domain;
- CONSTITUTIVE: the relation between an object and its constituent parts;
- TELIC: its purpose and function;
- AGENTIVE: factors involved in its origin or "bringing it about".

Qualia structures, which provide a multi-dimensional representation of concepts in the form of well-defined types and relational structures (Pustejovsky 1996: 78), interact with three other levels of representation: an argument structure, an (extended) event structure, and a lexical inheritance structure. As far as argument structures are concerned, Pustejovsky (1996: 63–64) distinguishes four types of arguments, namely TRUE ARGUMENTS, which are syntactically realized, DEFAULT ARGUMENTS, i.e. "parameters which participate in the logical expressions in the qualia, but which are not

⁴ It should be noted that qualia structures do not necessarily specify all of these generative factors. For example, since natural types such as stone or water are not inherently associated with a function, the TELIC quale remains unspecified (unless some purpose or intention is being imposed on these concepts); cf. Pustejovsky (2003: 375–379).

necessarily expressed syntactically”, e.g. *John built a house out of bricks* (1996: 63), SHADOW ARGUMENTS, which are semantically incorporated into the lexical item (e.g. *to butter*), and TRUE ADJUNCTS, which allow for temporal or spatial modification and which are not part of a lexical item’s semantic representation.

As pointed out already by Moravcsik, qualia structures, which define the prototypical meaning components of concepts, constitute a suitable device for representing fine-grained contrasts between semantically similar lexical items. For example, according to Pustejovsky (1996: 77–79), the English nouns *novel* and *dictionary* are semantically related (i.e. both objects are books in a general sense). Nevertheless, they differ not only with respect to their content, but also with respect to their purpose (we *read* a novel, but we *consult* a dictionary) and their coming into being (i.e. a novel is *written*, whereas a dictionary is *compiled*). These differences, which are reflected by collocational patterns, are represented at the CONSTITUTIVE, TELIC and AGENTIVE quale respectively, as illustrated below.

$$(1) \left(\begin{array}{l} \mathbf{novel} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = y: \text{physical_object} \\ \text{D-ARG1} = x: \text{human} \\ \text{D-ARG2} = z: \text{human} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = y \\ \text{CONSTITUTIVE} = \text{narrative (y)} \\ \text{TELIC} = \text{read (e, x, y)} \\ \text{AGENTIVE} = \text{write (e, z, y)} \end{array} \right] \end{array} \right)$$

$$(2) \left(\begin{array}{l} \mathbf{dictionary} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = y: \text{physical_object} \\ \text{D-ARG1} = x: \text{human} \\ \text{D-ARG2} = z: \text{human} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = y \\ \text{CONSTITUTIVE} = \text{information (y)} \\ \text{TELIC} = \text{consult (e, x, y)} \\ \text{AGENTIVE} = \text{compile (e, z, y)} \end{array} \right] \end{array} \right)$$

Both *novel* and *dictionary* have an argument (y) in their argument structure that is of the ontological type ‘physical_object’ and corresponds to the referential argument <R> introduced by Williams (1981) for nouns. The argument structures additionally display

two default arguments D-ARG1 and D-ARG2 of the type ‘human’, whose variables (x) and (z), like (y), are bound in the qualia structures. In (1), (x) is related to a reading event which defines the function of the concept NOVEL, whereas (z) actively participates in the writing event that causes the coming into existence of (y). Thus, (x) and (z) correspond to the reader and the author respectively. The structure of the text associated with novel is specified in the CONSTITUTIVE quale. The qualia value ‘narrative’ contrasts with the value ‘information’, which defines the internal structure of a dictionary. Further contrasts emerge from the TELIC and AGENTIVE quales, where (x) is the agent of a consulting event and (z) actively participates in an event of compiling (y).

Qualia structures have been designed to account for context-specific interpretations of lexical items. For example, a sentence like *Mary began the novel* is perfectly interpretable in GL theory because the contextual sense ‘began reading’ is inferable from the reading event in the TELIC quale of *novel* – the complement of the polysemous verb *to begin* (cf. (1)). This mechanism is referred to as true complement coercion by Pustejovsky (1996: 115–117).

As pointed out above, the dynamics of intensional knowledge does not generally affect the exchange of information. According to Moravcsik (1981: 24), “there is enough overlap among the situational schemes of speakers to make communication, in most cases, possible.” If speakers share information as to situational schemes (or qualia structures in Pustejovsky’s terminology), we should expect that indicators of this common intensional knowledge are retrievable from corpora.

Evidence for this assumption comes from computational linguistics and empirical ontology. An early attempt to extract qualia-related information from an electronic corpus is made by Anick & Pustejovsky (1990). Their search is based on “collocational patterns” which linguistically reflect the content of qualia structures. For example, the collocations *read a book*, *read a tape*, *use the mouse* or *with the mouse* are suggestive of the TELIC role associated with the concepts BOOK, TAPE, or MOUSE. Anick & Pustejovsky’s results were designed for information retrieval applications.

More recent works have shown that an automatic extraction of qualia relations from large corpora is possible as well. In particular, Cimiano & Wenderoth (2005), whose intention is to reveal the impact of qualia structures on natural language processes and to support the lexicographer’s work, developed a system which automatically produces the qualia structures (it acquired by learning) from the World Wide Web. Poesio & Almuhareb (2008) attempt to evaluate the relevance of attributes extracted from the Web for the description of concepts, with attributes being conceived of as essential properties of concepts as defined in AI, linguistics and philosophy – including qualia roles.

Although the aims of these works are quite different from the aims of the present article, it becomes clear that situational schemes anticipated by Moravcsik (1981: 17) “as a psychological and semantic claim” are an integral part of language use rather than theoretical constructs. The case studies in section 4 it will show that qualia-related information is also traceable in co-occurrence matrices and helps to formalize semantic contrasts between anglicisms and German near-synonyms.

2.2 Non-denotative information

Lexical items do not only have a referential function, but also convey rather heterogeneous facets of meaning which are traditionally referred to as connotations. Following Erdmann (1900), Ludwig (1986, 2002), Yang (1990: 45–46) and Fries (2007) it is assumed here that connotations comprise (a) all kinds of associations which make up the ‘Nebensinn’ (‘by-sense’) of a word, (b) its ‘Gefühlswert’ (‘emotional value’), (c) stylistic information, and (d) communicative-pragmatic information.

If associations generally arise in the context of a lexical item, i.e. if they are not formed in the minds of individual speakers, they are part of its connotative information content and thus constitute its ‘Nebensinn’ in the sense of Erdmann (1900: 82). Erdmann illustrates this by-sense for the pair *Krieger* “warrior” and *Soldat* ‘soldier’. While the former is associated with fight and battle, the latter rather evokes the image of barracks and parade ground. Additionally, the noun *Krieger* is archaic and only of historical relevance.

Emotions are defined by Fries (2007: 298) as clusters of subjective-psychological experience and motor behaviour which are encoded in linguistic or indexical signs. As far as the emotional predisposition (abbreviated as EM) is concerned, Fries (2007: 309) introduces three independent dimensions, namely (a) emotional polarity EM_{pol+}, EM_{pol-}, EM_{pol0}, (b) emotional expectation EM_{exp-}, EM_{exp+}, and (c) emotional intensity EM_{int+}, EM_{int-}. Although this classification was developed for German linguistic signs such as *leider* ‘unfortunately’, *hoffentlich* ‘hopefully’, *pfui* ‘ugh’, *Bewunderung* ‘admiration’ etc., it is also suited to describe the emotional potential of anglicisms. Emotional values are either word-inherent (e.g. *Wellness* EM_{pol+}, *stalken* EM_{pol-} “to stalk”, *Wow!* EM_{exp-}) or unfold in the context. For example, the anglicism *Crash* has a pejorative connotation in colloquial speech, where it refers to a collision, or in the jargon of Stock Exchange, but it is an emotionally neutral technical term in linguistics. Lexical items are predisposed to occur in particular domains of communication, and their use may be determined by the social norm. Lexicographically, Ludwig (1986: 187–193) distinguishes between the communicative predisposition, which determines stylistic properties at the word-level (e.g., *Gesicht* ‘face’ is stylistically neutral, whereas *Antlitz* ‘countenance’ and *Visage* ‘mug’ range above and below ‘neutral’ respectively), and communicative-pragmatic markers, which assign lexical items to specific domains of communication (e.g. ‘technical’, ‘group-specific’), signal a speaker’s or author’s emotional attitude, or predict temporal or regional restrictions for the use of a lexeme (e.g. ‘archaic’, ‘Afro-American English’).

In German, anglicisms are often used for stylistic or communicative-pragmatic reasons. Stylistic analyses of anglicisms in German have been strongly influenced by Galinsky (1963). An important notion is that of colouring (‘Kolorit’), which is comparable to timbre in music. If the author of a text uses an anglicism in order to evoke a particular setting, to characterize a certain group of speakers or depict a new trend, local, social, or technical colouring⁵ is being created (e.g. *Sheriff*, *cool*, *Wellness-Shop*). These stylistic devices enable him or her to convey a more vivid impression of his/her subject (cf. Pfitzner 1978, chapter III, Yang 1990, chapter 4, Donalies 1992:

⁵ The corresponding German notions are ‘Lokalkolorit’, ‘Sozialkolorit’, and ‘Fachkolorit’.

104–106, Leutloff 2003, Götzeler 2008: 282–289). In German youth language, anglicisms are primarily associated with social functions, which are discussed in detail by Androutsopoulos (1998, chapter 7).

