UNIVERSITÄT
HEIDELBERG
ZUKUNFT
SEIT 1386

Modeling Lexical Fields for Translation: a Corpus-Based Study of Armenian, German, and English Culinary Verbs

Dissertation
zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.)
der Neuphilologischen Fakultät
der Ruprecht-Karls-Universität Heidelberg,
Institut für Übersetzen und Dolmetschen

vorgelegt von Meri Dallakyan

Erstgutachter: Prof. Dr. Vahram Atayan
Zweitgutachter: Prof. Dr. Bogdan Babych

Tag der mündlichen Prüfung: 26. Juli 2024

Heidelberg 2025

To my family

Ընտանիքիս

Acknowledgement

Writing a doctoral thesis is undoubtedly an emotional journey, filled with a wide array of experiences at every stage. However, the most meaningful part for me is reflecting on the journey itself—on the memories and moments that made this work possible. My thesis spans through a global pandemic, an ongoing war in my homeland, moments of joy like my son’s first day at kindergarten and my daughter’s first day at school, as well as personal challenges like my hip replacement surgery. These years have been full of highs and lows, with many small victories along the way. The countless hours spent on this dissertation have been marked by both moments of disappointment and joy. Despite the difficult times, I have no regrets about embarking on this journey to earn a PhD from Heidelberg University. As I look back, I want to express my deepest gratitude to the people who made this achievement possible.

First and foremost, I owe my deepest thanks to my primary supervisor (I actually prefer the German *Doktorvater*), Prof. Dr. Vahram Atayan, whose unwavering patience and support have been instrumental throughout my entire PhD journey. This work would not have been possible without his guidance and his commitment to providing multiple solutions to a single problem. I am especially grateful for the way he taught me to approach research in his own mathematical style—sometimes making it feel lighter through his witty examples. On a personal level, I am deeply thankful for his attention to my well-being, both physical and mental. I am grateful for his encouragement to “limit myself to stress,” which prevented me from falling into despair. I don’t master the English language to the extent to express what I feel in Armenian. *Հոգաչափ շնորհակալ եմ:*

I am also immensely grateful to my second supervisor, Prof. Dr. Bogdan Babych, for agreeing to examine my work and offer invaluable insights. His guidance helped me establish clear boundaries for my research, steering me away from being overwhelmed by the vast sea of scientific literature on the topic. I sincerely appreciate his support, his encouraging words, and his positive belief that “the work is good enough to earn a PhD.”

My doctoral studies would not have been possible without the financial support of the DAAD. I also want to thank the *Graduiertenakademie der Universität Heidelberg* for their generous contribution in the form of an *Abschlussbeihilfe* through the *Landesgraduiertenförderung*.

I would also like to express my gratitude to my colleagues and now friends, Dr. Annika Straube and Bettina Fetzer, for their constant support and encouragement. Special thanks to Vanessa Šorak, Daniele Moretti, and Eva Klüber for reading different chapters of this work and providing insightful feedback. I am particularly grateful to my two soulmates, Astghik and Varduhie, for their unwavering patience and dedication throughout these years. Astghik, my German tutor, *wir werden hoffentlich bald auch deine Dissertationsabgabe feiern!* Vard, *ջան*, it would be impossible to express in words how thankful I am for always being there for me, especially when I needed someone who almost understood what I've been researching for the last few years!

My deepest gratitude goes to my parents, Anush and Armik, for never allowing my circumstances to define my potential. Your unwavering support has been my foundation in all my endeavors. *Երախտապարտ եմ ձեզ:* To my sister, Mariam, thank you for always believing in me and for having more faith in my abilities than I ever did myself. And to my youngest sister, Marine, thank you for providing a refreshing perspective on life, often asking about recipes instead of research! I also want to thank my in-laws for reminding me that there's a life outside of my dissertation.

Finally, I dedicate this work to my family. I owe you more than words can express. I cannot undo the time I've spent on this dissertation, but I hope I've managed to be a caring mother and wife as well. To my husband, Arman, *thank you* for letting me be myself. Thanks for always listening to my complaints about things that you did not get, even though you never admitted it! I really appreciate it! To my lovely daughter, Nane, thank you for teaching me the importance of balancing work, study, and rest (though I'm still working on it!). And to my son, Narek, you've probably endured the most, seeing me "always glued to this dumb computer." I hope you're both proud of mama, and if it weren't for you, I could never have completed this work. I did my best. *Ես ձեզ շատ եմ սիրում. Երեքդ էլ իմ գանձերն եք*: I promise this was my last exam... or maybe the second to last. You know me better than anyone!

