CHAPTER 2

Methodology, Data and Terminology

Most studies on lexical borrowing are limited to the analysis of the proportions of loans in different semantic fields without comprehensively examining the various effects of their adoption on the dynamics of the vocabulary of the receiving language. The various social and cognitive factors that influence the borrowing and dissemination of a foreign-derived lexical item have not yet been sufficiently researched either. The present study aims to overcome the shortcomings of existing studies in two respects:

1) by applying and expanding a coherent integrative socio-cognitive model (Schmid 2018) in order to illustrate the complexity of borrowing processes (according to different levels of investigation), which leads to more complex detailed descriptive results and implies a theoretical-methodological gain for the mapping and precise documentation of historical borrowing processes on the basis of Schmid’s (2018) model;

2) Thanks to a systematic extension of the primary data retrieved from dictionaries such as OED (henceforth also referred to as the primary corpus) by means of corpus material and linguistic evidence from social media (yielding the secondary corpus), a systematic analysis of the morphological, semantic and contextual variation of the examined borrowings has been carried out against a complex socio-cultural background.

1 Methodology

The findings presented in the present study are due to an empirical analysis; they rely on the evaluation of corpus data and the relevant research literature; in all, the results are based on a detailed qualitative analysis of a total of 1,247 borrowings (for details, see 2 below, table 1).

Using an integrative socio-cognitive model of the dynamic lexicon (see Schmid 2018: 215–231), the borrowing processes as well as their linguistic effects on English are to be illustrated. The model will be applied to lexical units which have been taken over from French, German, Spanish and Yiddish since 1801 and thereby refined and extended for borrowing processes both at the historical level as well as for more recent tendencies reaching up to the present.

1 For details on the structure of the primary and secondary corpora see 2.
Schmid’s model is a suitable basis for taking into account the abovementioned summarized points of criticism concerning existing studies since it depicts the complex interaction of usage, cognition and convention in its integrative character. To do so, it draws on pertinent research from a range of research traditions outlined in a preliminary draft (Schmid 2015: 5), the most significant of which are:

– usage-based and emergentist studies of grammar (e.g. Bybee 1985, Hopper 1987, Langacker 1988, MacWhinney 1999, Barlow and Kemmer 2000), of language change, such as the analyses by Bybee 2010, Bybee and Hopper 2001, and of language acquisition (e.g. Tomasello 2003),

– approaches to construction grammar (e.g. Fillmore et al. 1988, Goldberg 1995, Goldberg 2006, Hilpert 2013, Traugott and Trousdale 2013, Hilpert 2014),


– socio-cognitive studies (e.g. Kristiansen 2008, Croft 2009, Harder 2010, Geeraerts et al. 2010)

– investigations of language as a complex adaptive system (e.g. Blythe and Croft 2009), studies about specific differences in native language attainment (Dąbrowska 2012) and about individual differences and usage-based grammar (Barlow 2013),

– exemplar-based research (e.g. Bybee 2001, Pierrehumbert 2001),

– studies on formulaic language (e.g. Wray 2002, 2008),

– work in cognitive neuroscience (e.g. Pulvermüller 2003, 2013),

– research on variation from a sociolinguistic perspective (e.g. Labov 2001, 2010; Eckert 2000).

Research on lexical borrowing based on the theories and approaches considered by Schmid (e.g. cognitive concepts of entrenchment with its related mechanisms; socio-cognitive investigations; studies on sociolinguistic variation) has the potential to go beyond the individual results of previous studies in essential aspects. Schmid (2018) provides the first integrative approach to linguistic structure and usage taking into account cognitive, social, pragmatic and neurolinguistic theories with the aim of accounting for the nature and variation of linguistic structure. He gives an overview of the components of an integrative socio-cognitive model of the dynamic lexicon and its interaction (2018: 215–231). Its three essential components include actions in current usage, cognitive and social processes. According to Schmid (2018: 216–217), the lexicon is *dynamic* since it always changes in its structure and constituents, the lexical units and is therefore variable. Schmid (2018: 221) describes the model he has designed as an *Entrenchment and Conventionalization Model* which illustrates
the reasons for the dynamics of the lexicon. According to Schmid, stability and variability of the lexicon can be modeled through the interaction of language usage as well as specific cognitive and social processes. Figure 1 is my own English translation of the model including the relevant terminology:

Communicative actions and statements, i.e., the concrete use of lexical units in written and spoken language, are at the centre of the model. In cognitive processing, linguistic utterances are linked with symbolic, pragmatic, syntagmatic and paradigmatic patterns of association. Depending on the intensity of the processing, these patterns are routinized to a greater or lesser extent, schematized, finally consolidated in the mental lexicon and represented there. The entrenchment influences the degree of activation and processing. Strongly solidified linguistic patterns are activated faster than less entrenched ones. The linguistic knowledge represented in the mental lexicon is in turn subject to constant change through the use of language.
Co-semiosis, the common understanding of the meaning of an utterance by the language users involved in the communicative action, plays a crucial role in social conventionalization processes. It can lead to co-adaptation, that is, to the resumption of the relevant speech contribution (e.g., an emphasis pattern, a sentence structure, a lexical entity) in a given language usage situation. The frequency of repetition is crucial for the degree of conventionalization, i.e., the diffusion and normalization and finally the licensing of the types of expression in a language community (see Schmid (2018: 221–226)).