3. Computerlinguistic methods and their application

As far as research on anglicisms is concerned, the studies described in this article are innovative in two ways. First, the contrastive analyses are performed within a well-established theoretical framework, and secondly, the relevant data were extracted from very large heterogeneous corpora. The texts were accessed via a searchable database (KANG) comprising ten subcorpora, which was built by Jürgen Rolshoven and his colleagues from the University of Cologne. Within the scope of a research project on anglicisms in German, co-occurrence matrices for anglicisms and semantically similar German words were automatically generated from five subcorpora – the Spiegel-, Focus- and Zeit-Archiv, Twitter, and WaCky. The Spiegel-Archiv (documented period: 1946–2015, data volume: 1.7 GB) is a collection of texts from the German newsmagazine *Der Spiegel*, which has always been a source of lexical innovation (cf. Carstensen 1965: 22, Onysko 2007: 98). A more recent newsmagazine is *Focus*, which was first published in 1993. The Focus-Archiv (data volume: 375 MB) comprises articles from the first edition up to 2015. The Zeit-Archiv (documented period: 1946–2016, data volume: 1.3 GB) is based on articles from the supra-regional German newspaper *Die Zeit*. The data for these three subcorpora were extracted using the open online access to the archive of each of the papers. Apart from the articles' raw text, the publishing year and month as well as the name of the source were saved in the corpus as meta data. The Twitter corpus (2011–2015), whose tweets were collected via the Twitter-API⁶, was kindly provided by Prof. Dr. Chris Biemann (University of Darmstadt, now University of Hamburg) and Dr. Martin Riedl (University of Darmstadt). It is regularly updated and used for further processing at the University of Cologne. After filtering it for relevant German texts it now consists of approximately ten million tweets with a data size of 3.5 GB including some minor meta data such as the publishing date. KANG also provides access to the German version of the Web-As-Corpus Kool Yinitiative (WaCky), which was designed by Marco Baroni and colleagues⁷ to crawl the web. The German WaCky corpus represents the largest part of the data (9.5 GB) used to build KANG.⁸ All the data were standardized and merged into the searchable database.

The co-occurrence matrices were generated separately for each corpus by the computer linguists from the University of Cologne. By generating co-occurrence matrices from journalistic texts and web-based corpora it was ensured that the co-occurrences are not text-type specific. Each matrix displays a hierarchy of thirty lexical items that occur most frequently in the context of (a) an anglicism and (b) a potential German equivalent in a given corpus. The following extract of the co-occurrence matrix

⁶ <https://dev.twitter.com/streaming/reference/get/statuses/sample>

⁷ http://wacky.sslmit.unibo.it/lib/exe/fetch.php?media=papers:wacky_2008.pdf

⁸ A collection of working papers on the Web as Corpus was published by Bernadini & Baroni (2006). It is online available at <http://wackybook.sslmit.unibo.it/>.

for the key word *Drink* from the Spiegel-Archiv is intended to provide a first impression.

Table 1: The first 15 co-occurrences of *Drink*

Kookkurenzen von DRINK (Spiegel-Archiv)		
	BAR	Häufigkeit: 34 Anteil: 0.06938775510204082
	SOFT	Häufigkeit: 19 Anteil: 0.03877551020408163
	ERSTEN	Häufigkeit: 18 Anteil: 0.036734693877551024
	HAND	Häufigkeit: 14 Anteil: 0.02857142857142857
	COLA	Häufigkeit: 12 Anteil: 0.024489795918367346
	UHR	Häufigkeit: 12 Anteil: 0.024489795918367346
	NEHMEN	Häufigkeit: 12 Anteil: 0.024489795918367346
	ENERGY	Häufigkeit: 12 Anteil: 0.024489795918367346
	MANN	Häufigkeit: 11 Anteil: 0.022448979591836733
	ALKOHOL	Häufigkeit: 10 Anteil: 0.02040816326530612
	WHISKY	Häufigkeit: 10 Anteil: 0.02040816326530612
	ESSEN	Häufigkeit: 10 Anteil: 0.02040816326530612
	LETZTEN	Häufigkeit: 10 Anteil: 0.02040816326530612
	ZEIT	Häufigkeit: 10 Anteil: 0.02040816326530612
	PROZENT	Häufigkeit: 9 Anteil: 0.018367346938775512
	[...]	

The co-occurrence values were determined by counting the occurrences of all the words in a certain context window around the term and sorting them by frequency. Semantically irrelevant words (stop words, i.e. functional items, as opposed to lexical items) and compounds were removed from the texts before applying the context window. The window breadth was fixed to ± 5 , i.e. the span to be considered consists of five words to the left and five words to the right of a key word. The given values are the ratios of found contexts containing the respective co-occurrence.

The ranking of the co-occurrences in terms of co-occurrence values is less relevant for the present study because it is assumed here that looking for common co-occurrences in the five corpora more efficiently contributes to the identification of concept-defining information. The basic criterion is that a word appears among the “Top 30” co-occurrences as described above, independently of its position in the matrix. As observed by Baroni & Lenci (2008), token frequency may be accidental or result from fixed expressions. Their analyses have shown for example that the fixed expression *year of the tiger* occurs much more frequently in their corpora than the pattern *tail* and *tiger*, which signals a semantic relation and allows for different realizations (e.g. *tail of the tiger*, *tigers have tails*, *tigers with tails* etc.). Thus, the analyses performed in this study are qualitative rather than quantitative.

Co-occurrence matrices deliver the raw material for knowledge representations. They do not immediately reveal recurrent patterns such as qualia relations or other pieces of lexical information, but each matrix contains clues as to the interpretation of its key word. If these clues recur in at least two co-occurrence matrices from the

corpora used for this study, we may assume that they are lexically relevant. The search for useful lexical information in the co-occurrence matrices was primarily guided by the qualia-based approach.⁹ Co-occurrences, which also include word-forms (e.g. *singt*, *sang*, *gesungen* in the case of *Lied* ‘Song’), may be suggestive of qualia relations, qualia values, ontological types, hyponyms, arguments, fixed collocations, idioms, or connotative information.

Lexical items that contribute to the definition of a key word were manually extracted from the co-occurrence matrices and make up the co-occurrence profile¹⁰ of the key word. While a co-occurrence matrix also contains irrelevant information, a co-occurrence profile is a subset of all the co-occurrences automatically identified for a key word which only contains interpretable ‘slices’ of knowledge.

Co-occurrences may be irrelevant for different reasons. For example, unlike *Bar*, *Soft* and *Energy* are irrelevant for the semantics of *Drink* (cf. Table 1) because they are part of the highly frequent compounds *Soft Drink* and *Energy Drink*. Although compounds were automatically excluded from the co-occurrence analyses, these sequences could not be prevented from entering the matrices because they do not follow the German orthographic convention to write compounds in one word (i.e. *Softdrink*, *Energydrink*). Despite their relative frequency, co-occurrences such as *ersten* ‘first’, *Uhr* ‘clock’ or *Mann* ‘man’ are irrelevant as well because they cannot be systematically related to the concept DRINK.

Co-occurrence profiles provide an appropriate starting point for the construction of knowledge representations and for the performance of contrastive analyses because they reveal contrasts as well as semantic overlap between loan words and semantically similar native equivalents. Significantly, competing near-synonyms do not imply lexical redundancy, but signal a need for semantic or connotative differentiation, which may be very subtle.

4. Contrastive analyses

Loan words frequently undergo meaning specification when they enter the receptor language. The reason is that borrowing takes place at the level of *parole*. In a concrete communicative situation, a particular meaning component of a loan word is required rather than its entire meaning spectrum. Thus, speakers of the receptor language tend to borrow a lexical unit, i.e. a form and one meaning component related to this form (cf. Yang 1990: 46, 167, Gévaudan 2002: 25, Onysko 2007: 16–17, Winter-Froemel 2011: 213). However, as observed by Carstensen (1965: 256), English loan words – especially older ones – do not preserve a static meaning in German, but rather undergo semantic change. In particular, they are subject to meaning extension, which manifests itself most obviously in compounding. In Baeskow (2018), the well-integrated loan words *Drink*, *Dealer* and *Job* were analysed as to their behaviour in the head position of hybrid N+N compounds, which were automatically extracted from the Spiegel- and

⁹ This approach also supplements generic knowledge that is not retrievable from co-occurrence matrices.

¹⁰ This notion was borrowed from Belica (2011), who uses it in a slightly different way. According to Belica, a co-occurrence profile of an object comprises all the quantitative results of a co-occurrence analysis performed for this object.

Zeit-Archiv and manually selected from Cosmas II, the linguistic corpus provided by the IDS Mannheim¹¹. It was shown that the specific meaning the head-forming nouns were associated with when they were borrowed from English may be contextually overridden by German modifiers and that meaning extension by modification coincides with a semantic approximation of the head-forming anglicisms and German near-synonyms, e.g. *Händler* ‘trader’, *Getränk* ‘drink’ and *Beruf* ‘profession’ in the case of *Dealer*, *Drink* and *Job*. A natural next step was to analyse the semantic behaviour of selected anglicisms beyond compounding. The basic question was whether there is a general tendency for anglicisms to extend their meaning in the course of time, so that the contrasts which distinguished them from their German equivalents are gradually blurred. Apart from two of the previously analysed nouns, namely *Dealer* and *Drink*, the analyses include the verbal anglicism *shoppen* examined by Kettemann (2006) in rather limited contexts, and two nominal anglicisms which have not yet been subject to contrastive analyses, namely *Song* and *Skyline*.

4.1 *Händler* vs. *Dealer*

According to the *Anglizismen-Wörterbuch* compiled by Carstensen & Busse (1993–1996)¹², the agent noun *Dealer*, which is polysemous in English, was borrowed with the very specific meaning component “someone who sells illegal drugs”.

In Baeskow (2018) it was shown that this meaning component, which is also prevalent in compounds of the type *Drogendealer*, *Rauschgiftdealer* (both meaning ‘drug dealer’) or *Heroindealer*, may be overridden by native modifiers. Examples are *Autodealer* ‘car dealer’, *Knoblauchdealer* ‘garlic dealer’, *Klingeltondealer* ‘ring-tone dealer’ or *Plattendealer* ‘record dealer’, which are not (or only metaphorically) related to the drug scene. Although compounds like these are stylistically marked so far, they suggest that the meaning of *Dealer* at least partially overlaps with the meaning of a native agent noun that has a more general meaning, namely *Händler* ‘trader’, but does this local, i.e. context-specific approximation also hold beyond compounding? Is it justified to state that *Dealer* has undergone semantic extension in the receptor language? Let us begin by looking at the co-occurrence profile generated for *Händler*, which provides significant clues as to the generic knowledge associated with this concept.