Table of Contents

Introduction

This work aims at contributing to one of the current trends in semantic analysis using statistical and computational data analysis tools to describe linguistic and world knowledge driven from authentic comparable corpora. It is a bottom-up[1] corpus-based[2] description of culinary lexical fields in Armenian, German, and English with the objective to also provide possible translation models, in particular, a methodology for systematically identifying translation equivalents from comparable corpora as one of its outcomes. Both language teachers and learners as well as translators can benefit from the methodology developed in this work. The proposed methodology will assist avoiding atypical uses of collocations in culinary text production and provide translators with specific contexts where the typical or most probable lexical choice is identified based on prior manual semantic annotation and intralinguistic analysis. The proposed translation model could also be used for evaluating and improving machine translation systems, specifying linguistic conditions for specific lexical choices in given linguistic contexts. The framework could also be used to evaluate, verify and correct the choices made by neural machine translation models for words, especially culinary verbs that belong to the analyzed semantic fields. In this respect, even with the drawbacks of the method due to objective and subjective reasons, e.g. manual annotation pitfalls and challenges, the model makes a contribution to evaluating and improving the quality of machine translation as a heuristic approach and avoids any severely wrong lexical choices in both the production and the translation of culinary texts in Armenian, English, and German. The methodology used in this work could easily be applied to other languages, as the parameters used for the annotation of the corpus-based examples are characteristic of culinary texts in general. For instance, the method has also been tested for the French language and has operated successfully, the results of which will presumably be published separately. Certainly, the application of the method to different languages might and probably would require amendments and/or addition and/or reduction of the annotation parameters, however, as a type of a specialist text most of them would remain the same and allow both intralinguistic analysis and language comparison. One of the relevant contributions of this work is the *interpretability* of the developed language model

[1] For the comparison of the top-down and bottom-up methods of linguistic analysis on the example of pandemic discourses see Hunston (2022:295).
[2] Corpus-based approach to analyzing linguistic data differs from corpus-driven in that the former methodology aims at reinvestigating theories and descriptions carried out before large corpus data and analysis tools were available “to expound, test or exemplify them” (Tognini-Bonelli 2001:65) while the latter views corpus as an embodiment of the language theory itself (ibid. 84, cf. McEnery and Hardie 2012:5).

based on manual semantic evaluation of concordances aimed at deriving the meaning of words in their use. The current language models used in natural language processing (NLP) and artificial intelligence (AI) rely on semantic similarity derived from automatically calculated contextual features that are not interpretable. A potential contribution of the proposed methodology could thus be a step towards developing explainable translation models for the lexicon that are able to motivate translation choices made by machine translation (MT), as the current neural MT models still lack interpretability and are therefore prone to errors that are difficult to predict or to correct.

This work also contributes to corpus-based lexicography in that it identifies which features should be included in the definition of the word in a given lexical field, in this case the culinary verbs, which would lead to the improvement of dictionary definitions, i.e. the intension, the identifying features of the word. In other words, this work develops a model of word meanings as a type of word embedding of culinary verbs in Armenian, English, and German in multidimensional space without any loss of interpretability. It is important to emphasize that this method does not merely identify all possible uses of the word in context based on quantitative methods of gaining data analysis, but also the typical ones supported by qualitative analysis, i.e. manual semantic annotation of concordances.

This method identifies word patters from large corpora. **The novelty of the work** is the linguistic interpretability of the word profiles through features resulted from the manual semantic annotation of the occurrences extracted from the respective corpora. Currently word vectors (word embeddings) model distributional semantic properties of the lexicon in modern AI/NLP applications, such as MT systems (e.g. DeepL, Google Translate, Bing) or conversational agents (e.g. ChatGPT). They are used in large neural language models, which are capable of translating or generating text, and are trained on considerably larger corpora; however, the outcomes of this work show that training models based on relatively smaller corpora by combining both statistical computation and manual semantic annotation could yield notable and semantically interpretable results. **The contribution of the work** is the suggested methodology of bridging the gap between logic-based semantic annotation and data-driven processing of language.

Other representations of semantic relations between words, e.g. WordNet and FrameNet, model them as hierarchies or networks of labelled relations between concepts. These networks are referred to as ontologies (which normally represent hierarchies of words belonging to semantic field of the same type) and frames (where the relations can span different types of words used in specific situations or scenarios).