In his most recent monograph from 2020, Schmid expands his Entrenchment and Conventionalization Model by transferring it to the entire linguistic system, illustrating how this system is created, perpetuated and regularly influenced by the continuing interplay between usage, conventionalization and entrenchment. He shows how linguistic usage is converted into both collective and individual grammars, which themselves interact with usage. Within Schmid’s exceptionally wide approach to the topic, the phenomenon of language contact is just one of many aspects that can induce variability at the lexical level. A short section of his study is dedicated to general theoretical statements on linguistic contact as a force which may affect language change (see Schmid 2020: 119–123). Schmid (2020: 120) emphasizes that

[f]rom the perspective of the ec-Model, it is important to clarify which components of the model can be affected by language contact and in what ways. The primary locus of contact and the source of contact-induced change clearly lie in usage, more specifically in face-to-face encounters [...]. However, these encounters make two different contributions which should be distinguished. On the one hand, they give rise to acts or events of innovation by giving multilingual speakers an opportunity to use material from different languages. [...] If these innovations are salient to monolingual speakers, which they often are, they can get diffused, usualized, and routinized fairly rapidly [...]. Essentially, they become established like salient native innovations.

He also draws attention to the fact that the intensity of a given contact situation and the frequency of repetition are crucial factors for the spread and entrenchment of foreign-derived units and structures:

On the other hand, repeated multilingual encounters, and even monolingual encounters involving multilingual speakers, facilitate the usualization, diffusion, and routinization of any kind of utterance type from other languages, from distinctors to units and patterns, and from salient to highly unobtrusive ones. (2020: 120)
Since the focus of the present study is on processes related to variation and change in vocabulary and not in the linguistic system as such, Schmid’s original 2018 model will serve as a frame of reference (due to the fact that it focuses on the dynamics of the lexicon of a language). This model has great potential to describe borrowing processes against the background of the interaction between use, society and cognition. This is an essential aim of the present study, which seeks to obtain a comprehensive picture of the social and cognitive forces of borrowing processes. The systematic interplay of these dimensions of lexical development has so far received little attention in existing research.

Borrowing is a typical example of a lexical ‘innovation’ from the perspective of the receiving language; it facilitates the dynamics of the lexicon, as described by Schmid. As mentioned earlier, the present study examines the variability of the various borrowings in terms of their meanings, morphological development and contextual use in English, covering informal language in the social media. It will be illustrated how Schmid’s model and the interplay of the components he emphasizes can account for variation and change in the lexicon. For the present study, the following steps for the operationalization and differentiation of the model have been developed:

(1) Identification of the relevant borrowed lexical items in the primary corpus Schmid’s model is used to illustrate borrowing processes and their linguistic effects by taking the example of the French, German, Spanish and Yiddish impact on English since the nineteenth century. To collect the relevant words, electronic dictionaries such as *OED Online* have been used. In addition to the linguistic evidence available in *OED*, the comprehensive lexicographical samples of twentieth-century French and German borrowings provided by Schultz in 2012 and 2016, as well as her collection of nineteenth- and twentieth-century Spanish-derived terms from 2018b were consulted in order to compile an exhaustive primary corpus.

*OED* is currently being revised. Its digitalized version is available online at http://www.oed.com. It includes the second edition published in 1989, the *OED Additions Series* from 1993 and 1997 and a large number of updated and new entries in the planned third edition.² The electronic form of *OED* makes it possible to capture all lexical units that have a ‘foreign’ word origin in their etymological description. With the Advanced Search, Entries containing “French” (or “German”, “Spanish”, “Yiddish” etc.) in “Etymology” and “1801-” in “Date of

² For details concerning the revision of *OED*, see Durkin (1999: 1–49).
Entry”, borrowings which have been introduced into English since 1801 can be identified in *OED Online*. As will be seen, different types of loan influences occur among these borrowings. A number of them can be classified as direct loans which still show foreign influence with respect to their spelling and pronunciation in English, such as the French-derived culinary term *plat du jour*, pronounced /ˌplɑːdəˈʒʊə/ in BE and /ˌplɑːduːˈʒʊə/ in AmE (see *OED*). As is evident, the spelling of *plat du jour* was not modified during the borrowing process, and its stress pattern points to its French provenance as well. The *OED* sample also contains words of foreign provenance which have become assimilated. An example is the German borrowing *superman*. It represents a loan translation of German *Übermensch*, originally used by Nietzsche in a philosophical context. In the present study, all the different varieties of lexical borrowing are taken into account. *OED* also includes possible borrowings, i.e., lexical items which might have been derived from a foreign language. Such words were excluded from the present analysis. The lexical items under review are classified as introduced from French, German, Spanish and Yiddish as the immediate donor languages. Thus *pretzel*, for example, is categorized as an adoption of German *Bretzel* (now usually spelt *Brezel*) although the item ultimately goes back to the Latin *brachiātus*, *bracchiātus* (see *OED*).

Schultz’ studies from 2012, 2016d and 2018b constitute detailed analyses of the total number of twentieth-century French, German and Spanish borrowings in English. In all, she examines 1677 French-derived and 1958 German-derived borrowings from the twentieth-century in addition to 1355 words of Spanish origin taken over since 1801. Among them are a significant proportion of highly specialized technical terms which are not documented in general English. An example is *main succulente*, a rare medical term for a certain condition of the hands taken over from French in 1900. To identify comparatively common borrowings, a usage-based perspective is paramount in the present study (see step 2 for details).

The data offered in Schultz’ studies was updated and complemented by the linguistic evidence recorded in *OED*. Some lexical items classified as twentieth-century borrowings in Schultz’ analyses have been antedated as a result of the *OED* revision work. This holds for the common French borrowing *aioli*, for instance, initially labelled as an adoption from 1914 in *OED* and subsequently provided with an earlier date of first attestation due to the identification of a usage example from 1846. Recently, some twenty-first century borrowings

---

3 A detailed description of the framework used to classify different types of loan influences is offered in 3 below.
have also been added, such as the French-derived sports term *parkour*. All the changes relevant to Schultz’ data have been considered in the present study. The data set collected by means of *OED* was last updated in March 2021. The sample of lexical items retrieved from *OED* also contains a number of false matches; they have been eliminated.