Table 2: Co-occurrence profile of *Händler*

HÄNDLER	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	9381	3208	3671	1009	10.148
Kunden	345	269	139	31	386
Kunde		50			151
verkaufen	287	86	100	16	210

¹¹ <http://www.ids-mannheim.de/cosmas2/projekt/einsteiger/was.html>

¹² This reference work, which presents anglicisms in context and provides them with precise definitions, will henceforth be abbreviated to AWB.

verkauft	163	65	75		183
Markt	214	69	88	11	153
Ware	186	65	79		183
Preis	171	80	55		142
Preise	175	67	57		108
Geschäft	203	62	56		
Geschäfte	113		50		
Geld	158	65	60		166
kaufen	114	60	58	11	197
Käufer		81	57		210

In particular, this profile reveals an activity prototypically associated with HÄNDLER. This activity is realized by the verb *verkaufen* ‘sell’ and the past participle *verkauft*, which are recurrent in the corpora. Since the activity of selling defines a professional occupation, it determines the content of the TELIC quale of HÄNDLER. Moreover, since this activity requires a human agent, the referent of this person-denoting noun must be of the type ‘Person (x)’, which is specified in the FORMAL quale.

As far as the object of transaction is concerned, four corpora provide the collective noun *Ware* ‘merchandise’, which fits the position of the internal argument (y) opened by *verkaufen*. Compounds like *Obsthändler* ‘fruiterer’, *Autohändler* ‘car dealer’, *Buchhändler* ‘bookseller’, *Antiquitätenhändler* ‘antique dealer’, *Haushaltswarenhändler* ‘hardware dealer’ and many others show that the very general qualia value ‘Ware’ allows for specification. However, since the syntactic realization of (y) is not obligatory in the context of *Händler* (in Twitter it is not among the “Top 30” co-occurrences), it is considered here to be a default argument as defined in section 2. Further co-occurrences which are logically related to the concept HÄNDLER are *Kunde(n)* ‘customer(s)’, *Markt* ‘market’, *Preis(e)* ‘price(s)’, *Geld* ‘money’, and *Geschäft(e)*. The noun *Geschäft* either refers to a shop or to a transaction business. *Markt* is polysemous, too, because it either denotes a concrete or an abstract location. In the first reading, it functions as an adjunct which spatially locates the event of selling goods associated with HÄNDLER. In its abstract reading, it refers to the interplay of supply and demand in which the trader is involved. The following sentences from WaCky illustrate certain constellations which the co-occurrences can enter with their key word (italics and translation by HB).

(3) a. *Auf dem Markt* in der Fußgängerzone *verkaufen* die *Händler* Krimskrams für wenig *Geld*.

‘In the market place in the pedestrian zone, the traders are selling odds and ends for little money.’

b. Zunächst muss ein *Händler* seine *Kunden* identifizieren.

‘First, a trader needs to identify his/her customers.’

c. Zu beiden Seiten boten *Händler* ihre *Ware* lautstark an.
 ‘Traders loudly offered their goods on both sides.’

d. Habt ihr gehandelt oder hat der *Händler* sofort den *Preis* genannt?
 ‘Did you bargain or did the trader quote the price immediately?’

(4) a. Wenn an den Wochenenden *auf dem Markt* rund 30 *Händler* ihre *Waren* anbieten, [...]
 ‘When approximately 30 traders offer their goods in the market place at weekends, [...]’

b. Zusätzlich entsteht ein neuer *Markt* für *Händler* von Emissionsberechtigungen.
 ‘Additionally, a new market for traders of emission rights is developing.’

(5) a. Viele *Händler* hatten ihre *Geschäfte* geöffnet.
 ‘Many traders had opened their shops.’

b. Bei uns trägt der *Händler* die Beweislast, mit wem er ein *Geschäft* eingegangen ist.
 ‘At our place, the trader bears the burden of proof with whom he has entered into a business.’

These findings suggest that automatically generated co-occurrence matrices do not display arbitrary information. The qualia-based framework facilitates the evaluation of co-occurrences and helps to establish semantic relations between these items and a given key word. In particular, it is a useful guide to the identification of contrasts between semantically similar key words. The following profile indicates that six co-occurrences, namely *Geld* ‘money’, *Geschäft* ‘business’, *Kunden* ‘customers’ *Ware* ‘merchandise’, *verkaufen* ‘sell’ and *verkauft* ‘sold’ are shared by the concepts HÄNDLER and DEALER. Thus, the referents of both nouns are readily identifiable as participants in a transaction event.

Table 3: Co-occurrence profile of *Dealer*

DEALER	<i>Spiegel</i>	<i>Focus</i>	<i>Zeit</i>	<i>Twitter</i>	<i>WaCky</i>
corpus freq.	1.608	454	320	316	2.386
Drogen	97	20	13	23	186
Droge			10	4	33
Polizei	71	42	17	10	115
Heroin	62	19	13		77
Stoff	65	16	8	6	44
Geld	55	9	9		41
Kokain	43	23		4	34

Junkies	40	12			29
Süchtige	30	12			
Szene	37	11			
Geschäft	36	10			
Rauschgift	35	9			27
Kunden	34	8			27
Fahnder	32	10			
Gramm	34	11			28
kaufen				6	26
Ware	28				29
verkauft				5	36
verkaufen					37

Even more importantly, Table 3 provides a reliable answer to the question whether *Dealer* underwent meaning extension beyond compounding. Apart from instances of semantic overlap, the co-occurrences suggest that *Dealer* has preserved the specific meaning component it was associated with at the time it was borrowed. Although *Ware* “merchandise” occurs in both profiles, the object of transaction is definitely restricted to drugs in the case of *Dealer*. While *Ware* is displayed only in the Spiegel-Archiv and WaCky, *Drogen* is robustly represented in each of the five corpora. In addition, there are various instantiations of this qualia value, namely *Heroin*, *Stoff* ‘dope’, *Kokain*, and *Rauschgift*. Even the customers to whom the drugs are sold are further specified by the nouns *Junkies* and *Süchtige* ‘addicts’.

Moreover, following Pustejovsky’s (1996: 229–230) distinction between ‘role defining’ nominals (e.g. *physicist*, *linguist*, *violinist*) and ‘situationally-defined’ nominals (e.g. *pedestrian*, *passenger*, *student*), it is argued here that *Händler* is interpreted generically and that *Dealer* has a specific interpretation. While *Händler* constitutes an occupational title, a person referred to as *Dealer* is identified as such only if he is engaged in selling illegal drugs. The terms ‘role-defining’ nominals and ‘situationally-defined’ nominals reflect the more established distinction between *individual-level nominals* (ILNs) and *stage-level nominals* (SLNs)¹³. While nouns of the former type define the role of an individual independently of the activity performed at the time of reference, the interpretation of nouns of the latter type require the actual performance of characteristic activities. In GL theory, this difference is accounted for by assigning the activities typically associated with ILNs and SLNs to the TELIC and the AGENTIVE quale respectively, as exemplified below for the concepts HÄNDLER and DEALER.

¹³ The distinction between stages and individuals goes back to Carlson (1977)

$$(6) \text{ QS}_{\text{HÄNDLER}} = \left[\begin{array}{l} \text{FORMAL} = x: \text{Person} \\ \text{TELISCH} = \text{verkaufen (e, x, y: Ware, z: Kunden)} \end{array} \right]$$

$$(7) \text{ QS}_{\text{DEALER}} = \left[\begin{array}{l} \text{FORMAL} = x: \text{Person} \\ \text{AGENT.} = \text{verkaufen (e, x, y: Drogen, z: Junkies)} \end{array} \right]$$

Note that the co-occurrence profile displayed in Table 3 also shapes the connotative information content of *Dealer*. The co-occurrence of this key word with nouns referring to drugs, with *Polizei* ‘police’, *Fahnder* ‘investigator’, *Szene* ‘scene’, and with the nouns *Junkies* and *Süchtige* (both of which are inherently specified for <EMpol-> in the sense of Fries 2007) suggests that the activity specified in the AGENTIVE quale is illegal. Here we are dealing with a connotation or *Nebensinn* “by-sense” in Erdmann’s (1900: 82) terminology, which is not inherent to *Händler* and which evokes negative emotions.¹⁴

The results presented in this section have shown that there is a partial overlap between the concepts associated with HÄNDLER and DEALER. Both nouns refer to transactions involving merchandise, customers and money, but the meaning of the latter is more specific because the prototypical object of transaction is specified in its qualia structure and conveys the impression of illegality (along with other co-occurrences which are not part of the profile of HÄNDLER). The contrastive analyses of *Dealer* and *Händler* in compounding (Baeskow 2018) led to similar results, but also revealed that *Dealer* is quite frequently substituted for *Händler* (and related native nouns) in the head position of N+N compounds for stylistic purposes in journalistic texts. If the qualia value ‘Drogen’ is overridden by a modifier that does not refer to drugs, but to objects of value, harmful substances or limited resources, the connotation ‘illegal’ is preserved by the compound (e.g. *Grundstücksdealer* ‘estate dealer’, *Sprengstoffdealer* ‘dealer in explosives’, *Elfenbeindealer* ‘ivory dealer’). If the modifier refers to everyday items or food, the compound is either metaphorically/humorously related to the drug scene, or the aspect of illegality is contextually suppressed (e.g. *Strumpfdealer* ‘hosier’, *Plattendealer* ‘record dealer’, *Sonnenbrillendealer* ‘dealer in sunglasses’). Beyond compounding, however, *Dealer* and *Händler* are not (yet) exchangeable, because *Dealer* is too strongly associated with the distribution of illegal drugs. The frequent occurrence of *Dealer* in the context of *Drogen* in all five corpora suggests that this anglicism largely preserved the specific meaning it displayed when it was borrowed from English.

¹⁴ Although the referent of the emotionally neutral noun *Händler* may also deal in drugs (cf. *Droghändler*, *Rauschgift Händler*), the range of goods to be distributed is in principle infinite.

4.2 *Getränk* vs. *Drink*

Like *Dealer*, the noun *Drink* was subject to meaning specification when it entered the German language in the 19th century. According to Stiven (1936: 71), it was exclusively used to refer to alcoholic beverages. In the AWB, it is primarily defined as “alkoholisches (Misch-)Getränk” (‘(mixed) alcoholic beverage’). Furthermore, it is pointed out that this anglicism extended its meaning to denote not only alcoholic beverages, but also “(Misch-)Getränke jeglicher Art” (‘all kinds of (mixed) beverages’). The contexts provided in this reference work suggest that instances of meaning extension in compounding (e.g. *Cola-Drinks*, *Suppen-Drink* ‘soup drink’, both 1978) and beyond are not attested before the late 1970s. In Baeskow (2018), new trends as to the use of *Drink* are identified. First, *Drink* has become a vogue word in the context of modifiers related to wellness and health, where it unfolds a pseudo-medicinal flavour that is not conveyed by its German near-synonyms (e.g. *Darmgesundheitsdrink* ‘intestinal-health drink’, *Fett-weg-Drink* ‘fat-away drink’, *Herzdrink* ‘heart drink’). Secondly, *Drink* is preferred over *Getränk* (and similar near-synonyms) if the beverage is prepared by mixing the (alcoholic or non-alcoholic) ingredients denoted by the modifier. The mixing event, which is represented as an option in the AWB, is becoming more salient in view of the numerous innovative compounds whose modifiers refer to natural ingredients (e.g. *Erdbeer-Bananen-Drink* ‘strawberry-banana drink’, *Kiwi-Apfel-Drink* ‘kiwi-apple drink’, *Karotten-Brokkoli-Drink* ‘carrot-broccoli drink’). In these compounds and in contexts like those in (8), which are attested even before the 1970s, semantic approximation of *Drink* and the native noun *Getränk* is most evident.