EuroWordNet includes more than 10 European languages, Global WordNet Associates currently aims at developing WordNets for all the languages, including, for the moment, besides European also Asian, African as well as Oriental languages http://globalwordnet.org/resources/wordnets-in-the-world/. However, this method of describing word relations is especially successful for nouns that could "fit" in some type of a hierarchy, but not for verbs (cf. Hanks 2013 for expressing the same concerns as a shortcoming of the hierarchical ontological representation of word frames). Verbs require a special consideration and method of frame description.

> What does work in practice for a very large number of words, especially verbs and adjectives, is the correlation of different meanings with different sets of collocational preferences and syntagmatic patterns. (Hanks 2013:19)

Moreover, choosing verbs over nouns for describing the culinary fields in Armenian, German, and English allows to provide much more information with just a limited number of lexemes. For instance, choosing eight verbs from the German culinary corpus over the same amount of nouns ensures the analysis of considerably larger data, since verbs in recipes involve processes objectively covering significantly bigger part of the culinary field, i.e. the verb *braten* covers all processes where *heat* is involved.

Therefore, collocational analysis comes to fill in the gap of presenting the frame, however, not in the conventional Fillmorian (Fillmore 1982) sense of frames describing the world (cf. the famous *Restaurant frame* of Barsalou (1992), Blank's (2001) contiguity) but the frames/fields in the corpus, derived from the corpus data, not imaginary, not invented but already existing there.

Irrespective of the adopted method of corpus-based analysis, this work does not aim at describing all possibilities of the language use, nor does it intend to show the extreme cases of the lexicon where the language loses its natural character. Instead, it has the objective to reveal how the languages in question actually work within the norm. As Roman Jakobson (1959/1971) states, "Languages differ essentially in what they *must* convey and not in what they *may* convey" (264). Therefore, our method of corpus-based collocation analysis with its modest contribution to the theory of language and translation is an attempt to show what languages convey and how they differ from one another. In other words, this work emphasized the importance of collocations in determining the meaning of the words (cf. Firth 1950, 1957a, 1957b, Sinclair 1987, 1991, 1998, 2004) using corpus evidence.

The following work is an attempt to provide a contribution to fulfilling the existing desiderata in corpus-based linguistic analysis, with the objective of finding the meaning of a word in its usage, identifying the prototypical uses of the word, but also the possible ones that might differ from the "correct" pattern of word use, but be completely meaningful and "normal"[3] in the speaking community. Being an advocate of Firthian phraseological and collocational meaning, as well as corpus-driven (and corpus-based) language acquisition and analysis, Hanks, in his thorough work *Lexical Analysis, Norms and Exploitation*, states the need for more efforts to be undertaken in research to discover what possible word uses the given language corpora identify, instead of just singling out the normal one:

> It is the task of descriptive linguists and lexicographers alike to discover the principles that govern both normal usage and the ways in which it can be varied. It is a task that, until recent years, they have performed spectacularly badly. The biggest single problem has been a failure to distinguish normal usage from unusual but possible usage. (Hanks, 2013:343)

It is hard to precisely define what is and what is not a conventionally prototypical use of a word, since "normal usage is based on social salience, frequency, whereas cognitively salient linguistic phenomena are often socially not salient" (also in Hanks 2013, cf. Giora 1997, 2003). We are interested in word use based on corpus evidence, considering not only salient usages of words but also cognitively plausible yet less prominent ones.

Twentieth-century lexicographers were left with the more logical-philosophical definition of meaning, in the Leibnizian sense of "necessary and sufficient conditions" to make dictionary word definitions, and, for the most part, closed their eyes to new linguistic theory that suggested a novel approach of defining word meaning.

> Discovering conventions requires searching for regular patterns of word use shared by different speakers and writers; this is a sociolinguistic task, which can only be achieved by comparing large numbers of uses of each word in different texts. Such a task was not possible until the

[3] Languages express the thoughts of the speech community. It is utterly important who uses the language. According to Hanks (2013:349-350), in order to be able to recognize the thoughts of the speaking community, one should differentiate between "normal patterns of word usage from idiosyncrasies", suggesting that his corpus-driven Theory of Norms and Exploitations (TNE) or an equivalent could undertake this task.

> development of corpus linguistic technology in the closing decades of the twentieth century. (Hanks 2013:3)

Thus, this work aims at finding the meaning of a word in its actual usage and not what is conventional in an idealized language environment, where word combinations become so bizarre that native speakers would never use them in that way or would find them artificial.

