A Corpus-assisted Approach to Paronym Categorisation

Ruth Maria Mell, Petra Storjohann

Institute for the German Language, R5, 6-13, 68161 Mannheim, Germany
E-mail: mell@ids-mannheim.de, storjohann@ids-mannheim.de

Abstract

In this paper, we will present a first attempt to classify commonly confused words in German by consulting their communicative functions in corpora. Although the use of so-called paronyms causes frequent uncertainties due to similarities in spelling, sound and semantics, up until now the phenomenon has attracted little attention either from the perspective of corpus linguistics or from cognitive linguistics. Existing investigations rely on structuralist models, which do not account for empirical evidence. Still, they have developed an elaborate model based on formal criteria, primarily on word formation (cf. Lăzărescu 1999). Looking from a corpus perspective, such classifications are incompatible with language in use and cognitive elements of misuse.
This article sketches first lexicological insights into a classification model as derived from semantic analyses of written communication. Firstly, a brief description of the project will be provided. Secondly, corpus-assisted paronym detection will be focused. Thirdly, in the main section the paper concerns the description of the datasets for paronym classification and the classification procedures. As a work in progress, new insights will continually be extended once spoken and CMC data are added to the investigations.

Keywords: paronyms; commonly confused words; e-dictionary; categorisation; semantic classification

1. Introduction

Paronyms are words that are similar in spelling, sound and / or meaning, i.e. *formell / formal / förmlich* (*formal*), *Technik / Technologie* (*technology*), *elektrisch / elektronisch* (*electric / electrical / electronic*), *Methode / Methodik / Methodologie* (*method / methodology*)[1] etc.[2] In this sense, paronyms are easily confused words which regularly cause problems for language learners and native speakers. Generally, such pairs of paronyms are not regarded synonymous although corpus analyses suggest that some items undergo meaning change due to the rivalry of two or more paronyms:

> "Sometimes, [paronyms] can develop synonymous notions and simply become lexical alternatives (cf. Storjohann, 2015). In other cases, they remain similar in meaning but show subtle differences and restrictions in usage. Inevitably,

[1] The first group are all essentially *formal*; the second are essentially *technology*; English has the same problem with *electric / electrical / electronic* and *method / methodology*.

[2] For more examples see Schnörch (2015).

situations of confusion arise when speakers' intuitions contradict information in existing reference works." (Storjohann, forthcoming)

So far, paronyms have been looked at only from a structuralist point of view and mainly from a language learners' perspective (cf. Lăzărescu, 1999). Up until now, the phenomenon has attracted little attention from the perspectives of corpus linguistics and cognitive linguistics. With the availability of diverse corpora, particularly spoken data and the development of new semantic approaches, only recently has paronymy become the focus of a new project ("Paronyme – Dynamisch im Kontrast"). The project lexicographically documents paronyms in a new corpus-based e-dictionary. Furthermore, it focusses on research on paronymy as a lexical-conceptual phenomenon and aims to develop an empirically-driven classification of paronyms using diverse genres of corpora including written and spoken texts as well as CMC data. In the past, investigations have relied on models accounting for language as a formal and logic system and not requiring empirical evidence in real communicative situations.

While a detailed description of the e-dictionary with respect to structure, content, navigation and visualisation is provided by Storjohann (in this volume), the central aim of this paper is to attempt to classify commonly confused words in German by consulting their communicative functions and semantic manifestations in written corpora.

2. The Paronym Dictionary ("Paronymwörterbuch")

The new German online dictionary "Paronymwörterbuch" (*Paronym Dictionary*) (Storjohann, 2014; Storjohann and Schnörch, 2017), which is currently being developed at the Institute for the German Language (IDS, Mannheim), breaks new ground by adopting a more conceptual and encyclopaedic approach to meaning by incorporating cognitive features. It will be published in 2017 and is publically accessible free of charge.[3] It is the very first corpus-assisted reference guide to the contemporary use of paronyms with regard to German.[4] The online dictionary strives to exploit the possibilities of using the electronic medium more effectively and in order to create an innovative, flexible and user-friendly instrument instead of listing traditional, linear and static entries. In doing so, this dictionary represents a step towards a dynamic, multi-functional cognitive-oriented online reference work with adaptive navigation (for details see Storjohann in this volume).