- (8) a. Sie verabreichte siebzehn Kraftfahrern *Drinks*. Acht erhielten reinen Fruchtsaft, die anderen ein alkoholisches Getränk. (*Zeit-Archiv*, 6/1952)
‘She gave drinks to seventeen drivers. Eight of them got pure fruit juice, the others an alcoholic drink.’
- b. So nahm er in der Milchbar des Bundeshauses einen *Drink* mit einem Parlamentsportier [...] (*Spiegel-Archiv*, 4/1956)
‘Thus, he had a drink with the parliamentary porter in the milk bar of the federal parliament building [...]’

In (9), *Mineralwasser* ‘mineral water’ is obviously referred to as *Drink* for stylistic purposes. If a perfectly appropriate native noun (in this case *Getränk*) is contextually replaced by an anglicism in order to render an everyday concept more attractive or prestigious, we are dealing with an instance of semantic-stylistic upgrading in the sense of Pfitzner (1978: 194–195). In journalese, this stylistic device may be accompanied by a touch of irony.

- (9) a. *Mineralwasser* ist der absolute *In-Drink* der Saison. - Dem Durstigen, der im Café des “Steigenberger Parkhotels” zu Hamburg ein Mineralwasser bestellt, reicht der Ober eine Karte: die Wasserkarte. (*Focus-Archiv*, 8/1994)
‘Mineral water is the absolute in-drink of the season. - The waiter passes a card to the thirsty guest who orders mineral water at the café of the ‘Steigenberger Parkhotel’ in Hamburg: the water menu.’

- b. Schon aus Solidarität mit Daniel Schreiber kann ich zu beiden Büchern nur einen *Drink* empfehlen: *Mineralwasser*. Sage niemand, das sei fantasietötend: Es gibt Mineralwasser zu äußerst fantasievollen Preisen. (Fokus-Archiv, 2/2015)
 ‘If only in solidarity with Daniel Schreiber I can recommend just one drink for both books: Mineral water. Don’t tell anyone this is killing fantasy. There is mineral water available for most fanciful prices.’

Despite these instances of semantic approximation, the co-occurrence matrices of both nouns, from which compounds were excluded, indicate that the basic contrast is still prevalent. The co-occurrence matrix generated for *Getränk* gave rise to the following profile:

Table 4: Co-occurrence profile of *Getränk*

GETRÄNK	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
Cola	47	6	26	10	74
Coca	28		15		
Bier	29	13	27	21	140
Wasser	25	10	16	27	52
Kaffee	22	10	23	19	94
Tee	8		11	7	113
Wein	25	16	18		153
Whisky	19		14		
Champagner	9		11		
Alkohol	16	4	8		
alkoholisches		5		14	
Flasche	22		13		
Glas	14	6	8		70
trinken	12	8	16	23	105
getrunken	8				

To begin with, *Getränk* is of the ontological type ‘Flüssigkeit’ (‘liquid’). This very general piece of information is not directly retrievable from the co-occurrence profile, but it is reflected by the co-hyponyms *Cola*, *Bier* ‘beer’, *Wasser* ‘water’, *Kaffee* ‘coffee’, *Tee* ‘tea’, *Wein* ‘wine’, *Alkohol*, *alkoholisches* ‘alcoholic’, *Whisky*, and *Champagner* ‘champagne’. Further support for the typing of *Getränk* as ‘Flüssigkeit’ comes from the co-occurrence partners *Flasche* ‘bottle’ and *Glas* ‘glass’, which denote

containers for keeping liquids in. The function typically associated with *Getränk*, namely *trinken* ‘to drink’, is directly retrievable from the co-occurrence profile, which displays the infinitive as well as the word forms *trank* [preterite] and *getrunken* [past participle]. Significantly, the co-occurrences indicate that *Getränk* refers to all kinds of beverages, including those containing alcohol. Thus, the constituency remains unspecified in its QS.

$$(10) \left(\begin{array}{l} \mathbf{Getränk} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = y: \text{Flüssigkeit} \\ \text{D-ARG1} = x: \text{Person} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = y \\ \text{TELISCH} = \text{trinken} (e, x, y) \end{array} \right] \end{array} \right)$$

As expected, the co-occurrence profile of *Drink* is very similar to that of *Getränk*. In particular, the referents of both nouns are of the same ontological type and share a common *telos*. Moreover, there are instances of *Whisky*, *Cola* and *Coca* in both profiles. In spite of these semantic similarities, subtle contrasts between *Getränk* and *Drink* become apparent. Consider the following co-occurrence profile:

Table 5: Co-occurrence profile of *Drink*

DRINK	<i>Spiegel</i>	<i>Focus</i>	<i>Zeit</i>	<i>Twitter</i>	<i>WaCky</i>
corp. freq	490	234	350	613	2.837
Bar	34	15	33	8	161
nehmen	12	6	26		81
nimmt	8		7		42
trinken				14	38
trinke				7	
getrunken				10	
Glas		8			64
Hand	14	7			43
Alkohol	10	5			43
Whisky	10		8		

Cola	12		9		
Coca			8		
abends	8		7		
Abend		5		8	55
Wodka	7		7		

This profile suggests that *Drink* typically refers to alcoholic beverages (beyond compounding). First of all, co-hyponyms like *Kaffee*, *Wasser* or *Tee* identified for *Getränk*, which denote common thirst quenchers, do not surface among the 30 most frequent co-occurrence partners of *Drink*. Secondly, this anglicism occurs in the context of the noun *Bar* in each of the corpora used for this study. Since it is part of the generic knowledge that a bar is a location where alcoholic beverages are served (unless the reference is restricted by a modifier, as in the compound *Milchbar* ‘milk bar’ (cf. (8b)), this relatively stable co-occurrence indicates that {Alkohol} continues to be a defining property which basically distinguishes *Drink* from its native competitor *Getränk*. Thirdly, the function associated with *Drink* (TELIC) is lexically realized by word forms of *nehmen* (rather than *trinken*). In German, the phrase *einen Drink nehmen* (literally ‘to take a drink’) is a fixed collocation in which *Drink* refers to an alcoholic beverage. Phrases like *?einen Softdrink/Energydrink/Frucht-drink nehmen* are definitely marked.

(11)

$$\left(\begin{array}{l} \mathbf{Drink} \\ \text{ARGSTR} = \left[\begin{array}{l} \text{ARG1} = y: \text{Flüssigkeit} \\ \text{D-ARG1} = x: \text{Person} \end{array} \right] \\ \text{QUALIA} = \left[\begin{array}{l} \text{FORMAL} = y \\ \text{KONST.} = \{\text{Alkohol}\} \\ \text{TELISCH} = \text{nehmen (e,x,y)} \end{array} \right] \end{array} \right)$$

Further collocations in which *Drink* exclusively refers to alcohol are *auf einen Drink* ‘for a drink’ and *bei einem Drink* ‘with a drink’. Since stop words (i.e. function words) were excluded from the automatic co-occurrence analyses, these constructions are not retrievable from Table 5, but they are listed in the AWB (entry for *Drink*), where they are associated with the image of a social gathering with friends or business partners at which alcoholic beverages are enjoyed. Similarly, a relaxing, informal atmosphere is evoked by the fixed collocation *mit einem Drink in der Hand* ‘with a drink in his/her hand’, for which there is an obvious indicator in the co-occurrence profile, namely the noun *Hand* (Spiegel, Focus, and WaCky). Attention should also be paid to the adverbial *abends* ‘in the evening’ and the noun *Abend* ‘evening’, which are to be found among

the co-occurrences of *Drink* in five corpora. Since these temporal relations are absent from the profile of *Getränk*, we may conclude that the social gathering typically (though not exclusively) takes place in the evening when work is done. Concrete examples are provided below:

- (12)a. Wenn er *abends* ausgehen will, hat er die Wahl zwischen 25 Klubs, in denen er seinen *Drink* nehmen kann. (Spiegel-Archiv, 5/1948)
'When he wants to go out in the evening, he may chose among 25 clubs to have his drink.'
- b. Er lädt sie zum Lunch und unterhält sie *des Abends* bei festlichem Dinner und *Drinks* vorm Kamin [...] (Spiegel-Archiv, 9/1987)
'He invites them for lunch, and in the evening he entertains them with a festive dinner and drinks in front of the fireplace.'
- c. Ein frischer, angenehm aufmunternder *Drink* für den späten *Abend*. "Zur Bar-Kultur", sagt Daun, "gehört das richtige Getränk zur richtigen Situation." (Focus-Archiv, 7/2015)
'A fresh, pleasantly encouraging drink for the late evening. "The right drink for the right situation makes up the bar culture", says Daun.'

To sum up, the co-occurrence profile generated for *Drink* from five extensive corpora suggests that this anglicism preserved its reference to alcoholic beverages in German. However, the qualia value {Alkohol}, which distinguishes it from *Getränk*, does not have absolute character, but constitutes a prototypical semantic property which may be overridden by a modifier (e.g. *Frucht-Drink*) or the larger context (cf. (8) and (9)). In spite of the partial semantic overlap of *Drink* and *Getränk*, the anglicism provides connotative information which is not inherent to its native competitor.