Step (2): Identification of potentially entrenched structures

Cognitive Linguistics takes a usage-based perspective of language. From a cognitive point of view, linguistic signs consisting of a certain form that reflects a particular meaning become entrenched as a result of repeated usage and thus give structure to a language. The mutual and common use of linguistic structures and patterns points to their entrenchment in a language. Within Cognitive Linguistics, salience is crucial for the formation and perception of conceptual structures. This salience depends on various factors, including frequency and several different cultural and perceptive aspects which may render a particular concept more prominent. An evaluation of the linguistic data available in corpora, for instance, may reveal the frequency of specific usage patterns. Hence, it allows conclusions about the spread of these patterns in a language community (see also Glyn 2010: 89–90).

Sources such as *OED* include a considerable number of technical terms that would not normally be understood by the ‘average’ native speaker of English and which are not included in balanced corpora representative of current English usage, such as the *British National Corpus* (*BNC*) or the *Corpus of Contemporary American English* (*Coca*). In contrast to Schultz’ investigations from 2012, 2016d and 2018b, the focus of the present study lies on high-frequency borrowings which are familiar to the ‘ordinary’ native speaker of English. To distinguish between rare borrowings and those which have become comparatively widespread, EFL dictionaries such as the digital versions of the *Oxford Advanced Learner’s Dictionary* (*OALD*) and the *Longman Dictionary of Contemporary English* (*LDOCE*) were consulted. These types of resources are fairly meaningful and objective sources of well-used words of French, German, Spanish or Yiddish origin. Of the 21,395 lexical items recorded in *OED* and in Schultz’ studies, only 1,247 (i.e., 5.8%) make up a certain core field of relatively common borrowings which are also documented in *OALD* and/or *LDOCE*. These words have been investigated in detail; they represent potentially frequent/common structures which have entered the entrenchment and conventionalization process in the sense of Schmid (2018).

Frequency and generality are essential criteria for the inclusion of a lexical item in the EFL dictionaries consulted. To determine these factors, the dictionary editors use corpora such as *BNC*, *Oxford English Corpus*, Longman
Corpus of Spoken American and the web. In all, the linguistic data investigated comprises several million words from a variety of sources. In contrast to LDOCE, OALD occasionally also comprises technical terms that are not high-frequency items par excellence, but are nevertheless relatively common in domains such as the natural sciences; in these cases it is assumed that the average native speaker might come across such vocabulary items in oral or written language, which is a criterion for recording them in OALD. Resilience is also an important factor, i.e., a lexical item has to be documented in language use for several years before it is included (p.c. with the editors of OALD and LDOCE).

EFL dictionaries are reliable sources for determining the degree of conventionalization and entrenchment of a word in common usage. Inclusion in Learner Dictionaries is first and foremost a sign of conventionalization (as assumed by lexicographers). As a second step, one might assume that a word recorded in these sources is entrenched in the mind of the ‘average’ speaker. Since OALD and LDOCE are themselves based on a diversity of linguistic evidence representative of general English, no further corpora were used to analyse the frequency of the borrowings under review. For example, Google Ngram Viewer (also referred to as Google Books Ngram Viewer) was not used to identify comparatively widespread foreign-derived items in the present study. Although its underlying corpus (twelve million books) is very large, it does not comprehensively reflect all levels of language use. In addition, there seems to be an imbalance of different types of books in the corpus. For example, it comprises only a few usage examples of fairly common French-derived culinary vocabulary such as haute cuisine and coq au vin, which indicates a lack of cookery books.

The focus is not on the total vocabulary borrowed from French, German, Spanish and Yiddish since the nineteenth century, but on relatively widespread lexical units that have undergone changes over the two centuries with respect to their meaning, word form and contextual use and thus bring about variation and change in the lexicon of the receiving language as analysed in steps 3–5.

---

4 Although it was not used in the present study, Google Ngram does not necessarily have to be completely excluded from linguistic research. In some cases it might serve as an additional tool to illustrate developments in the vocabulary. Owing to changing fashion trends, for instance, there was an increasing use of French fabric terms in English in the earlier decades of the twentieth century. An analysis of Google Ngram reveals that in the case of georgette, for instance, the number of usage examples of the borrowing peaked in the 1920s, which confirms the findings of the present study (see chapter 3, section 1.5 below).
Step (3) Semantic analysis and recreation of the rich cultural contexts for semantic fields motivating cognitive, pragmatic, social and emotive-affective forces

To determine the various cognitive, pragmatic, social and emotive-affective forces which enhance the entrenchment and conventionalization of borrowings from the perspective of Schmid's (2018) model, the rich cultural contexts for semantic fields influenced by the four donor languages have been recreated (see chapter 3).

According to their meaning in current English, the various fairly common borrowings have been assigned to different semantic domains in order to capture the diversity of socio-cultural contexts influenced by French, German, Spanish and Yiddish since the nineteenth century. Sometimes a borrowing can be allocated to several different semantic fields. An example is *motif*, which belongs to ‘architecture’, ‘literature’ and ‘music’. *OED* usually provides the earliest documented use(s) of a borrowing in the relevant field(s) it belongs to. In some cases, however, the earliest usage of a word in a particular field cannot be clearly determined. An example is *troupe*, which has been recorded in *OED* since 1825, designating “[a] company, band, troop; esp[ecially] a company of players, dancers, or the like.” As to *troupe*, *OED* does not further differentiate between its earliest recorded uses in the various fields. For example, it might well be that it initially served as a theatre term before it came to be used in dancing contexts. In such cases, no further differentiation will be made either since analysing the earliest attestation of a loan in a given field would be too complicated because of the lack of comprehensive access to reliable sources to be used for this type of investigation.