3. Corpus-assisted paronym detection and paronym analysis

Language data used for compiling dictionaries is often outdated, or lexicographic

[3] It will be published in the dictionary portal OWID (www.owid.de) in 2017.

[4] To our knowledge there is no corpus-guided, electronic reference guide of easily confused pairs in any other language.

practice is rather conventional and does not take advantage of corpus-assisted approaches to semantic analysis. The objective of the "Paronymwörterbuch" is to compile a new kind of dictionary with contrastive entries which will be a useful reference tool in situations of language doubt. At the same time, it aims to sensitise users to context dependency and language change.

> "As the subject of paronyms has not been revisited with empirical, data-driven methods either in terms of semantic theory or in terms of practical lexicography suitable corpus methods for contrastive investigation needed to be tested. Currently, complementary software-driven resources facilitating the search for similarity and difference are being exploited, each of which is based on the analysis and interpretation of contextual profiles, collocations and colligations, corresponding semantic roles and syntactic functions." (Storjohann, forthcoming)

To create the new online dictionary "Paronymwörterbuch", innovative approaches to empirical lexicographic work that pave the way for a new data-driven, descriptive reference work of confusable German terms have been adopted. An index (lemma list) is an essential pillar of every type of dictionary. For this reason, the concept, corpus extraction and compiling of a lemma list is a key task in the initial phase of every lexicographic project (cf. Schnörch, 2015: 16).

The first step in the paronym dictionary project was to find potential candidates for a paronym index. Consulting traditional print dictionaries such as Pollmann & Wolk (2010), Duden 9, and Müller (1973) provided us with typical pairs and their morphological features. We were then able to establish groups of candidates based on a variety of formal patterns (Schnörch, 2015), e.g.:

-al/-istisch (*natural*/*naturalistisch*)

-end/-lich (*dringend*/*dringlich*)

-ig/-lich (*fremdsprachig*/*fremdsprachlich*)

-sam/-lich (*betriebsam*/*betrieblich*)

Approximately 154 such formal categories were detected through the study of texts and dictionaries.

With the help of large corpora, all pairs which differed with respect to such patterns (often regular suffixes) but were identical in their root were automatically extracted using the 'string comparison' method. As a database, we used DeReWo (version derewo-v-ww-bll-320000g-2012-12-31-1.0). DeReWo consists of frequency-based rankings of lemmata and word forms on the basis of virtual corpora. These lists of

lemmata and word forms in use in the German language (for example the lemma candidate list with 350,000 entries for *elexiko*[5], the online dictionary of contemporary German) are generated by applying the methods for creating frequency-based rankings of lemmata and word forms on DeReKo – the German Reference Corpus (cf. http://www1.ids-mannheim.de/direktion/kl/projekte/methoden/derewo.html?L=1).[6]

In the next step, all automatically retrieved pairs were analysed manually. Overall about 9000 cases were scrutinised, 2000 were considered potential candidates. They were then categorised according to frequency (Storjohann & Schnörch, 2017). Two years ago, semantic analyses and lexicographic descriptions of the most frequent pairs started using different analysing tools and methods. An examination of the paronym list reveals a remarkable attribute of all these words. The candidates of the index are not an arbitrary jumble of words; by segmenting the character strings, morphological patterns and regular occurrences can be found. Among them is the study of significant collocations as identified by the corpus tool COSMAS II[7] – the Corpus Search and Management Analysis System. A further effective procedure is the use of the contrasting-near-synonym-method (CNS). This is profitably employed for contrastive analyses.