Unlike Leibniz (1646–1716, through the works of Couturat 1901, 1903), Russel (1905), Frege (1892), and even later linguists such as Fillmore (1982) and Jackendoff (1983, 1990, 1991, 2002), who concentrated on finding the definitions for concepts and describing the concept frames with logical relations, we are interested in the meaning of a word in its use, thus the culinary fields are not described on the conceptual level but on word level, and if viewed as culinary frames, the aim of the following work is not to describe the logical relations connecting the items of the frame, but rather reveal their collocational behavior.

Corpus-based language analysis comes to fill in these gaps of revisiting already existing theories, supported by large amount of authentic data. [4] Corpus-based or corpus-driven evidence would certainly have provided more *real* data and yielded more realistic outcomes with a better picture of the language being examined. A corpus-based word field analysis approach also contributes to contrastive lexical studies by singling out normal word usage as well as possible but less frequent ones. It emphasizes the importance of real authentic data[5] instead of invented ones to yield tangible results in linguistic analysis. For this purpose, corpus linguistics provides authentic data in large quantities and analysis tools to describe the language in its use.

The following corpus-based descriptive study of the culinary fields in Armenian, German, and English is grounded on the statistically significant collocational analysis as well as computational heuristics, which emphasizes not only the typical syntactic combinability of words coinciding with their encyclopedic meaning, but also explores those excluded from the conventional dictionary entry of the word, which are nevertheless proven by corpus evidence.

Chapter 1 provides the research methodology, accompanied by the most significant theoretical and practical contributions to the study of lexical semantics and word usage. **Chapter 2** gives a comprehensive description of the German sub-corpus, as well as the introduction of the annotation parameters underlying the linguistic analysis of entire work. This chapter is then followed by the description of the culinary field in German, based on eight German culinary

---

[4] Hanks suggests that all works "dating from before 1987 must be reevaluated systematically in the light of corpus evidence" (2013:361).

[5] Sinclair (1991) emphasizes the lack of "authentic data" before 1987.

verbs, i.e. *backen, braten, dünsten, grillen, kochen, rösten, schmoren,* and *toasten*, and their collocations derived from corpus evidence. The collocations[6], extracted from the corpus through quantitative methods (frequency, statistical significance) and linguistic analysis tools, are then manually semantically annotated. This is followed by various visualizations of data analysis, illustrating the combination of quantitative and qualitative research methods to present deeper insights. **Chapter 3** introduces the English sub-corpus, and subsequently describes the English culinary field based on the collocation analysis of six English culinary verbs, i.e. *bake, braise, cook, fry, roast,* and *sauté*. Several data analysis methods and their visualizations, such as *Correspondence Analysis* (CA) and *Multiple Correspondence Analysis* (MCA), *Conditional Inference Trees* (CIT) as well as *Mosaic-Plots and Context-Conditional Correlation Graphs* (CCCG), guide the reader in the evaluation of considerably large, complex linguistic data with numerous variables that would be impossible for a human being to analyze without such visualization methods based on statistical significance. **Chapter 4** introduces the Armenian sub-corpus, focusing on the thorough description of the Armenian food field based on the collocational analysis of six Armenian culinary verbs, i.e. *բովել [bovel]*, *եփել [ephel]*, *թխել [thxel]*, *խաշել [xashel]*, *շոգեխաշել [shogexashel]*, and *տապակել [tapakel]*. Throughout the work, both the Armenian and the German examples are translated into English word by word to provide the reader with the necessary understanding of the context. However, as a type of specialized text, culinary texts are rich in culture-specific items, the translation of which could itself be a topic for further research. Moreover, the translation of the above-mentioned culinary verbs into the respective languages represents a significant challenge, which is addressed in **Chapter 5** of this work. Therefore, verbs with potentially more than one translation possibilities are marked with an asterisk (*) indicating all possible equivalents, while the whole sentence is enclosed in square brackets as [].[7] The outcomes of the linguistic analysis of the Armenian culinary field are also illustrated by different graphical visualization methods, followed by their interpretation. Thus, **Chapter 5** of this work attempts to provide the reader with possible translation equivalents for the culinary verbs across the three languages in comparison. The

[6] Several scores had been previously developed to determine the association between two words seen as collocations, among them *T-score*, *MI-score*, *MI³-score*, *Minimum sensitivity*, *MI log Freq*, and *Dice* (cf. Evert 2005).