4.3 *Lied* vs. *Song*, *Chanson*, *Arie*

Lexical co-occurrence is a suitable device not only for comparing word pairs, but also helps to identify contrasts and similarities between members of a lexical field. For example, the noun *Musik* 'music' and the Verb *singen* 'to sing' (including various word-forms such as *singt* or *sang*) are frequently represented in the context of the anglicism *Song*. These lexical items also co-occur with at least three other key words, namely with the native noun *Lied* (which corresponds with English *Song*) and two further loanwords – *Chanson* and *Arie* 'aria'. Each of these nouns is of the ontological type 'Musikstück (y)' 'piece of music' (FORMAL), and in each case, 'singen (e, x: human, y)' determines the TELIC quale. However, despite the semantic overlap, *Lied*, *Song*, *Chanson* and *Arie* are not freely exchangeable. Indicators of subtle contrasts are provided by the individual co-occurrence profiles. Let us begin by comparing *Song* with the more general native noun *Lied*.

Table 6: Co-occurrence profile of *Song*

SONG	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	3.162	1.444	1.706	7.380	10.382
Lied	94	34	44		200
Album	70	61	61	211	418
Band	115	59	44		381
Musik	74	46	43		285
Pop	60	29			
Rock	66		26		149
schreiben	47				
geschrieben	54	37	26		158
singen	72	30	27	85	
singt	63	29	48		
sing		32		97	
Sänger	47		27		
Platte	54		35		122
heißt	110	30	55	86	161
hören			33	144	182
gehört				132	138
hört			27	79	

Table 7: Co-occurrence profile of *Lied*

LIED	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	1.668	1.367	2.485	5.992	875
singen	122	569	320	195	62
singt	60	39	116	103	20
sang	57		54		13

sangen	36		46		
gesungen	48		75		14
Musik	26	23	46		23
Text	29		35		17
alte	35	20	90		
alten					14
heißt		22	70	156	15
Melodie	27		37		20

The first noteworthy contrast revealed by these profiles is related to the coming-into-being process, which is defined at the AGENTIVE quale. In addition to the 47 occurrences of the infinitive *schreiben* ‘to write’ in the Spiegel-Archiv, the past participle *geschrieben* ‘written’ systematically occurs in the context of *Song*. It is missing only in Twitter. Although *Lied* collocates with *schreiben*, too (*ein Lied schreiben*), this verb is not among the thirty most frequent co-occurrences of the native near-equivalent. This contrast allows for the conclusion that the concept denoted by *Song* is more readily associated with the composer (or *songwriter*, which is another anglicism in German) than the concept denoted by *Lied*. Similarly, the relatively frequent occurrence of *Band* in the context of *Song* indicates that the anglicism is more closely related to the performers than *Lied* (hence *ein Beatles-Song*, *ein Abba-Song*, *ein Song von den Stones* etc.).

Further co-occurrences which allow for a distinction between *Song* and *Lied* are *Pop* and *Rock*. These items do not contribute to the semantic structure of *Song*, but rather provide connotative information because they suggest that we are dealing here with a modern piece of music of anglo-american origin or influence, which may be of international popularity.¹⁵ Like *Band*, *Pop* and *Rock*, the noun *Song* found its way into other languages, too. Thus, each of these anglicisms has a high international recognition value. On the other hand, *Lied* co-occurs with two word forms of the adjective *alt* ‘old’, namely *alten* and *alte*, in four corpora. Depending on the linguistic context, the resulting phrase either receives a literal or a metaphorical interpretation. If a piece of music for singing is old in the literal sense, it is referred to as *Lied* rather than *Song* in German, as shown below:

- (13)a. Und auch sonst stimmt am *alten Lied* vom braven Landmann kaum noch eine Verszeile. (Spiegel-Archiv, 11/1977)
‘And in other respects, too, hardly any verse line of the old song of the brave countryman is still correct.’

¹⁵ Of course, song texts are not restricted to English. In KANG, the anglicism *Song* frequently occurs in the context of names of German bands or musicians.

- b. Und da war Amir so gerührt, daß er auf der Stelle “Ich schwebe über den Bäumen” sang, ein *altes* arabisches *Lied*. (Spiegel-Archiv, 1/1996)
 ‘And then Amir was so strongly moved that he instantaneously sang ‘I am floating above the trees’, an old Arabian song.’

A metaphorical interpretation is required for the idiom *Es ist das alte Lied* ‘It’s the (same) old story’, which means that a well-known problem recurs frequently. In this context, *Lied* is not replaceable by *Song* (*Es ist *der alte Song*).

The use of *Song* may also be motivated by communicative-pragmatic and stylistic considerations. In the Twitter-Korpus, it frequently occurs in the context of modifiers or intensifiers that are typical of German youth language (e.g. *krass* ‘wicked, cool, awesome’ in (14a), *Hammer*¹⁶ in (14b) and *cool* in (14c)), and quite a few utterances are accompanied by emoticons, e.g.

- (14)a. Noch nie erlebt, dass jemand so krass einen *Song* angeteasert hat 😊 aber hat sich gelohnt (Twitter 8/2015)
 ‘I never heard someone tease a song so wickedly 😊 but it was worthwhile.’
- b. der [sic!] *song* dein herz trägt felsen fand ich wahnsinnig gut. der text ist toll und der refrain ist *hammer!* (Twitter 2012)
 ‘I considered the song ‘Dein Herz trägt Felsen’ [‘Your heart is bearing rocks’] to be incredibly good. The text is great and the refrain is awesome.’
- c. bitte nehmt Jedward’s [HASHTAG293601174] in eure playlist auf! das ist so ein *cooler song* und verdient echt eine chance! (Twitter 11/2012)
 ‘Please add Jedward’s to your playlist! It’s such a cool song and really deserves a chance!’

Tweets like these signal that young speakers in particular consider *Song* to be less conventional or even more prestigious than *Lied*. This is what Pfitzner (1978) refers to as ‘affect’, which typically manifests itself in semantic-stylistic upgrading mentioned already in section 4.2.¹⁷ Furthermore, it is noteworthy that *Song* also surfaces in a collocation in which *Lied* is less usual, namely *einen Song performen* ‘to perform a song’. While *Lied* is absent from the thirty most frequent co-occurrences of *performen*, *Song* occurs in the context of this verbal anglicism in three corpora:

¹⁶ The idiom *Das ist/war der Hammer* either refers to something extremely positive or extremely negative.

¹⁷ According to Pfitzner (1978: 211), the use of anglicisms to create semantic-stylistic downgrading is relatively restricted. In Author1 (2018), this device is exemplified by expressive compounds like *Knochenjob* ‘back-breaking job’, *Sklavenjob* ‘slave job’, *Routinejob* ‘routine job’, *Stressjob* ‘stress job’, *Höllensjob* ‘hell’s job’, *Mistjob* ‘crap job’, or *Billigjob* ‘low-paying job’ in which *Job* is usually not replaced by *Beruf* ‘profession’.

Table 8: Co-occurrence profile of *performen* (extract)

PERFORMEN	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	8	16	19	189	189
SONG			2	13	16
SONGS				6	9

A Google-based search restricted to German websites yields 2.010 hits for *einen Song performen* and only 279 hits for *ein Lied performen*.¹⁸ In this collocation, *Song* unfolds a (pseudo-)technical aura which is not conveyed by *Lied* in its basic reading ‘(short) piece of music to be sung’. On the other hand, *Lied* constitutes a real technical term used by music experts to refer to the art form developed by Franz Schubert and continued by composers like Hugo Wolf or Richard Strauss. In this particular reading, *Lied* was borrowed into other languages such as English or Spanish.¹⁹ This contrast illustrates the distinction between *Fachsprache* (‘technical terminology’) and *Fachkolorit* (‘technical colouring’). While *Lied* in the sense of ‘art form’ denotes a musicological concept, *Song* rather constitutes a vogue word that is suited to evoke the impression of expertise.

Interestingly, *Song* seems to have largely abandoned an early meaning component. According to the AWB (entry for *Song*), this loanword is first attested in 1798, but only occurred sporadically until the beginning of the 20th century, when it was used to refer to socio-critical cabaret songs in the Brechtian style and hence assumed a very specific meaning as compared to its English model. Although the compound *Protestsong* still exists in German, there are no indicators in the co-occurrence matrices generated from the five modern corpora which point towards a critical connotation. Thus, we may conclude that the narrow reading imposed on this loanword by Brecht was abandoned in favour of a more general meaning, which entails a semantic approximation of *Lied* and *Song*. However, despite this meaning extension, *Lied* is not completely synonymous with *Song* because of its connotative shift towards a vogue word in the jargon of light music.

As indicated above, semantic overlap is observable not only for *Song* and *Lied*. Consider the co-occurrence profile of *Chanson*, which belongs to the same lexical field.

Table 9: Co-occurrence profile of *Chanson*

CHANSON	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	431	62	256	22	1.352

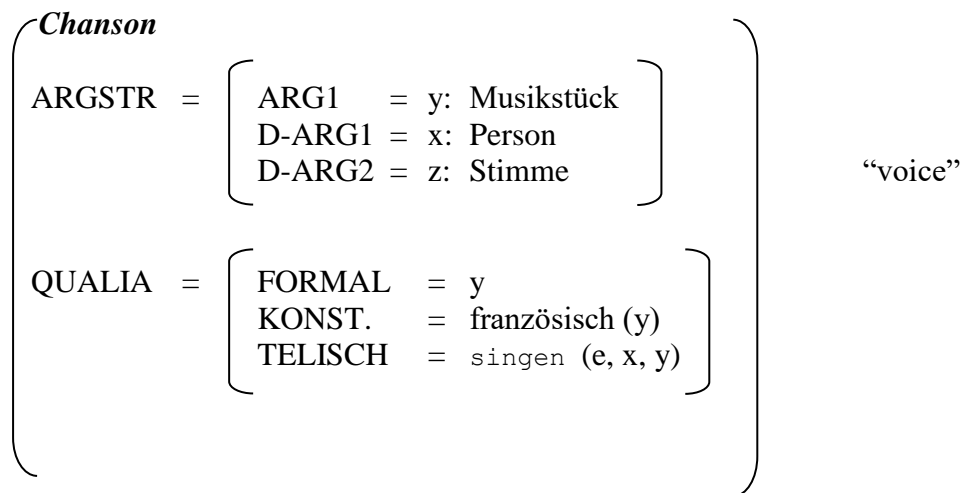
¹⁸ Last access June 20th, 2017

¹⁹ For English, *the Oxford English Dictionary* (OED) also lists the hybrid compounds *lieder-singer* and *lieder-singing*. The Corpus del Español constructed by Mark Davies even provides a Spanish agent noun *liederista*.

singen	13	3	7		
singt	13		17		
Sänger	29	3			
Sängerin	35		11		40
Song		2			38
sang			8		
gesungen			7		
französische	24	4	9		40
französischen	18		7		68
Frankreich			6		37
Lied	10		13		55
Lieder					34
Musik	13	4	10		84
deutschen	11	3			35

The profile of *Chanson* shares with the profiles of *Song* and *Lied* the co-occurrences *singen* (along with related word-forms) and *Musik*, which determine the TELIC and the FORMAL quale respectively. More importantly, however, a prototypical property of CHANSON that distinguishes this concept from the concepts SONG and LIED is retrievable from the co-occurrence profile as well, namely its French origin. Considering the fact that the overall corpus frequency of *Chanson* is relatively low (in the Spiegel-Archiv, for example, there are 3162 instances of *Song*, but only 431 instances of *Chanson*), the word-forms *französische* and *französischen* of the adjective *französisch* “French” are well represented in the context of this loanword. The Zeit-Archiv and WaCky additionally display the proper noun *Frankreich* ‘France’. Semantically, the French origin of the concept CHANSON is represented at the CONSTITUTIVE quale.