4. Datasets / Corpora for paronym classification

In this chapter, we will describe the corpora we are currently using for the analysis of paronyms. We will also present further options using different corpora for a future comprehensive classification of paronyms, paying particular attention to our base corpus "Paronymkorpus" (which is the basis for detailed paronym analysis). These different data resources will hopefully enable us to define a wider spectrum of variational properties and specific communicative idiosyncrasies otherwise not detected through the sole use of newspaper texts.

4.1 Paronymkorpus

As all analyses are guided by large corpora, for our initial investigations we have compiled a special, publically accessible corpus (the so-called Paronymkorpus) that contains written texts from between 1990 and 2015, comprising around 2.3 billion tokens. We have built a corpus based on DeReKo (the German Reference Corpus Collection, hosted by the Institute for the German Language (IDS) in Mannheim). DeReKo includes vast amounts of texts from genres as diverse as newspapers, fiction, parliamentary debates, and specialised text with different terminologies from more technical language use (cf. Kupietz & Lüngen, 2014).

---

[5] *elexiko*: http://www.owid.de/wb/elexiko/start.html.

[6] COSMAS II: https://cosmas2.ids-mannheim.de/cosmas2-web/.

[7] For details on analysing methods see Storjohann and Schnörch (2017).

With respect to German, the Paronymkorpus is the first lexicographic data resource that is completely open to the public. As it contains texts without restrictions of copyright it allows lexicological investigations and lexicographic documentation to be completely transparent. Concerning the regional distribution of the newspaper data (Figures 1 and 2), the corpus can be defined as relatively well-balanced (Paronyme – Dynamisch im Kontrast: project description, http://www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonymkorpus.html) compared to others, e.g. elexiko http://www.owid.de/wb/elexiko/glossar/elexiko-Korpus.html).

Figure 1: Regional distribution of newspapers in the Paronymkorpus

Currently, the main focus of the project is on the analysis and description of the most frequent paronyms in written language data, especially in newspapers. Besides dialectal diversity of smaller regional newspapers and standardised nation-wide reception of larger journals, one major advantage of this text type is its variety of authors and subjects and genre (e.g. weather forecasts, adverts, political and scientific reports, readers' letters etc.). The underlying paronym corpus consists of the following texts in more detail (see Figure 2):