In order to determine the extent of the influence of French, German, Spanish and Yiddish on the different semantic fields and spheres of life, all borrowings in the different fields were counted. Polysemous borrowings belonging to more than one semantic field had thus to be counted several times. Similarly, a differentiation was made between borrowings which belong to more than one word class. An example is the colour term *beige*, which shows both a nominal and an adjectival use in English. Such borrowings were counted twice.

The presentation of the individual domains to which the various donor languages have contributed in the form of new words and senses in chapter 3 depends on the degree of the influence of the foreign language in question on English: domains reflecting a relatively strong impact of French, German, Spanish or Yiddish follow domains where the influence of the respective donor language is less intense. This allows a comparison between the proportions of potentially entrenched structures from the different donor languages. A socio-cognitive perspective is taken to identify the various social, emotive-
affective, etc. factors which might have contributed to the borrowing, spread and entrenchment of the loans (in the sense of Schmid 2018).

Using \textit{BNC} and \textit{COCA}, the present study will give a detailed picture of the morphological and semantic change of the various high-frequency borrowings, which contributes to the dynamics and variation of the lexicon of the receiving language according to the underlying logic of Schmid’s model.

Step (4) Analysis of morphological change
Borrowings manifesting a long-term morphological and/or semantic development that has not been influenced by their donor language and has nothing to do with the (initial) socio-historical reason for their inclusion in English are of particular interest. Both \textit{BNC} and \textit{COCA} reflect authentic language usage in a relatively balanced manner since they include linguistic evidence from a variety of genres. \textit{BNC} consists of 100 million words of current BE. It is based on a wide range of text types such as fiction, newspapers, periodicals, academic texts and spoken data. With a total of one billion words, \textit{COCA} is a huge, well-balanced corpus which documents contemporary American usage comprehensively. It covers the period from 1990 to 2019 and includes linguistic evidence from a variety of sources (such as fiction, newspapers, academic publications, popular science texts and blogs.) A close perusal of the linguistic evidence recorded in the two corpora allows the analysis of the variation and change of borrowed vocabulary in different genres/registers. Based on the results obtained in this way, a differentiation of Schmid’s analytic model is to be undertaken with respect to registers and modes of language usage (e.g. in literary texts, press language and spoken language).

In order to identify the borrowings which have undergone morphological change, an open corpus search including wildcards was carried out. With the queries of corpora such as \textit{BNC}, the derivatives of the various borrowings can easily be detected by placing an asterisk before or after a given search term. Only derivatives from \textit{BNC} and/or \textit{COCA} that are also included in \textit{OALD} and/or \textit{LDOCE} (and thus belong to the core area of relatively widespread formations), have been taken into account. Idiosyncratic uses have been excluded from the corpus collected in this way.

A specific focus is on cognitive approaches to word formation, considering Langacker’s (1987) influential reference frame of cognitive grammar and concepts such as \textit{Figure/Ground Alignment}. From a cognitive perspective, word formations encode semantic extensions. Meaningful constituents (i.e., lexical units and affixes) are combined with simple lexical units to generate complex words. Working with corpora and cognitive approaches allows a comprehensive analysis of all derivatives of French, Spanish German and Yiddish borrow-
ings in line with Schmid's 2018 model since it comprises both a usage-based and a cognitive perspective.

Step (5) Analysis of semantic change
To document semantic change, the present project will offer a comprehensive overview of the diachronic sense development of the borrowings identified by means of corpora, such as the Corpus of Historical American English (COHA). Comprising more than 400 million words from the 1810s to the 2000s, COHA is the largest structured corpus of historical English, providing valuable insights into linguistic variation. Its genre balance is more or less identical each decade, which allows for a comparison between the use of borrowings in the various genres (including spoken language) from decade to decade. In the analysis, Blank's empirical, cognitive-based 1999 approach, which offers an innovative typology of motivation for lexico-semantic changes such as socio-cultural change, close conceptual or factual relationship, complexity and irregularity in the vocabulary of a language and emotionally marked concepts (see 1999: 70–83) will be considered. By means of a detailed qualitative analysis, the various borrowings were assigned to Blank's categories according to their use in the various decades (for details, see chapter 5). Blank's categories (e.g. socio-cultural change and emotionally marked concepts) are also relevant to the underlying logic of Schmid (2018) since they imply essential socio-cognitive factors and aspects that account for changes and variation in linguistic usage.

Step (6) Analysing patterns of usage
The different patterns of usage have been investigated on two levels: (1) current usage in different discourse domains (as reflected by social media and additional corpora) and (2) in terms of speakers' (pragmatic) choices between synonyms revealing socio-cognitive attitudes. In order to compare the treatment of the words of French, Spanish, German and Yiddish origin in dictionaries such as OED and Learner Dictionaries with their actual usage in various discourse domains including informal language in social media contexts, additional linguistic evidence from a variety of corpora and Internet blogs has been used.

Corpora which comprehensively document everyday language, such as the Spoken British National Corpus 2014 (Spoken BNC 2014), the TV Corpus, the Movie Corpus and the Corpus of American Soap Operas (Soap Corpus), make it possible to identify the typical informal contexts in which a borrowing is used after its first documentation. Of these, the Spoken BNC 2014 contains 11.5 million words of transcripts from 1,251 authentic informal conversations by 627 native British
speakers. The *TV Corpus* has 325 million words collected from 75,000 TV programmes between the 1950s and the early 2000s, and the *Movie Corpus* consists of 200 million words from approximately 25,000 movies from the 1930s to the present. Both corpora cover very informal usage in *BE* and *AmE*. The *Soap Corpus* makes it possible to evaluate informal usage in *AmE*. It contains 100 million words from 22,000 transcripts of American soap operas of the early 21st century.