(15)



Of course, *französisch* is merely a default value, i.e. a prototypical value which may be contextually overridden. For instance, the co-occurrence profile of *Chanson* also displays the word-form *deutschen* ‘German’ even if this adjective is underrepresented. This finding does not come as a surprise because this type of song has also been cultivated in Germany at least since the Weimar Republic. However, although the *chanson* is not regionally restricted to France, it is definitely inspired by a French model (16a) and thus suitable to create local colouring (16b).

(16)a. Aznavour: Nein! Das französische *Chanson* ist ein typisch französisches Produkt, weil es die Liebe zur französischen Sprache ist, die diese Autoren [Gainsbourg, Moustaki, Piaf etc.; HB] erst hervorgebracht hat. (Zeit-Archiv 5/2014)

‘Aznavour: No! The French *Chanson* is a typically French product because it is love for the French language which eventually brought forth these authors.’

b. Dann trällert die Empfangsdame ein *Chanson* – Erlebnisgastronomie im Stil der “Bouffes Anversoises – Cafés chantants” des 19. Jahrhunderts. (Zeit-Archiv, 2/2007)

‘Then the receptionist warbles a *chanson* – event gastronomy in the style of the 19th century “Bouffes Anversoises – Cafés chantants”.’

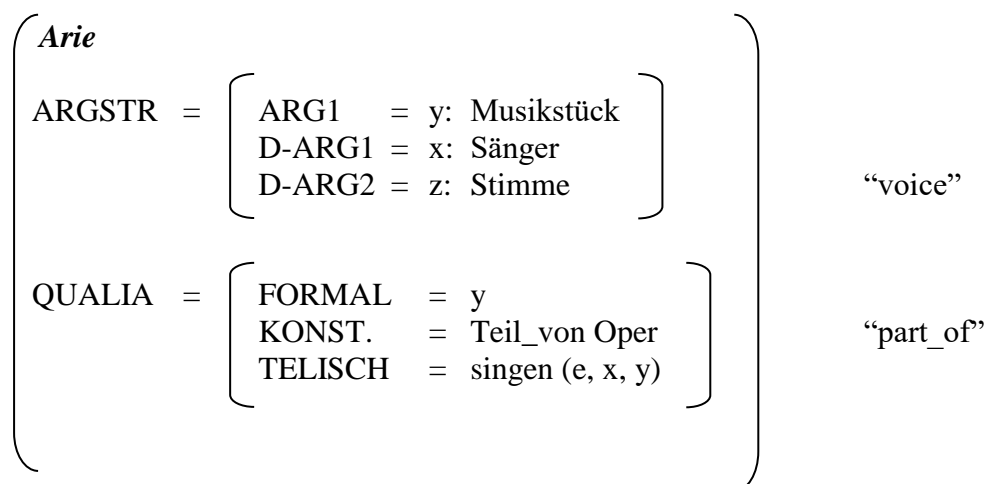
Another member of the lexical field examined in this section is *Arie* ‘aria’, whose distinctive property is also retrievable from the corpora.

Table 10: Co-occurrence profile of *Arie*

ARIE	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	300	83	341	33	190
Oper	16	5	19		6
singen	9	6	12	3	9
singt	13	4	23	1	9
sang		4		2	
Rezitativ			15		15
Musik	5	5	5		6
Bühne	7		7		

Like *Lied*, *Song* and *Chanson*, *Arie* displays the co-occurrences *singen* (as well as related word-forms) and *Musik*, which provide the most basic components for its qualia structure. One co-occurring lexical item not shared by the other field neighbours is *Oper* ‘opera’. This co-occurrence partner signals that *Arie* is a relational noun, which is typically part of an opera²⁰ – even if it is performed independently of the complete work. Following Bouillon et al. (2012: 1529), it is assumed here that part-of relations are represented at the CONSTITUTIVE quale, as shown in (17).

(17)



Further co-occurrences which are represented less frequently but nevertheless fit the profile of *Arie* are *Rezitativ* (*Zt* 15, *WaC* 15) and *Bühne* ‘stage’ (*Sp* 7, *Zt* 7). While a

²⁰ Beyond the opera, an aria is also part of a cantata or an oratorium; cf. Seeger, H. (1966): *Musiklexikon in zwei Bänden*. Leipzig: VEB Deutscher Verlag für Musik.

recitative – like an aria – is typically part of an opera (except that it is spoken instead of sung), *Bühne* is interpretable as an adjunct that spatially locates the event of singing.

While Yang’s (1990: 49–55, 92–93) pre-digital identification of semes and distinctive connotative features largely depended on speakers’ judgements and dictionary entries, the present analyses benefit from automatically performed co-occurrence analyses which provide clues as to the semantic and/or connotative information content of lexical items and thus help to reveal not only lexical contrasts and semantic overlap between members of word pairs, but also to determine the distribution of members of lexical fields. Moreover, since the co-occurrence matrices were generated from large quantities of text, they ensure that the relations and values that are eventually mapped onto qualia structures are empirically motivated.

4.4 Skyline vs. Stadtsilhouette

Although proper nouns are not part of knowledge representations, they may contribute to the connotative information content of a lexical item. As will be shown in this section, the systematic co-occurrence of a keyword with geographical names may be an indicator of local colouring (‘Lokalkolorit’). Evidence comes from the profile of the anglicism *Skyline*, which is represented below:

Table 11: Co-occurrence profile of *Skyline*

SKYLINE	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corp. freq.	299	92	205	154	1.602
New	36	11	14	9	183
York	25	6	11	6	112
Yorker	8	3		4	35
Stadt	36	12	19	4	154
Manhattan	29	9	11	3	105
Manhattans	14	9	5		35
Blick	28	13	22	10	16
Wolkenkratzer	22		7		43
Frankfurt	8	3	12	9	105
Frankfurter	16	3			78
Hochhäuser	12	6	7		28
City	10	4	7	5	45

Significantly, the noun *Skyline*, which roughly corresponds with the less frequent German-French hybrid compound *Stadtsilhouette* (literally ‘city silhouette’), co-occurs with *New, York, Yorker* and *Manhattan* in all the corpora used for this study. This constellation, which is not available for *Stadtsilhouette* (*die Skyline* /? *Stadtsilhouette von Manhattan*)²¹, locates the concept SKYLINE in a typically American setting and makes it a suitable device for local colouring – a stylistic nuance which according to Galinsky (1963: 101) conveys “the impression of American ‘atmosphere’ on the German listener’s or reader’s mind.” The following sentences from the print media illustrate this effect:

- (18)a. Am Fenster zieht die *Skyline* von Manhattan vorbei. Sein Ziel ist downtown, eine Bank hat ihn zum Lunchvortrag eingeladen. (Spiegel 4/2006)
 ‘The Skyline of Manhattan passed the window. His destination is downtown, a bank invited him for a lunchtalk.’
- b. Die *Skyline* definierte nicht nur die Stadt, sondern auch die Menschen, die zu ihren Füßen leben. Wolkenkratzer wirken plötzlich zerbrechlich. (Focus 3/2002)
 ‘The skyline defined not only the city, but also the humans living at its feet. Skyscrapers suddenly appear fragile.’

In (18a), the American atmosphere is conveyed by an accumulation of three anglicisms: *Skyline*, *downtown*, and the hybrid compound *Lunchvortrag*. In (18b), the noun *Skyline* has a strongly symbolic function because it is representative of life in Manhattan before and after 9/11. Both sentences would be less vivid and less authentic if *Skyline* was replaced by *Stadtsilhouette*.

Table 11 indicates that the connotative relation between *Skyline* and *New York* or *Manhattan* is transferable to megacities displaying a similar architecture. In particular, *Skyline* co-occurs with *Frankfurt*, whose financial district with all its skyscrapers is strongly reminiscent of Manhattan. On account of this similarity, Frankfurt is humorously referred to as *Mainhattan* (i.e. Manhattan on the Main). The names of other megacities only occur sporadically in the context of *Skyline*, as the following frequencies show: *Hongkong* (Sp 9, Zt 4), *Londoner* (Sp 6), *London* (Sp 6), *Dubai* (Foc 4), *Shanghai* (Tw 3), *Chicago* (WaC 23).

The link for the connotative transfer is provided by two common nouns which regularly co-occur with *Skyline* in the corpora, namely *Wolkenkratzer*²² and *Hochhäuser* ‘high-rise buildings’. Although *Hochhäuser* (unlike *Wolkenkratzer*) also co-occurs with *Stadtsilhouette* (Foc 2, Zt 3, WaCky 6), modern buildings of a considerable height do not seem to be salient defining components of this concept. Instead, *Stadtsilhouette* tends to co-occur with various nouns referring to or related to churches, namely *Frauenkirche* (Sp 1), *Kirchenschiff* ‘nave’ (Foc 1), *Kirchen* (Zt 1, WaC 8), and *Dom* (WaC, 8). Another co-occurrence partner is *Altstadt* “historic city” (Tw 1, WaC 6), which refers to a part of a city that usually consists of historic buildings. In the contexts provided in (19), the substitution of *Skyline* for *Stadtsilhouette* would result in stylistic markedness:

²¹ The profile of *Stadtsilhouette* will not be displayed here because this noun has a very low corpus frequency. Instead, reference will be made to co-occurrences in the individual corpora.