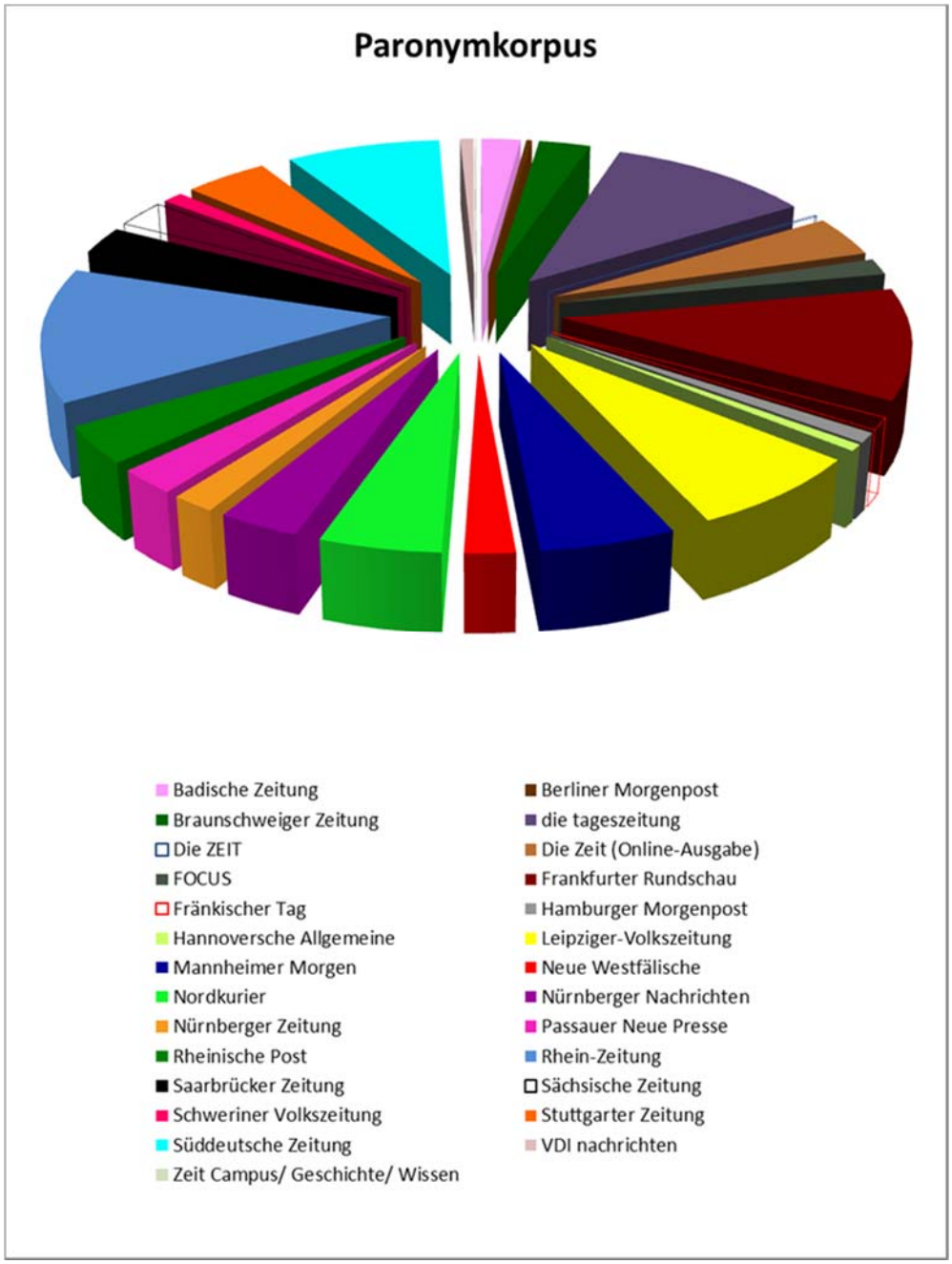

Figure 2: Percentage of newspapers in the Paronymkorpus

4.2 FOLK – the Research and Teaching Corpus of Spoken German

In a further step, we will look at technical terms and easily confused pairs in spoken data as a lexical database for expert communication and the Datenbank für Gesprochenes Deutsch (DGD-IDS 2012-2017) (Database of Spoken German) as a resource for spoken communication. Specifically, FOLK, the Research and Teaching Corpus of Spoken German, which is part of the DGD will be used for our linguistic

research and lexicographic investigations (cf. Stift & Schmidt, 2014; Schmidt, 2016: 398). FOLK is a large corpus of spontaneous verbal interactions in German (Schmidt, 2016: 396–397), containing a growing number of TV-interviews and conversations. As Schmidt (2016: 397) points out, FOLK

> "i. covers a broad range of interaction types in private, institutional and public settings;
>
> ii. is sufficiently large and diverse and of sufficient quality to support different qualitative and quantitative research approaches;
>
> iii. is transcribed, annotated and made accessible according to current technological standards;
>
> iv. is available to the scientific community on a sound legal basis and without unnecessary restrictions of usage." (Schmidt, 2016: 397)

Another reason for using FOLK is that it is a balanced corpus. Schmidt writes: "FOLK also attempts to control for some secondary variables, like regional variation, sex and age of speakers, in order to achieve a balanced corpus" (Schmidt, 2016: 398). FOLK currently contains data from 259 different conversations. This makes 202 recorded hours and 1.95 million tokens. (DGD: New version of DGD, http://www1.ids-mannheim.de/prag/artikelansicht/article/neue-version-der-dgd-3.html). Unfortunately, so far spoken and written data cannot be analysed using one and the same corpus tool since they are incorporated into different systems. As a consequence, results have to be individually interpreted and their underlying data need to be explicitly mentioned in order to relate findings to their source of information. Hopefully, the corpus systems of the next generation will be able to process both types of data.