[7] The square brackets [] throughout this work represent an interlinear gloss. Since a direct equivalent of the culinary verbs, in most cases, is not possible and is not the objective of this research, which rather aims to provide a comprehensive comparison of culinary verbs in German, English, and Armenian, with one of the outcomes being the *identification* of possible translation equivalents derived from corpus-based linguistic data analysis, the translation of the examples serves primarily to provide the reader with at least some semantic understanding of the examples in question. Possible equivalent verbs are marked with an asterisk *.

identified translation models are (among other factors) the outcome of comparable corpora analysis, where deep *intra*-linguistic investigation serves as the basis for *inter*-linguistic comparison. The same visualization methods demonstrate the comparison between the culinary verbs in six potential translation directions, i.e. German <-> Armenian, English <-> Armenian, and German <-> English. Both the intra- and interlinguistic analyses serve as test cases for the quantitative and qualitative methods and the overall methodology adopted within the framework of this research. The conclusions sum up the highlights of the following work and emphasize its own contribution to the current trends in corpus analysis.

The **objective** of the following corpus-based research is to give a comprehensive description and modeling of Armenian, English, and German culinary lexical fields using considerably large data sets by combining qualitative (manual semantic annotation) and quantitative (computational analysis based on statistical significance) methods. The present research also envisages to compare the aforementioned culinary lexical fields and to suggest possible translation models as one of its outcomes. The intralinguistic analysis serves as the basis to also identify possible translation models, thereby contributing to language acquisition and teaching, lexicographic works, and the enhancement of human-oriented monolingual as well as bilingual dictionaries. In addition, the suggested model could be integrated into computational linguistics and computational lexicography research in the form of specific tag sets and linguistically informed annotations, contributing to improving the quality of machine translation and machine learning. It also offers a way to specify differentiating features in the word entries in such projects as FrameNet (https://framenet.icsi.berkeley.edu/), established by Charles Fillmore.[8]

The central part of this work is primarily methodological in pursuing its objectives by adopting the collocational analysis method and bridging the conventional logic-based semantics, such as the componential analysis and the distributional approach to analyzing linguistic data. Thus, the objects of study within the scope of this work – eight German (*backen*, *braten*, *dünsten*, *grillen*, *kochen*, *rösten*, *schmoren*, and *toasten)*, six English *(bake*, *braise*, *cook*, *fry*, *roast*, and *sauté*), as well as six Armenian culinary verbs (*բովել [bovel]*, *եփել [ephel]*, *թխել [thxel]*, *խաշել [xashel]*, *շոգեխաշել [shogexashel]*, and *տապակել [tapakel]*) – are investigated with regard to their collocational distribution in the respective corpora, which are elaborated upon in separate chapters of this work mentioned above. The statistically significant co-occurrences of the investigated culinary verbs and the manual semantic annotation of the context examples

[8] For more information on FrameNet see Baker, Fillmore, and Cronin (2003); Baker, Fillmore, and Lowe B. (1998); Ruppenhofer et al. (2010).

extracted from corpora guide the reader throughout the work – from data collection and corpus description in the respective language to annotated data analysis and visualizations. The work concludes with the **References**, providing the list of works cited, followed by the **Data Management and Appendices**.

Chapter 1. State of the Art and Methodology

The initial objective of this work was to describe the culinary fields of Armenian, English, and German, relying on conventional logic-based lexical semantic analysis. Therefore, lexical field theory and frame semantics in general were considered highly relevant for describing the lexical, pragmatic, and cognitive properties of the culinary fields. However, as will be observed throughout this work, the dilemma of field theory and frame semantics, as well as the ambiguity of the lexical field itself, has resulted in going beyond conventional semantics toward distributional approaches to bridge the two. This integration aims to derive interpretable word profiles based on corpus evidence, accompanied by manual semantic annotation, as opposed to the vector space meaning elaborated upon further in this chapter.

The chronology of mainstream linguistics from the seventeenth to the twentieth century has left us with several prominent linguists who, among other things, made considerable efforts to set the boundaries between classical logic and philosophy on one hand and language on the other.[9] Representing the European structuralism, Wilhelm von Humboldt claimed that language "makes infinite use of finite means." He, in fact, is considered the precursor of Ferdinand de Saussure (1857–1913), who himself played a major role in the theory of lexicography and word meaning, focusing on the meaning of words that were not previously considered by Latin grammarians. Lexical field theory was developed as one of three main approaches to structural semantics, inspired by de Saussure (1916), who viewed language as a system and not a haphazard pile of words. Weisgerber[10] (1927) became one of the earliest linguists influenced by structural semantics as well as the philosophy of Wilhelm von Humboldt, viewing language not only as a system but also as a means of conceptual perception of the world. In describing structural semantics, Geeraerts describes Weisgerber as the most prominent researcher in the lexical field theory, with his idea of language constituting "a conceptual layer between the mind

[9] In this work I do not explicitly discuss the alternative approaches to meaning developed within formal semantic theory, which originates in the works of Frege (1892), Russell (1905), Montague (1970) and others, primarily focusing on phenomena beyond lexical semantics.