Researching vocabulary in social media contexts requires the collection of linguistic data that includes evidence from a series of discussion forums and blogs concerning domains such as politics, gastronomy or fashion relevant for the vocabulary identified in the study. Balanced corpora such as *COCA* also include usage examples taken from a variety of blogs and other web pages. Apart from the relevant sections compiled in *COCA*, a comprehensive analysis of *Twitter* posts (available at www.twitter.com/trustedblogs) was undertaken in order to collect additional usage examples reflecting typical uses of borrowings in social media. The evidence collected on the basis of these sources illustrates Internet communication and thus represents a significant addition to the language use documented in corpora such as *BNC* or *COHA*.

Among the various borrowings under review, several items already have a synonym or near-synonym in the receiving language. To account for speakers’ preferences for synonyms in the different discourse domains, borrowings with (native) synonyms for all four donor languages have been analysed qualitatively to identify potential socio-cognitive motivations for their development and use. As to Schmid’s model, it has been essential to determine what is felt to be ‘French’, ‘German’ or ‘Spanish’ when carrying out contrastive analysis of the usage of borrowings in various discourse domains. In order to identify connections between the (conscious) use of a borrowed lexical item instead of a native synonym and social attitudes, precise descriptions of very particular and culturally embedded attitudes have been necessary.

The analysis of synonymy has long been a key issue in the field of Cognitive Linguistics. For example, Dirven et al. (1982), Lehrer (1982), Schmid (1993), Geeraerts, Grondelaers and Bakema (1994) and Rudzka-Ostyn (1995) are significant (early) empirical studies. Recent research covers lexical items which are near-synonymous (e.g. Newman and Rice 2004a, 2004b; Divjak 2006; Divjak and Gries 2006) as well as syntactical variability (e.g. Gries 1999; Grondelaers et al. 2007; Speelman and Geeraerts 2009).

Within a cognitive approach, a comparison of the use of semantically related lexical items is important since similarities or differences reveal their conceptual structure relevant for linguistic usage. Various extralinguistic and social variables including register have to be considered in order to yield a usage-
based account of the relevant linguistic patterns.\(^5\) Glynn (2010: 91) rightly points out that

[i]f we assume that speakers have knowledge of their language and culture and make their judgements based on that knowledge, this entails that their choices will reflect such knowledge. In Cognitive Linguistics, where entrenched language structure (or knowledge of language use) equates conceptual structure, by identifying the patterns of similar and distinctive usage, we chart the conceptual structure that motivates those patterns.

The investigation of lexical synonymy or near-synonymy is significant in a cognitive approach as it provides a technique for the identification of conceptual structures due to speaker choice and variation between different lexical items. These aspects have to be interpreted against the background of the social contexts in which the relevant lexical items are documented.

To identify native English equivalents which belong to the same lexical-semantic domains as the relevant borrowed lexical items, the *Historical Thesaurus of OED (HTOED)* was consulted. This reference work is now fully integrated into *OED Online* and groups the lexical entries from the dictionary according to similarity of meaning. Allan and Kay (2016: 222) outline that

[t]he combination of HTOED with the more detailed information about individual words available in the OED makes a powerful tool for lexical analysis, especially when reinforced by the increased availability of historical online corpora large enough to enable lexical research by allowing further scrutiny of contexts [...].

**HTOED** presents the semantic evolution of English vocabulary in all its diversity as it describes uses from Old English to the present day. With 800,000 words and meanings, the thesaurus includes a full set of meaning for English, based on entry categories. It is based on the second edition of OED and the *Thesaurus of Old English*. By means of this source, synonyms or near-synonyms of individual lexical entries were examined in combination with the linguistic documentary evidence in OED, which reflects the semantic and contextual variability of the corresponding vocabulary over time.

---

\(^5\) Extralinguistic aspects relevant for a usage-based description have moved into the focus of linguistic concern in cognitive approaches, see Kristiansen and Dirven (2008), Geeraerts et al. (2010) for discussion and illustrative examples of this domain of research.
According to Glynn, who assumes a cognitive-functional perspective, synonymy can be defined as “different forms for a concept-function” (2014: 10). In the present study, the notion of synonymy corresponds to Glynn's understanding of the term.

Schmid’s (2018) model shows how the lexical system is created, sustained and regularly refreshed by the continuing interaction between usage, conventionalization and entrenchment. Steps 1–6 of the present analysis serve both to demonstrate and enhance the role of borrowings in a dynamic lexicon. Hence, in a final step, they have been used to elaborate and specify Schmid’s current model at the levels of borrowing processes, their linguistic outcomes and the various socio-cognitive factors to which they are related (see chapter 7).

The original and innovative aspects of the research conducted for the present study can be summarized as follows:

1. The presentation of a new approach to loanword study which combines a sociolinguistic and a cognitive perspective to describe borrowing processes and their linguistic outcomes;
2. innovative methodological approaches (including the use of an integrative, socio-cognitive model of the dynamic lexicon) which have as yet been used rarely in research on lexical borrowing;
3. an analysis of an important part of the English foreign vocabulary that has been neglected in earlier studies (i.e., the investigation of the various French, German, Spanish and Yiddish borrowings, their derivatives and their semantic development);
4. the provision of a reconstruction of the manifold contexts for the different domains and their related fields (e.g. humanities, gastronomy, people and everyday life);
5. an evaluation of the impact of French, Spanish, German and Yiddish on English, explained against a complex and multi-layered socio-cultural background.