4.3 Wikipedia Corpus

In a final step, we will additionally use the German Wikipedia Corpus[8] (hosted at the Institute for the German Language) for analysing the use of paronyms in computer-mediated communication (CMC). It is through the research of paronyms in a third textual variety that our findings can cover a larger spectrum of the German language than would be possible by looking at written corpora only. Margaretha & Lüngen (2014) describe Wikipedia as a large and rich online encyclopaedia that covers an unbelievably wide range of subjects including history, sport, arts and culture in articles and talk pages (discussions). As a language repository, Wikipedia provides a wealth of multilingual natural language data, also useful for the analysis of knowledge concepts and ontological categories. Since the content of Wikipedia has not been

[8] Available under http://www.ids-mannheim.de/cosmas2/.

written by a single author, but collaboratively by many users, it is particularly interesting for the study of computer-mediated communication (CMC) (Margaretha & Lüngen, 2014: 59), as aspects of dialog and mediation need to be considered. Of particular importance, might be the Version Control System (VCS) for documenting the various versions of an entry, including editorial comments and remarks.

Analytical relevance is given, as this kind of corpus data gives us the opportunity to analyse CMC language data spontaneously and dialogically. The Wikipedia corpora are also available as a virtual corpus in the COSMAS II corpus search and analysis system. Currently, only research of written texts is being conducted; this will be followed by further investigations of spoken data and analyses of Internet texts in the following years. The findings will be documented as part of the dictionary in different sections.

5. Paronym Categorisation and Classification Procedures

At the moment, our paronym classification is solely based on written corpora and it only relates to analyses of roughly a hundred paronym pairs. Needless to say, it cannot lead to a sufficient classification model but has already provided us with valuable insights into functions in thematic domain, discourse and style, text types, and degrees of semantic similarity or contrast of easily confused words. It is expected that in the future we might be able to come up with a detailed terminology covering paronyms from different angles.

A closer look at the different communicative and discursive functions of paronyms has so far suggested the following cases:

i. general (non-technical) paronyms with some conceptual overlap but individual constructional preferences, e.g. *praktisch / praktikabel* (*practical*), *nötig / notwendig / notwendigerweise* (*necessary / necessarily*),

ii. discourse-identifying word pairs, i.e. paronyms strictly determined through specific (critical) discourse, e.g. *national / nationalistisch* (*national / nationalistic*) in political discourse; *unehelich / nichtehelich* (*illegitimate / out of wedlock*) in official language discourse. The wrong choice between them can lead to politically incorrect use,

iii. pairs with different connotations with the tendency to be misused more frequently in spoken conversations, e.g. *bäuerlich / bäurisch* (*rural / peasant*), *weiblich / weibisch* (*feminine / effeminate*); one item has a neutral connotation while the other is of negative pragmatic value,

iv. opposites denoting similar concepts but with concrete contrary specifications, e.g. *konkav / konvex* (*concave / convex*), *Stalagmit / Stalaktit* (*stalagmite /* stalactite); users are usually aware of a distinction but lack factual knowledge

in specific situations,

v. paronyms with strong similarities in spelling but no semantic closeness, e.g. *ethisch / ethnisch* (*ethnic / ethical*); There is no overlap on the designated concept and confusion leads to clear mistakes,

vi. pairs with different syntactic functions, e.g. *fraglich / fragwürdig* (*questionable / dubious*); there are restrictions of grammatical usage for one member of a pair, such as adverbial, attribute or predicative role of adjectives,

vii. synonyms which specifically occur in different thematic domains, e.g. *sportlich / sportiv* (*athletic / sporty*); these are identical in meaning but are preferably used in different subjects,

viii. pairs with a very different distribution and frequency pattern, e.g. *Adaption / Adaptation* (*adaption / adaptation*), *herzlich / herzig* (*warm, lovingly / cute, heart-shaped*).