²² *Wolkenkratzer* (literally ‘cloud scraper’) is a loan rendition of the English noun *skyscraper*.

- (19)a. Steile Staffeln, enge Gassen und spitze Giebel prägen die *Stadtsilhouette* von der Tübinger Altstadt bis hinauf zum Schloss. (Twitter 11/2012).
 ‘Steep steps, narrow lanes and pointed gables determine the skyline of Tübingen’s historic city up to the castle.’
- b. Aufragend aus der Masse, geben einige wenige Gebäude der *Stadtsilhouette* ihren unverkennbaren Schnitt: rechts, im Osten, der mächtige Kubus des Alcázar, der Stadtfestung, in der Mitte der Turm der gotischen Kathedrale.
 ‘Looming out of the mass, a few buildings shape the unmistakable skyline: to the right, in the east, the mighty cube of the Alcázar, the town fortress, in the middle the tower of the Gothic cathedral.’

The only place names co-occurring with *Stadtsilhouette* in more than one corpus are *Hamburg* (Sp 2, WaC6) and *Köln* “Cologne” (Zt 2, WaC 5) – two German cities which lack skyscrapers of a dimension typically associated with the United States. Since recently, Hamburg’s highest building is the Elbphilharmonie (110 metres)²³, which is still relatively small in comparison with the Empire State Building (381 metres)²⁴. In Cologne, there are some high-rise buildings, too, but since its panorama is inextricably linked to the two pointed towers of the famous cathedral – the Kölner Dom – the use of *Stadtsilhouette* constitutes an option.

From a semantic point of view, however, *Skyline* and *Stadtsilhouette* are in principle exchangeable because both nouns refer to the outline of a number of buildings seen against the sky and thus exclude the shape of natural objects such as rocks, hills or trees.²⁵ As shown in this section, the contrast is rather connotatively motivated, and the systematic occurrence of *New York* and *Manhattan* in the context of *Skyline* allows for the conclusion that these proper nouns trigger the use of this anglicism.

4.5 *Shoppen vs. einkaufen*

Co-occurrence matrices also allow for a comparison of lexical items other than nouns. In this section, two near-synonyms analysed by Kettemann (2006: 177–178) will be revisited, namely *shoppen* and *einkaufen* ‘to do the shopping’. As indicated in the introduction, Kettemann performed his analyses on the basis of concordances from COSMAS II. Although the contexts were extracted non-automatically, they make interesting predictions. As far as *shoppen* is concerned, co-occurrences like *gerne* “with pleasure”, *relaxen* ‘to relax’, *Spaß* ‘fun’, *gemütlich* ‘leisurely’ or *Freudinnen treffen* ‘meet (female) friends’ suggest that this anglicism refers to a pleasant leisure activity which is not necessarily a targeted activity. In terms of the Generative Lexicon, the TELIC quale of *shoppen* remains unspecified because there is not always a concrete need for the goods to be purchased. By contrast, the co-occurrences identified by Kettemann for *einkaufen* signal that this activity is part of the household chores, which

²³ <http://www.hamburg.de/sehenswuerdigkeiten/4401242/hochhaeuser-in-hamburg/>

²⁴ https://de.wikipedia.org/wiki/Empire_State_Building

²⁵ In English, by contrast, the denotation of *skyline* also includes “the line at which the earth or a part of the landscape appears to meet the sky” (OED), e.g. *I do love horses moving slowly against a skyline of trees* (a1933). Once again, the borrowing process led to meaning specification.

have to be done quickly and efficiently. Examples are *müssen* ‘have to’ *kochen* ‘to cook’, *putzen* ‘to clean’, *günstig* ‘cheap’, or *Stress*. Kettemann’s findings were checked against the automatically detected contexts which gave rise to the following co-occurrence profiles:

Table 12: Co-occurrence profile of *einkaufen*

EINKAUFEN	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	2.210	871	1.534	4.060	14.956
gehen	148	64	146	554	1.566
Geld	87	21	52	48	443
Supermarkt	52	24	39		405
Kunden	57	37	36		339
billig(er)	85	24	74		
kochen		23	46	49	361
Arbeit			32		284
arbeiten			30		262

Table 13: Co-occurrence profile of *shoppen*

SHOPPEN	<i>Sp</i>	<i>Foc</i>	<i>Zt</i>	<i>Tw</i>	<i>WaC</i>
corpus freq.	268	297	155	2.579	2.109
gehen	31	30	20	383	353
kaufen		13	11		49
Internet	13	16	5		51
Stadt			8	49	87
gern(e)	9	10	4	45	100
Freundin(nen)			4	51	
Freunde(n)					91
Spaß				67	84
Geld	14	10	7	59	61
Klamotten			4	37	

While both verbs enter a fixed collocation with *gehen* ‘to go’ (*shoppen gehen*, *einkaufen gehen*)²⁶ and share the co-occurrence *Geld* ‘money’, only *shoppen* displays the positively connotated items *gern(e)*, *Spaß* and *Freundin(nen)*, *Freund(en)* in its co-occurrence profile. In the context of *einkaufen*, there are co-occurrences such as *billig(er)* ‘cheap(er)’, *kochen* ‘to cook’, *Arbeit* ‘work’, or *arbeiten* ‘to work’. These constellations, some of which are exemplified below, confirm Kettemann’s observation that *shoppen* – unlike *einkaufen* – refers to a pleasant and relaxing activity.

(20)a. *Mit meinen Freunden* gehe ich *gern* in die Stadt *shoppen*, ins Kino oder ein Eis essen. Das macht immer riesen *spaß* [sic!] und überhaupt keine Langeweile. (WaCky)

‘I like to go shopping, to go to the cinema or to eat an ice-cream with my friends_[+fem]. This is always great fun and never boring.’

b. Morgen wahrscheinlich *mit meiner besten Freundin* *shoppen* und stöbern bei Hugendubel. Ich freu mich jetzt schon. (Twitter, 8/2011)

‘Tomorrow probably shopping with my best friend_[+fem] and rummaging at Hugendubel’s. I am already looking forward to it.’

(21)a. Auf der anderen Seite könnte man sagen, daß ich Künstler bin und als solcher nicht meine ganze Zeit mit *Einkaufen* und *Kochen* verbringen will [...] (Focus, 6/1995)

‘On the other hand, one could say that I am an artist and thus do not want to spend all my time doing the shopping or cooking.’

b. „Wenn ich nach der *Arbeit* noch *einkaufen* muß, dann Kläuschen aus der Tagesstätte abholen und mit ihm spielen – das schaffe ich nicht“ (Zeit 1/1990)

‘If I have to do the shopping after work, to fetch Kläuschen from the day-care centre afterwards and to play with him – I don’t manage it.’

At the lexical level, this contrast is accounted for by emotional markers in the sense of Fries (2007). While *shoppen* is inherently specified for <EMpol+>, its native competitor should be considered emotionally neutral. Although *einkaufen* may be more closely related to household chores, it is in no way a pejorative verb. Following Fries, emotional neutrality will be represented by the feature <EMpol₀>.

Another revealing, though less frequent co-occurrence of *shoppen* is *Klamotten*, which is colloquial for articles of clothing. Although the goods to be purchased are not restricted to clothes, it is a matter of fact that *shoppen* is generally avoided in the context of food or domestic articles in German unless a particular stylistic effect is to be achieved. For example, the use of this anglicism in the slogan *Die ganze Woche Frische shoppen* ‘buy freshness throughout the week’, which covers the weekly advertising leaflet of a German discounter, is clearly marked and intended to attract the customers’ attention. In this case, *shoppen* is contextually assigned the connotative feature <EMexp–>, which encodes the emotion of unexpectedness.

²⁶ A slightly archaic variant listed in the AWB is *shopping gehen*.

Contrasts are also observable as to the locations in which the events denoted by *shoppen* and *einkaufen* typically take place. In the corpora, the noun *Supermarkt* “supermarket”, which is associated with a large supply of goods, self-service and special offers, only occurs in the context of *einkaufen* – the activity mainly driven by need. On the other hand, the co-occurrence profile of *shoppen* displays *Stadt* “city” and – more importantly – the noun *Internet*. Although the virtual location denoted by *Internet* is also compatible with *einkaufen* (*im Internet einkaufen*), the verb *shoppen* unfolds its positive connotation even in this context and thus implies the advantages associated with online shopping (sitting comfortably in front of the computer in order to choose a particular article independently of opening hours, using the delivery service, paying by online banking etc.).

To sum up, *shoppen* differs semantically from *einkaufen* in that it denotes an activity which is not necessarily targeted. While *einkaufen* constitutes a telic verb, *shoppen* is inherently atelic although a transitive, telic use is not generally precluded (e.g. *Klamotten*, *Frische shoppen*). From a connotative point of view, *shoppen* is associated with a pleasant activity, whereas *einkaufen* is emotionally neutral. Computerlinguistic evidence for this contrast, which was anticipated by Kettemann (2006) on the basis of a small set of data, comes from automatically identified co-occurrence partners like *gern(e)*, *Freundin(nen)*, or *Spaß*.

5. Conclusion

In this article, the anglicisms *Dealer*, *Drink*, *Song*, *Skyline* and *shoppen* were compared to semantically close German equivalents. The case studies have shown that contextual information from very large and heterogeneous corpora is a reliable basis for information retrieval and for the performance of contrastive analyses. Although we cannot expect co-occurrence matrices to provide full-fledged semantic representations or encyclopaedic details, prototypical properties of the key words under examination are made accessible to some degree. In order to avoid arbitrariness, the selection of semantically relevant co-occurrences was guided by Pustejovsky’s (1996) Generative Lexicon whose multi-dimensional qualia structures provide basic parameters for generic knowledge. These parameters account not only for contrasts between anglicisms and native near-synonyms, but also for instances of semantic approximation.

The automatically performed co-occurrence analyses also revealed lexical properties that are much more difficult to capture out of context, namely connotations, which comprise stylistic, emotional and communicative-pragmatic facets of key words as well as images we typically associate with them. Although connotative nuances are not part of semantic representations, they enrich our knowledge about concepts and allow us to work out very subtle contrasts between loanwords and semantically close native equivalents.