[10] By criticizing the historical-philological semantics, Weisgerber (1927) brings about the kinship names in German to state that the absence of names for the maternal brother and paternal brother does not depend on the psychology of mankind nor is it influenced by reality, therefore the lexicon of a language should be observed from an onomasiological rather than a semasiological perspective. In the semasiological perspective, first the linguistic expression is present, then the meaning is sought, while in onomasiology the meaning already exists, then the expression is thought. The structuralist view of word meaning also stressed the importance of shifting from the meaning of the individual sign to the system as a whole and the relations it enters into in the overall field. Thus, the meaning of the word strictly depended on the meaning of the words as a whole. Besides, according to Weisgerber, language description should be synchronic and not diachronic (cf. Geeraerts 2010 for a more detailed introduction to the chronological sequence of language description and different perspectives of the studies on word meaning).

and the world" breaking down the field into "conceptual plots," where the meanings of the constituents are mutually interdependent, thereby presenting "conceptual structure for a certain domain of reality" (Geeraerts 2010:52). Weisgerber's assertion that, in order to fully master a language and understand how words properly function in that specific language, one also needs external knowledge of the language community itself, was restated by the American anthropological linguists Edward Sapir (1926 [1960]) and further developed by Benjamin Lee Whorf (1940) in what became known as the Sapir-Whorf hypothesis. The latter, which is more widespread in the English-speaking linguistic community, claims that language greatly influences a person's world perception.[11] However, while Weisgerber laid the grounds of the lexical field theory, Jost Trier (1931) developed it further and fully formulated its underlying principles and standards.

Trier's (1931) concept of lexical field draws parallels between a mosaic image and the contents of human cognition. Only semantic relations are included within Trier's fields without formal relations.[12]

> …, daß man bei der Beschreibung von Wortbedeutungen und deren Veränderungen die Wörter nicht isoliert, sondern in Zusammenhang mit ihren Nachbarn betrachten soll. Der Gegenstand der Bedeutungsforschung sollen deshalb nicht Wörter, sondern ganze „Felder" sein.
> (Trier, qtd. in Leisi 1985:105)
> […that when describing the meanings of words and their changes, words should not be considered in isolation but in connection with their neighbors. Therefore, the object of semantic research should not be individual words, but entire "fields"]. (Translation mine)

Trier's image of the mosaic of lexical fields has been criticized from various perspectives. His lexical fields did not suppose any gaps. All parts of the mosaic covered the whole lexical field, with no parts missing (see also in Geeraerts 2010:65).

1.1 On the Ambiguity of the Term *lexical field*

The vagueness of the term *lexical field* both in the English- and German-speaking area has led to the interchangeable use of *lexical*, *semantic* and *conceptual field*. However, certain

[11] Ogden and Richards's (1923) model is also based on the ideas and conclusions of Weisgerber (1927). Ogden and Richards's Semiotic Triangle illustrates that the relation between the word and the thing is mediated through thought or reference. The meaning of the word is not necessarily identical to the objects in the world but is conveyed through personal associations of the utterer, the listener, etc.

[12] De Saussure (1916) envisages the description of the lexical field with both semantic and formal relations.

distinctions have also been attempted in an effort to separate them. Lyons (1977:268) distinguishes between a *lexical field*, consisting only of words, and a *semantic field*, including other-than-word units, too, such as idiomatic expressions, to cover a conceptual field. Lyons's distinctions overlap with Trier's (1934) *"Wortfeld"* and *"Bedeutungsfeld."*[13] Lipka differentiates *word field* with "morphologically simple items" from *lexical field* with "complex items next to simple ones" (Lipka 2002:168). His *word field* corresponds to Lyons's (1977) *lexical field*; however, Lipka also distinguishes between linear and hierarchic fields, the latter being based on one of the purely semantic relations, viz. hyponymy. Synchronic and diachronic lexical fields cover the same conceptual field.[14]