Taking the class of thematically related synonyms (vii) as an example, the differences between the adjectives *sportlich / sportiv* can be summarised as follows: Generally, both denote a person as physically fit, healthy and athletic. Hence, they can be used synonymously. Still, they differ with respect to their collocates.

Collocates of *sportlich* are, for example, *Figur* (*figure*), *Fitness* (*fitness*), *Statur* (*stature*), *Mann* (*man*), *Täter* (*culprit*), *Pensionär* (*pensioner*) (all of which refer to people and their appearances). Contexts in which *sportlich* occurs together with these collocates are predominantly found in police reports, illustrating the thematic domain of descriptions of criminal offenders (see examples 1, 2 and 3) [9]:

1. *Ein **Täter** soll 18 bis 20 Jahre alt und 1,65 Meter groß sein. Er soll eine **sportliche** muskulöse **Figur** und kurze schwarze leicht gelockte Haare haben. Bekleidet war er mit weißem T-Shirt, dunklen Jeans und weißen Schuhen.* (Frankfurter Rundschau, 29.05.2007, S. 36)

2. *Nach übereinstimmenden Aussagen mehrerer Zeugen ist er 20 bis 22 Jahre alt, 1,80 Meter groß, hat kurze Haare und eine **sportlich**, kräftige **Statur**. Bekleidet war er mit schwarz-weiß karierten Bermudashorts, dunkelblauem T-Shirt und Basecap.* (Leipziger-Volkszeitung, 31.05.2014, S. 19)

3. *Freitagvormittag sah Schiefer zufällig, wie ein Einbrecher in das Haus seines Sohnes auf der anderen Seite der Gustav-Mahler-Straße einstieg. Seine Schwiegertochter mit ihrer kleinen Tochter war glücklicherweise nicht mehr im Haus, stellte er nach einer Schrecksekunde mit Blick auf den Parkplatz*

[9] The examples are taken from the Paronymkorpus.

fest. Der ***sportliche Pensionär*** *alarmierte die Polizei über Handy, bewaffnete sich mit einem Golfschläger und filmte das Haus von der anderen Straßenseite aus.*(Rheinische Post, 16.11.2006, Diebe bei Einbruch gefilmt)

Collocates of *sportiv* are, for example, *Typ* (*type*), *Menschen* (*people*), *Erscheinung* (*appearance*), *Biker* (*biker*), *Models* (*models*), *Damen* (*ladies*), all of which refer to general denotations of humans. Frequently, these can be found in contexts of sports and health issues (see citations 4 and 5):

4. "*Fit for Life*" *lautet das Motto zweier Grundlagenseminare, die "rz sporty" am Mittwoch, den 7. bzw. 14. Februar, zwischen 18 und 21 Uhr im RZ-Haus in Koblenz veranstaltet. Sportmediziner Prof. Dr. Peter Billigmann und die Diplom-Ernährungsberaterin Birgit Binninger-Heid vermitteln dabei Ernährungstipps für* ***sportive Menschen****. Folgende Themenkomplexe werden behandelt: Weg mit dem Winterspeck - wie nehme ich gesund ab; Fitnesssport und Ernährung - zehn Prinzipien für Essen und Trinken im Sport; Herzkraft und sportliche Leistung; Träge im Winter, topfit im Sommer - das wichtigste über das Immunsystem.* (Rhein-Zeitung, 25.01.2001, Die letzten Reste werden gesucht.)