2 Data

In the existing literature, Schmid’s model has only been tested by means of individual examples (see, for example, Schmid 2018: 226–230). The following gives a concise overview of the primary corpus compiled for the present analysis in steps 1 and 2 as well as secondary corpora scrutinized for the analysis of their use in current English in steps 4 to 6. Drawing on step 1, the present study yielded a corpus of 16,405 lexical borrowings distributed among the donor lan-
The data illustrated in table 1 as a result of the first and second analysis steps create a special corpus of the core of lexical borrowings from French, German, Spanish and Yiddish and their historical development, by means of which the dynamics of the development of this area of the English vocabulary can be traced precisely from a socio-cognitive and historical perspective. This special corpus forms the basis for the analyses presented in steps 3–6.

For the proportions of borrowings from the nineteenth and twentieth centuries and their distribution among the various semantic fields, see chapter 3.
On this basis, the relevant semantic fields to which the individual borrowings can be assigned are analysed with their respective motivating cultural contexts (chapter 3), their morphological and semantic development in English (independent of the donor language) (chapters 4 and 5) as well as their use in various discourse domains extending to the present day (chapter 6), in order to capture the details of this dynamic in detail within the framework of the present socio-cognitive approach.

New media, such as corpora and Internet forums, have made it possible to examine how and to what extent the relatively widespread borrowings show morphological, semantic and contextual variability, and thus produce variation and change in the lexicon of the receiving language. Table 2 gives an overview of the structure of the secondary corpus used to analyse these aspects.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type of analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BNC</strong></td>
<td>morphological and semantic analysis</td>
</tr>
<tr>
<td><strong>Coca</strong></td>
<td>morphological and semantic analysis</td>
</tr>
<tr>
<td><strong>COHA</strong></td>
<td>diachronic semantic analysis</td>
</tr>
<tr>
<td><strong>Spoken BNC 2014</strong></td>
<td></td>
</tr>
<tr>
<td><strong>TV Corpus</strong></td>
<td>analysis of usage patterns including informal language in social media contexts</td>
</tr>
<tr>
<td><strong>Movie Corpus</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Soap Corpus</strong></td>
<td></td>
</tr>
<tr>
<td>Internet blogs and web pages in <strong>COCA</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Twitter posts</strong></td>
<td></td>
</tr>
</tbody>
</table>

3 Terminology

The following gives an overview of the terminology employed in the present study. I shall begin with the definitions of word, lexical item and term, all of which represent essential concepts in a lexical analysis.

3.1 Word, Lexical Item and Term

3.1.1 Word
The term *word* has been defined from many points of views. An early syntactic understanding of the word is provided by Sweet (1875: 474), according to whom it is “an ultimate, or indecomposable sentence.” Sapir (2007: 37) assumes
a syntactic and semantic perspective and defines the term as “one of the smallest, completely satisfying bits of isolated ‘meaning’ into which the sentence resolves itself.” There are also approaches which consider semantic, grammatical and phonological criteria, as is reflected by Arnold’s study *The English Word* (1973: 9):

The term word denotes the basic unit of a given language resulting from the association of a particular meaning with a particular group of sounds capable of a particular grammatical employment. A word therefore is simultaneously a semantic, grammatical and phonological unit.

In the present investigation, the understanding of word relies on the notion of the term as revealed by Arnold.

### 3.1.2 Lexical Item
Quirk et al. (1985: 2.35) define lexical item as “a word as it occurs in a dictionary.” The term thus refers to a meaningful vocabulary unit in its most abstract sense without an assigned grammatical function. In the present study, the use of lexical item relies on the notion of the term offered by Quirk et al.

### 3.1.3 Term
*Term* designates a lexical item which manifests a specific sense in a particular domain, such as the humanities or the natural sciences.

### 3.2 Classification of Loan Influences
One of the generally accepted classification schemes for categorizing types of loan influences was developed by Haugen in 1950. It represents an influential twentieth-century study on language contact which is still considered a standard reference work, as pointed out by other researchers on contact phenomena (e.g. Clyne 1987; Winford 2003; Matras 2009). The studies by Betz (1949, 1959) have equally become widely acknowledged approaches to the classification of lexical borrowing. Betz’s categorization scheme is based on his evaluation of the impact of Latin on Old High German. These typologies of loan influences also inspired a number of further analyses including Gneuss’ (1955) survey of loan formations and semantic loans in Old English as well as Carstensen’s 1965 and 1968 studies on the German-English language contact scenario. Haugen’s

---

7 In her study on the history of English, Strang (1970: 368) uses Gneuss’ (1955) approach as a frame of reference since it was the most recent study on loan influences at the time of her research.
approach is still considered an established frame of reference despite the manifold studies on language contact phenomena which have suggested a variety of modified terminologies to distinguish between the different types of loan influences. Thomason (2001: 61), for instance, juxtaposes the term borrowing, which is generally used to refer to the adoption of a foreign-derived lexical item into a specific target language, with the expression imperfect learning to refer to the transfer of structural or semantic features from one language into another. There is also Johanson (2002), who investigates structural aspects in Turkic language contact. He proposes to replace the term borrowing with the term copying since the latter seems to be the more adequate designation of the process occurring in a linguistic contact scenario.

A plethora of studies offer attempts to systematize the terminology for identifying language contact phenomena. Recent examples are Onysko’s monograph from 2007 and Haspelmath’s paper from 2009, both of which comprise a critical analysis of the existing terminology. Onysko’s paper from 2019, which proposes a more general view by re-conceptualizing conventional notions of borrowing, code-switching, transfer and calquing with a specific focus on their cognitive interrelation.