Porzig (1934) brings about the idea of lexical field and combinability: syntagmatic relations, e.g. *Gehen Sie oder fahren Sie nach Hause?* Helmut Gipper (1959) introduces the German *Sessel* (armchair) und *Stuhl* (chair) field, with its central and peripheral items, by identifying semantic features differentiating the various types of sitting items expressed in two words covering the semantic area of sitting in German. The core of the field constitutes the prototypical separate items for that category, while the peripheral ones exhibit more fancy, unusual, specific items. However, the boundaries of the core and the peripheral items in the field seem to be difficult to draw. Otto Ducháček (1959) introduces the star-like lexical field of *beauty* in French. At the center of the field, the French word *'beau'* ('beauty') is placed with beams to its surrounding fields, introducing the borrowed words that exist in the conceptual field of *beauty* in French. Nonetheless, Ducháček's representation of the star-like lexical field of *beauty* was also criticized, as the borrowed peripheral terms surrounding the core might, in fact, separately belong to two distinctly different fields. Pottier presents (1964, 1965) the field of sitting furniture in French by analyzing it with various semantic features distinguishing one type of sitting furniture from the other: *sèmes* making up the *sémème* of a lexeme. The drawbacks of the lexical field theory were then overcome by another methodology of structural semantics, viz. the *componential analysis*, marked by Adrienne Lehrer's (1969, 1972, 1974, 1993) remarkable contribution to the field, particularly her description of the culinary verbs in several languages, including German and English, as well as indigenous languages of the United States. Lehrer's (1974) extensive analysis on culinary verbs reveals the so-called lexical

---

[13] Trier's *"Bedeutungsfeld"* goes back to Ipsen (1924), as well as Jolles (1934) and Porzig (1934). According to Trier, the vocabulary of every language, regardless of the historical period, consists of "word-fields", in other words, lexical items with each item being separate parts of the mosaic picture, semantically fitting into the conceptual structure (Trier 1931, 1934, cf. Hanks 2013:352).

[14] Lyons describes a conceptual field "as a structure of concepts on the semantic level", while a lexical filed is "a set of lexical items to cover a specific conceptual field" (1977:253). "Lexical gaps occur when the coverage of the conceptual field by the lexical field is not complete" (Geeraerts 2010:56).

gaps, “which occur when a concept – that for reasons of systematicity seems to be a bona fide member of the conceptual field – is not lexicalized” (Geeraerts 2010:65). Lipka (2002) speaks about “lexical sets” which, on a paradigmatic axis, represent a lexical field based on “either association and intuition, or on objectively verifiable extralinguistic relationships captured by encyclopedic knowledge” (Lipka 2002:173).

A closer look at the classification of the German and English culinary verbs by Lehrer is provided later in this work, serving as the initial basis for bridging the logic-based approaches – field and frame semantics, componential analysis, prototype semantics (Putnam 1975a, 1975b, Rosch 1973a, 1973b) – with the distributional approach to language analysis (collocation meaning based on statistical significance and manual annotation).

1.2 Field vs. Frame Semantics (Fillmore, Lehrer)

The so-to-say father of frame semantics, Charles Fillmore. states in one of his renowned works, *Frame Semantics:*

> … that “with the term *frame* I have in mind any system of concepts related in such a way that to understand any of them you have to understand the whole structure it fits; when one of the things in such a structure is introduced into a text, or into a conversation, all of the others are automatically made available. (Fillmore 1982a:111)

Various authors have given different names for the term *frame* with a slight difference either in the structure or usage of the frame, viz. *schema*, *script* (Barsalou 1992, Schank and Abelson 1977), *frame* (Lakoff 1987), *cognitive model*, *scenario*, *global pattern*, *pseudo-text*, *experiential gestalt*, *base* (Langacker 1991, cf. Cruse and Croft 2004), *scene*, etc. Schank and Abelson (1977) first mentioned a *script* as a “structure that describes appropriate sequences of events in a particular context made up of slots and requirements of what can fill those slots” (1977:41). We encounter an identical, nearly word-for-word definition of a *frame* as a *script* also in Allan, who defines it as “structured representation[s] of event sequences” while contrasting it to *frames*, which “identify the structural relations of listemes and the concepts they name” (Allan 2001:247). Croft and Cruse see the *frame* as “a coherent region of human knowledge, or as a coherent region of conceptual space” (2004:14). Langacker illustrates the notion of a *frame* as a *base-profile* relation (Langacker 1991), using the word *radius*, which symbolizes the *concept* RADIUS, and placing it against the *base* CIRCLE. There does not seem to be a sharp and strict distinction between the bases, domains and frames, since Langacker uses the term base and domain interchangeably. Fillmore introduces the idea of a domain but also as a reference to what others call a *script, cognitive model, scenario* (Fillmore 1982a:111). However, Croft and Cruse seem to shed light on the distinction between a *base* and a *domain*. For example, if a base like CIRCLE defines a single concept profile, such as RADIUS (of course, with all the preconditions of our background knowledge of what the circle is), then it is considered a *base.* A base becomes a *domain* when it supports and defines multiple concepts, such as a diameter, a chord, etc., meaning that "several different concepts have it as a base". In this respect, a *domain* is then identical to the Fillmorian frame.