5. *Petras ist nicht nur äußerlich, als notorischer Baseballkappenträger, der* ***sportive, kämpferische Typ****, er ist es auch in seinem Verständnis vom Theatermachen.* (Die Zeit, 12.10.2006, S. 53, Im Hagel der Stücke.)

As emphasised before, in a second step, spoken data and CMC data will be investigated in terms of paronym behavior. We have indicative evidence that specific aspects occur in different genres, styles and registers only or preferably. For instance:

i. There are paronyms that are more typically confused in spoken communication, e.g. *anvisieren / avisieren* (*to target / to notify*). In such situations, mistakes occur more frequently as "side effects" of spontaneous, unreflected speech. These are particularly revealing in terms of cognitive processing.

ii. There is a class of technical terms, i.e. paronyms originally from expert communication, mostly in written language, but also in spoken language, e.g. *Parodontose / Parodontitis* (*periodontosis / parodontitis*), *Arthrose / Arthritis* (*arthrosis / arthritis*). Confusion occurs in everyday language but not in technical terms. In public discourse, such terms are treated differently from medical contexts.

The list is neither complete nor homogeneous, but it accounts for some formal and linguistic elements. Without doubt, these distinctions and classes listed above are only a first sketch approaching the phenomenon of paronymy from a usage-based perspective. These first findings do not constitute a uniform classification but suggest that different linguistic aspects need to be taken into account and any adequate

approach to grouping paronyms requires a multi-layered, cross-classification. Hopefully, on the one hand, the features mentioned above enable us to find usage-based definitions and restrictions of paronyms, and on the other hand, they inform us about guiding principles of semantic change in authentic language in use.[10] In order to be able to identify classes and to be able to provide an adequate and comprehensive description of the various kinds of paronyms, it is, however, necessary to use different data sets for a more refined classification model: As results vary according to which corpus we use for our analysis, we distinguish between paronyms that are most frequent in written, spoken and CMC-language data.

Overall, the findings concerning the classification of paronyms are not only theoretically relevant. They help us to find criteria which reflect usage behavior, context-dependent functions and cognitive principles rather than formal, logical distinctive aspects isolated from contexts. As a result, information on their features as described here are implemented in the dictionary entries (or will be in the future) in different ways, e.g. through specific sections, guidewords, explicit reference in the paraphrase.

6. Conclusion

The focus of this paper was to present a first attempt to classify commonly confused words (so-called paronyms) in German by studying their communicative and discursive functions in written corpora. These unveil different categories compared to traditional models and principles. Paronyms have not been studied empirically in language use so far. Sound corpus-guided studies of paronyms show different meanings from traditional dictionaries; a contextual usage-based approach leads to different categories of classifications than structural accounts. Our categories of paronyms exemplify text-functional aspects with regard to contextual relations as illustrated by collocation constructions. These uncover complex semantic structures and relational networks and we are able see how paronyms behave differently in contextual patterns and discourse.

At the moment, the bases of our investigations are very large written corpora. In the future, additional text types and genres of written as well as spoken language (see section 4) will play a vital role in defining paronyms and in embedding the phenomenon into a larger semantic framework. This necessarily has to imply approaches to real language in use and a variety of registers for a more objective view on communication and language in general.

[10] Another interesting aspect of research implies the rivalry of paronyms and their mutual contextual as well as cognitive influence on each other.

7. References

Duden 9. Richtiges und gutes Deutsch (2007). *Wörterbuch der sprachlichen Zweifelsfälle*. Edited by P. Eisenberg, F. Münzberg & K. Kunkel-Razum (Dudenredaktion). 6. ed. Mannheim et al.: Dudenverlag.

Kupietz, M. & Lüngen, H. (2014). Recent Developments in DeReKo. In N. Calzolari, K. Choukri, T. Declerck, H. Loftsson, B. Maegaard, J. Mariani, A. Moreno, J. Odijk & S. Piperidis (eds.) *Proceedings of the Ninth International Conference on Language Resources and Evaluation* (LREC'14), Reykjavik, Iceland. European Language Resources Association (ELRA).