The terminology which will be used in the present study relies on Carstensen’s (1968) approach because it comprises the essential types of lexical borrowing (e.g. direct loan, loan translation, semantic loan, pseudo-loan) and is applicable to various loan influences and language contact situations. Here is an overview of the terms to be used:

3.2.1 Borrowing

*Borrowing* is the common term for the process by means of which a language adopts a word or a meaning from another language. It can also specify the word or sense borrowed in this process (see also Bußmann 1996: 55).

3.2.2 Direct Loan

*Direct loan* specifies the acquisition of a foreign-derived lexical item showing no or only slight orthographical and phonological adaptation to the linguistic system of the receiving language. The term can also designate the relevant adopted lexical item. *Poltergeist*, which was taken over from German in 1848, can serve as an example. It corresponds to the German *Poltergeist*, pronounced /ˈpɔltɐɡaɪ̯st/ in the donor language (see *Duden Online*). Wells (2005: 590) offers the following pronunciation variants: /ˈpɒltəɡaɪst/ for BE and /ˈpoʊltərgaɪst/ for AmE. As can be seen, the spelling of the word has been maintained in English, and the pronunciation variants of the borrowing still point to its German origin; it can be classified as a direct loan.
3.2.3 Foreign Word, Loanword
A number of German-language studies on lexical borrowing differentiate between *Lehnwörter* ‘loanwords’ and *Fremdwörter* ‘foreign words’ as further subcategories of direct loans. Carstensen points out that a *Lehnwort* is assumed to be more strongly assimilated to the phonological, orthographical and inflectional system of the receiving language than a *Fremdwort*. This differentiation has not been made in the present study as the direct loans borrowed from French, German, Spanish and Yiddish since 1801 diverge only slightly with respect to their adaptation to the linguistic rules of the receiving language.

3.2.4 Exoticism
The sample under analysis also contains a number of culture-specific terms. Some of them are designations of referents which only exist in the foreign-language community from which they were borrowed. In the literature, these borrowings are described as exoticisms (what some German linguists also refer to as *fremde Wörter*, e.g. Steinbach 1984: 35–36). An example of an exoticism is the French borrowing *souk*, which is used to designate “[a]n Arab market or marketplace, a bazaar” (*OED*) in English. Its French source term *souk* ultimately reflects the Arabic form *sūḳ* ‘marketplace’ (see *OED*).

3.2.5 Adaptation
*Adaptation* is used to designate the process by which a foreign-derived lexical item is assimilated in line with the linguistic rules of the receiving language. It can also be used to denote the word naturalized during this process. The French borrowing *évacuee* serves as an example. It was taken over from French *évacué* in 1934. During the borrowing process, the accents were dropped, and the French ending was replaced by an indigenous one, yielding a naturalized form with respect to both spelling and pronunciation.

3.2.6 Loan Translation
*Loan translation* is the process by which the various elements of a foreign-derived item are literally translated into the receiving language. The term can also relate to a word produced in this manner. An example is *blue blood*, a loan translation of Spanish *sangre azul*.

3.2.7 Loan Rendition
*Loan rendition* designates the process in which a complex lexical item is coined where one constituent of the foreign model is translated literally and another part is substituted by an element lacking a formal relation to the relevant source term. It can also relate to the word created during this process. The nineteenth-
century borrowing *Prince Charming* is an example. It is a loan rendition of French *roi charmant*, literally ‘king charming’.

3.2.8 Loan Creation
A *loan creation* occurs when a complex word is created on the model of a foreign lexical item with no formal correspondence to the original term. *Epistemology* is such an example. It was formed from ancient Greek ἐπιστήμη ‘knowledge’ and the combining form -ology, after the German compound Wissenschaftslehre.

3.2.9 Semantic Loan
A *semantic loan* occurs when an already existing native word assumes a particular meaning of a foreign lexical item. *New Wave*, for example, can be categorized as such. It has been attested in English since 1960 as a general designation of a new trend (e.g. in the arts). Owing to the influence of the French expression nouvelle vague, *New Wave* expanded its sense content by adopting the additional sense of “[a] movement in French filmmaking from the late 1950s to early 1960s, characterized by an emphasis on individual directorial style, innovative editing and filming techniques, and a preference for existentialist themes” (*OED*). *Semantic loan* can refer to both the adoption of a foreign-derived meaning and the relevant sense transferred in this way.

3.2.10 Hybrid (Hybrid Compound, Hybrid Phrase)
A hybrid refers to a compound or phrase which includes a foreign-derived and a native element. Hybrids may be formed on the model of a foreign-language item. This is true of *liverwurst*, for instance, a partial translation of the German compound *Leberwurst*. Some hybrids lack a foreign-language model, as is the case with the phrase *borscht belt*. It was formed from *borscht*, a spelling variant of Yiddish *borsht* and English *belt*, relating to “([t]he hotels of) a resort area in the Catskill Mountains, New York State, popular with Jewish people of Eastern European origin” (*OED*). The phrase does not show a corresponding equivalent expression in Yiddish.

3.2.11 Pseudo-Loan, Pseudo-Compound, Pseudo-Phrase
*Pseudo-loan* refers to the creation of a lexical item in the receiving language from linguistic constituents of a foreign language or the lexical item thus coined. The word produced in this manner only looks like an acquisition from a particular donor language. Yet, an equivalent term does not exist there. The sample of borrowings under scrutiny comprises pseudo-compounds such as *bon-viveur* and pseudo-phrases like *crème de la crème*. These lexical items were
formed by means of French elements within English (see also chapter 6 for a
detailed discussion.)