On the whole, the contrastive analyses have shown that despite semantic overlap, the distribution of the anglicisms under consideration and their German near-synonyms is non-arbitrary and largely predictable from their co-occurrence profiles. While *Dealer*, *Drink*, *Skyline* and *shoppen* preserved the specific meaning components with which they were borrowed from English and unfold connotations that are not conveyed by the native near-synonyms *Händler*, *Getränk*, *Stadtsilhouette* and *einkaufen* respectively,

semantic approximation is most obvious for the pair *Lied* and *Song*, the latter of which abandoned its political, socio-critical flavour. Nevertheless, these nouns do not constitute complete synonyms either because *Song* gained the status of a (pseudo-)technical vogue word which distinguishes it from *Lied* in its basic and technical reading. As predicted by Weisgerber (1962: 167), there are no instances of complete synonymy in the examples selected from KANG.

Acknowledgements

The results presented in this study were obtained during the research project “Anglizismen im Deutschen: Kontextbasierte Interpretation, dynamische Restrukturierung und Generalisierung” kindly supported by the Deutsche Forschungsgemeinschaft (DFG) and reflect the co-operation of the authors. While Jürgen Rolshoven’s contribution involves data collection, data processing and the generation of co-occurrence matrices as a type of vector semantics, Heike Baeskow is responsible for the linguistic analyses. We would like to thank Børge Kiss, Peter Seipel and Ipek Cengiz for their assistance in data processing.

References

- Androutsopoulos, Jannis. 1998. *Deutsche Jugendsprache. Untersuchungen zu ihren Strukturen und Funktionen*. Frankfurt a. M. [etc.]: Peter Lang.
- Anick, Peter & Pustejovsky, James. 1990. An application of lexical semantics to knowledge acquisition from corpora. In Karlgren, Hans (ed.), *Proceedings of the 13th International Conference on Computational Linguistics*, 7–12, University of Helsinki, Finland.
- Baeskow, Heike (2018): “-Drink or -trunk, that is the question: The distribution of selected anglicisms and German near-synonyms in compounding”. *Studia Linguistica* (Online publication ahead of print; <https://doi.org/10.1111/stul.12086>).
- Baroni, Marco & Lenci, Alessandro. 2008. Concepts and properties in word spaces. In Lenci, Alessandro (ed.), *From context to meaning. Distributional models of the lexicon in linguistics and cognitive science. Special Issue of the Italian Journal of Linguistics* 20(1). 55–88.
- Belica, Cyril. 2011. Semantische Nähe als Ähnlichkeit von Kookkurrenzprofilen. In Abel, Andrea & Zanin, Renata (eds.), *Korpora in Lehre und Forschung*, 155–178. Bozen: Bozen-Bolzano University Press. Freie Universität Bozen-Bolzano.
- Bernadini, Silvia & Baroni, Marco (eds.). 2006. *Wacky! Working papers on the Web as corpus*. Bologna: GEDIT. Available online at <http://wackybook.sslmit.unibo.it/>
- Bouillon, Pierrette, Jezek, Elisabetta, Melloni, Chiara & Picton, Aurélie. 2012. Annotating qualia relations in Italian and French complex nominals”. In Calzolari, Nicoletta, Chukri, Khalid, Declerck, Thierry, Uğur Doğan, Mehmet, Maegaard, Bente, Mariani, Joseph, Moreno, Asuncion, Odijk, Jan & Piperidis, Stelios (eds.), *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC)*, 1527–1532. Istanbul: European Languages Resources Association (Elra).

- Carlson, Gregory Norman. 1977. *Reference to Kinds in English*. Massachusetts, Amherst: University of Massachusetts. (Doctoral dissertation).
- Carstensen, Broder. 1965. *Englische Einflüsse auf die deutsche Sprache nach 1945*. Heidelberg: Carl Winter Universitätsverlag.
- Carstensen, Broder & Busse, Ulrich (eds.): *Anglizismen-Wörterbuch. Der Einfluß des Englischen auf den deutschen Wortschatz nach 1945*. Berlin, New York: De Gruyter. Volume 1 (1993), Volume 2 (1994), Volume 3 (1996); reprinted 2001.
- Cimiano, Philipp & Wenderoth, Johanna. 2005. Automatically learning qualia structures from the Web. *Proceedings of the ACL-SIGLEX Workshop on Deep Lexical Acquisition*, 28–37. Ann Arbor: Association for Computational Linguistics.
- Davies, Mark. (2002-) *Corpus del Español: 100 million words, 1200s-1900s*. Available online at <http://www.corpusdelespanol.org>.
- Donalies, Elke 1992. Hippies Hopping und toughe Trendies: Über neumodische, noch nicht kodifizierte Anglizismen in deutschsprachigen Female-Yuppie-Zeitschriften". *Deutsche Sprache: Zeitschrift für Theorie, Praxis, Dokumentation* 20 (2). 97–110.
- Erdmann, Karl-Otto. 1900. *Die Bedeutung des Wortes*. Leipzig: Eduard Avenarius.
- Firth, John R. 1957. Modes of meaning. In Firth, John R. (ed.), *Papers in Linguistics 1934–1951*, 180–215. London: Oxford University Press.
- Fries, Norbert. 2007. Die Kodierung von Emotionen in Texten. Teil 1: Grundlagen. *Journal of Literary Theory* 1(2). 293–337.
- Galinsky, Hans 1963. Stylistic aspects of borrowing. A stylistic and comparative view of American elements in Modern German and British English. In Fraenkel, Ernst, Galinsky, Hans, Gerhard, Dietrich & Lang, H.J. (eds.), *Jahrbuch für Amerikastudien*, 98–135. Band 8. Heidelberg: Universitätsverlag Winter.
- Gévaudan, Paul. 2002. *Klassifikation lexikalischer Entwicklungen. Semantische, morphologische und stratische Filiation*.
[\[http://homepages.uni-tuebingen.de/paul.gevaudan/Filiation.pdf\]](http://homepages.uni-tuebingen.de/paul.gevaudan/Filiation.pdf)
- Götzeler, Christiane. 2008. *Anglizismen in der Pressesprache. Alte und neue Bundesländer im Vergleich*. Bremen: Hempen Verlag.
- Jones, Susan & Sinclair, John M. 1974. English lexical collocations. *Cahiers de Lexicologie* 24. 15–61.
- Kettemann, Bernhard. 2006. Morphologische Integration und semantische Differenzierung durch Anglizismen im Deutschen. In Mair, C. (ed.), *Corpora and the history of English. Papers dedicated to Manfred Markus on the occasion of his 65th birthday*, 169–182. Heidelberg: Winter.

- Leutloff, Kristina. 2003. Anglizismen in deutschen Jugendzeitschriften. Zu ihren stilistischen und kommunikativen Funktionen am Beispiel von *Bravo* und *Yam! Lebende Sprachen* 48. 5–13.
- Lipka, Leonhard. 2002. *English lexicology*. Tübingen: Gunter Narr Verlag.
- Ludwig, Klaus-Dieter. 1986. Nicht-denotative Informationen lexikalischer Einheiten als Wörterbucheinträge. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 39(2). 182–194.
- Ludwig, Klaus-Dieter. 2002. Registerkonzepte: Ein Überblick. In Cruse, D. Alan, Hundsnurscher, Franz, Job, Michael & Lutzeier, Peter Rolf (eds.) *Lexikologie. Ein internationales Handbuch zur Natur und Struktur von Wörtern und Wortschätzen*. 1. Halbband, 784–793. Berlin [etc.]: de Gruyter.
- Moravcsik, Julius. 1975. Aitia as generative factor in Aristotle's philosophy. *Dialogue* 14. 622–636.
- Moravcsik, Julius. 1981. How do words get their meanings? *Journal of Philosophy* 78. 5–24.
- Onysko, Alexander. 2007. *Anglicisms in German. Borrowing, lexical productivity, and written codeswitching*. Berlin [etc.]: Walter de Gruyter.
- Onysko, Alexander & Winter-Froemel, Esme. 2011. Necessary loans – luxury loans? Exploring the pragmatic dimension of borrowing. *Journal of Pragmatics* 43. 1550–1567.
- Oxford English Dictionary*. <http://dictionary.oed.com/>
- Pfitzner, Jürgen. 1978. *Der Anglizismus im Deutschen. Ein Beitrag zur Bestimmung seiner stilistischen Funktion in der heutigen Presse*. Stuttgart: J.B. Metzlersche Verlagsbuchhandlung.
- Poesio, Massimo & Almuhareb, Abdulrahman. 2008. Extracting concept descriptions from the Web: the importance of attributes and values. In Buitelaar, Paul & Cimiano, Philipp (eds.), *Proceedings of the 2008 conference on ontology learning and population: Bridging the gap between text and knowledge*, 29–44. Amsterdam: IOS Press.
- Pustejovsky, James. 1996. *The Generative Lexicon*. Cambridge, MA: The MIT Press.
- Sosnizka, Julia. 2014. *The collocational behavior of anglicisms in German and American business and news magazines*. Göttingen: Cuvillier Verlag.
- Stiven, Agnes B. 1936. *Englands Einfluß auf den deutschen Wortschatz*. Doctoral dissertation, Marburg.
- Weisgerber, Leo. 1962. *Von den Kräften der deutschen Sprache. Band I: Grundzüge der inhaltbezogenen Grammatik*. 3. Auflage. Düsseldorf: Pädagogischer Verlag Schwann.
- Williams, Edwin. 1981. Argument structure and morphology. *Linguistic Review* 1. 81–114.

Winter-Froemel, Esme. 2011. *Entlehnung in der Kommunikation und im Sprachwandel: Theorie und Analysen zum Französischen*. Berlin/Boston: de Gruyter.

Yang, Wenliang. 1990. *Anglizismen im Deutschen: am Beispiel des Nachrichtenmagazins Der Spiegel*. Tübingen: Max Niemeyer Verlag.

Heike Baeskow
Bergische Universität Wuppertal
Fakultät für Geistes- und Kulturwissenschaften
Anglistik/Amerikanistik
Gaußstraße 20
42119 Wuppertal
Germany
baeskow@uni-wuppertal.de

Jürgen Rolshoven
Universität zu Köln
Institut für Linguistik
Abteilung Sprachliche Informationsverarbeitung
Albertus-Magnus-Platz
50931 Köln
Germany
rols@spinfo.uni-koeln.de

In SKASE Journal of Theoretical Linguistics [online]. 2018, vol. 15, no. 1 [cit. 2018-18-06]. Available on web page http://www.skase.sk/Volumes/JTL36/pdf_doc/02.pdf. ISSN 1336-782X.