Lăzărescu, I. (1999). Die Paronymie als lexikalisches Phänomen und die Paronomasie als Stilfigur im Deutschen. Bukarest: Anima Verlag.

Margaretha, E. & Lüngen, H. (2014). Building Linguistic Corpora from Wikipedia Articles and Discussions. Available at: http://www.jlcl.org/2014_Heft2/3MargarethaLuengen.pdf.

Müller, W. (1973). *Leicht verwechselbare Wörter*. Duden-Taschenwörterbücher, Bd. 17. Mannheim: Bibliographisches Institut.

Pollmann, C. & Wolk, U. (2010). *Wörterbuch der verwechselten Wörter*. Stuttgart: Pons.

Schmidt, T. (2016). Good practices in the compilation of FOLK, the Research and Teaching Corpus of Spoken Language. In *International Journal of Corpus Linguistics* 21: 3 (2016), pp. 396–418.

Stift, U.-M., & Schmidt, T. (2014). Mündliche Korpora am IDS: Vom Deutschen Spracharchiv zur Datenbank für Gesprochenes Deutsch. In Institut für Deutsche Sprache (ed.) *Ansichten und Einsichten. 50 Jahre Institut für Deutsche Sprache*. Mannheim: Institut für Deutsche Sprache, pp. 360–375.

Schnörch, U. (2015). Wie viele Paronympaare gibt es eigentlich? Das Zusammenspiel aus korpuslinguistischen und redaktionellen Verfahren zur Ermittlung einer Paronymstichwortliste. *Sprachreport*, 4/2015, pp. 16–26.

Storjohann, P. (forthcoming): Cognitive descriptions in a corpus-based dictionary of German paronyms. In Anatol Stefanowitsch & Stefan Hartmann (eds.) *GSCL Yearbook 2017*. Berlin: de Gruyter.

Storjohann, P. & Schnörch, U. (2017). Sprachlicher Wandel der Gegenwart und seine Dokumentation in einem Wörterbuch. In M. Vachková, M. Šemelík & V. Kloudová (eds.) *Themenheft Lexikographie, Germanistica Pragensia, AUC Philologica*, 4/2016, pp. 133–172.

Websites:

Cosmas II. Accessed at: https://cosmas2.ids-mannheim.de/cosmas2-web/ (18 May 2017)

DGD-IDS (2012-2017). DGD. Datenbank für Gesprochenes Deutsch. Accessed at: http://dgd.ids-mannheim.de/dgd/pragdb.dgd_extern.welcome?v_session_id (17 May 2017)

DGD: New version of DGD. Available at: http://www1.ids-mannheim.de/prag/artikelansicht/article/neue-version-der-dgd-3.html (17 May 2017)

DEREKO: Deutsches Referenzkorpus. Available at: www1.ids-mannheim.de/kl/projekte/korpora/ (10 May 2017)

DeReWo: version derewo-v-ww-bll-320000g-2012-12-31-1.0. Download: http://www1.ids-mannheim.de/direktion/kl/projekte/methoden/derewo.html?L=1 (5 May 2017)

elexiko. Accessed at: http://www.owid.de/wb/elexiko/start.html (30 April 2017)

FOLK: Forschungs- und Lehrkorpus gesprochenes Deutsch. Available at: http://dgd.ids-mannheim.de (26 April 2017)

Paronyme – Dynamisch im Kontrast. Project description. Accessed at: http://www1.ids-mannheim.de/lexik/paronymwoerterbuch/dasparonymkorpus.html (1 May 2017)

PARONYM-Korpus. Available at: https://cosmas2.ids-mannheim.de/cosmas2-web/faces/investigation/corpus.xhtml.

WIKIPEDIA-Korpus (2013/2015). Available at: https://cosmas2.ids-mannheim.de/cosmas2-web/faces/investigation/archive.xhtml.