3.2.12 Double and Multiple Loans
Within his classification scheme, Carstensen differentiates between double
and multiple loans. According to Carstensen (1968: 44), this type of loan in-
fluence occurs when a lexical item is borrowed in various senses two or more
times from a particular donor language. However, the borrowings analysed
in the present study have not been divided into double and multiple loans.
Carstensen’s differentiation, as Duckworth (1997: 55) emphasizes, might be dif-
ficult since a borrowed foreign-language item might continue to be influenced
from a relevant donor language for a long time after being introduced. The pro-
cess described by Carstensen will simply be considered a sense development
of a borrowed lexical item and not as a double or multiple re-borrowing of an
identical foreign-language item. In some cases a new sense of a borrowing also
exists in the donor language. Yet, it would be too difficult to ascertain whether
it is due to the continuing impact of the relevant foreign language or whether it
results from an independent semantic development within the receiving lan-
guage.

The different categories outlined by Carstensen generally correspond to
Haugen’s reference frame from 1950. However, Carstensen’s approach is slightly
more comprehensive since it includes further subcategories not made explicit
in Haugen (1950). Haugen (1950) differentiates between loanwords (i.e., bor-
rrowed words which are identical with their source terms in form and meaning
but may show phonological adaptation—cf. Carstensen’s division into foreign
words and loanwords), loanblends (hybrid formations, i.e., lexical items com-
prising a foreign-language constituent and a native one) and loanshifts, i.e.,
cases where the entire foreign-language item is rendered or substituted by indi-
genous elements (what Carstensen refers to as loan translations and semantic
loans). Another difference between the two classification schemes is that Haug-
en only pays attention to hybrid constructions that “involve a discoverable
foreign model” (1950: 215), while Carstensen also considers hybrids without a
foreign-language model as types of loan influence.

The work of Haspelmath (2008, 2009), which deals in particular with the
motivation for borrowing in contrast to other, comparable processes such as
code switching, will also be considered as a theoretical foundation in the
present study. Haspelmath’s (2008) study provides an overview of types of bor-
rowing, aspects affecting borrowability as well as aspects influencing the loan
processes of several different languages from a comparative perspective. It is
based on the findings related to the study on borrowings and borrowability
carried out at the Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany, from 2007 to 2008. As far as borrowability is concerned, Haspelmath (2008: 49) points out that “[i]t is widely acknowledged that lexical items are more likely to be borrowed than grammatical items, and that words are more likely to be borrowed than bound morphemes [...].” It is also well-known that nouns are adopted more frequently from one language into another in comparison to other parts of speech. According to Van Hout and Muysken (1994: 42),

[a] very important factor involves one of the primary motivations for lexical borrowing, that is, to extend the referential potential of a language. Since reference is established primarily through nouns, these are the elements borrowed most easily.

Another essential factor motivating the borrowing of vocabulary is the prestige attached to a particular foreign language. Haspelmath (2009: 48) draws attention to the fact that

[t]he way we talk (or write) is not only determined by the ideas we want to get across, but also by the impression we want to convey to others, and by the kind of social identity that we want to be associated with.

It is generally acknowledged that French, for instance, has long functioned as the prestige donor language per se. Haspelmath (2008: 52–53) also adduces further factors which might influence borrowability, such as structural similarity of donor and receiving languages, which usually motivates the assumption of especially grammatical features.

The integration of borrowings into the recipient language is generally considered a gradual process. Foreign-language items are assumed to be adopted initially as instances of code-switching before they become more and more naturalized over time, provided that they are used often enough by speakers of the language community borrowing them (see, for example, Van Coetsem 2000, Myers-Scotton 2002, Thomason 2003). Yet it is also argued that the distinction between code-switches and borrowed foreign-language items is not possible. Some scholars claim that such a distinction is problematic since a suitable method has not yet been developed for this (e.g. Eliasson 1989, Eastman 1992, Johanson 1993, Thomason 2001, Winford 2003, Gardner-Chloros 2009). Other researchers point out that there is no need to differentiate as the two phenomena represent different instances or phases within the borrowing process regarded as a continuum anyway (e.g. Eliasson 1989, Heath 1989, Ben-

Poplack and Dion’s study “Myths and facts about loanword development” (2012) is an innovative approach to theories related to the borrowing process. Looking at the use of items of English origin in Quebec French over a time span of 61 years, the authors come to the conclusion that “[l]inguistic integration is abrupt, not gradual” (Poplack and Dion 2012: 279). They (2012: 296) claim that

[...] [W]hen speakers access a lone other-language item, they make an instantaneous decision about whether to treat it as a borrowing or a code-switch. If they opt to borrow it, they produce it with all the requisite recipient-language morphosyntactic trappings, variability included, independent of considerations of frequency, diffusion, or listedness. What if they opt to code-switch? Based on the criterion of retaining donor-language grammar, speakers apparently do not avail themselves of this option with respect to lone other-language items. Of the 601 tokens of unambiguous nonce lone other-language items studied diachronically, only one could qualify as a code-switch [...]. In other words, lone other-language items tend to be borrowed, and borrowings are introduced already as borrowings, even if the vast majority of them will not go on to become bona fide loanwords.

These results are illuminating in many respects. Yet, Poplack and Dion focus on grammatical features such as verb inflection and plural marking to examine the development of the adaptation of the linguistic items under review. They do not look at further aspects related to the borrowing process, such as the semantic or morphological integration of foreign-language words.

The focus of the present study will be on the interrelation between use, meaning and the mind as a central issue of contact-induced linguistic variation and change from a historical point of view. Further frameworks are employed which combine a cognitive perspective with a usage-based approach, to analyse the semantic and morphological development of the various French, German, Spanish and Yiddish borrowings. The respective typologies with their related terminologies are described in detail in the chapters 4 and 5 below.