> A domain is a semantic structure that functions as the base for at least one concept profile. (Croft and Cruse 2004:15)

We encounter the notion of semantic co-occurrence within the frame in Geeraerts; however, this does not extend to syntagmatic or word-level co-occurrence.

> A semantic frame … is a coherent structure of related concepts where the relations have to do with the way the concepts co-occur in real world situations. (Geeraerts 2006:16)

Thus, the Filmorian *Frame semantics* comes to fill in the gaps of a "dictionary view of linguistic meaning," which is just a very limited part of the concept profile, or, as Croft[15] and Cruse quote Fillmore: "the frame semantic model of linguistic meaning highlights the failings of the dictionary view" (Fillmore 1982:134, 1986:233). Fillmore's *Frame semantics*, however, was much focused on the description of the conceptual rather than lexical semantics.

Traditional logic-based semantics would describe food frames with possible dictionary definitions allowing for limited coverage of meaning, followed by the intention and the

[15] Cognitive linguistics, however, according to Hanks does not distinguish between "normal and abnormal" use of the authentic language (cf. Hanks 2013:381). Hanks (2013) links his Theory of Norms and Exploitations (TNE) to Croft's (2000, 2004) approach to the meaning as a cognitive linguist: TNE's notion that a central feature of the language is the relationship between normal, conventional uses and innovation involving exploitation of those conventions finds an echo in the work of another cognitive linguist, William Croft (Hanks 2013:383).

extension[16] of meaning. A lexical field description of food items could be either hierarchic or non-hierarchic.[17]

Fillmore's, Barsalou's and Atkins' frame semantics is linked to Pustejovsky's 'lexical semantic structure' (Pustejovsky 1995).[18] In the theory of the *Generative Lexicon* (GL), Pustejovky identifies semantic types that are determined by the context. Following the onomasiological approach of naming an object based on its definition, lexical items are then attached to these semantic types. Pustejovsky sees the lexicon within the generative framework, where words have the potential to generate infinite and meaningful nuances of meaning in different contexts, despite the lexicon of a language consisting of finite lexical items (Pustejovsky 1995).[19]

The present work, therefore, attempts to link conventional frame semantics to the word meaning in actual usage by combining quantitative computational analysis methods and qualitative manual semantic annotation. This approach serves as a tested model on more or less *monosemous* words, i.e. the culinary verbs, which we treat as specialized vocabulary. This also envisions the potential of the model to be expanded to ambiguous words, provided the necessary annotation is performed. Culinary verbs, however, could themselves be ambiguous and/or have connotational meanings. For instance, the German verb *kochen*[20] also carries the sense of "sehr erregt, aufgebracht, wütend sein; einen hohen Grad an Erregung aufweisen," as in "der Volkzorn kocht" *(DWDS)* [to be very upset, furious, angry; exhibiting a high degree of temper, e.g. the fury of the nation boils*] (translation mine). Similarly, the English verb *cook*[21] has other meanings outside the culinary frame, i.e. "to falsify (information) with the aim of hiding a certain type of activity" (Antidote 9). Here, *monosemous* should then be restricted to the central meaning of the word. Thus, this work aims to contribute to bridging the gap between

---

[16] Extension of an expression (Carnap 1956), e.g. of *red* is the class of all red objects and intention is the property of being red. Lyons uses *denotation* to cover both the extension and the intension (Lyons 1996:208).

[17] "When two meanings of the same form belong to different lexical fields" we have homonymy (also Kastovsky 1982:123). When "the archilexeme fails, we have a linear hierarchy but it must have a common semantic feature, i.e. red, blue, yellow, and grey for color" (Lipka 2002).

[18] Pustejovsky et al. introduce in 2004 the methodology of Corpus Pattern Analysis (CPA) the central idea of which is the claim that word senses "are not directly encoded in the lexicon" but rather "each word is associated with one or more stereotypical syntagmatic patterns" which they call "selection contexts" (2004:1). The CPA Methodology thus identifies those selection contexts from large corpora, gathers them in selection contexts dictionaries where both syntactic and semantic stereotypic information is encoded.

[19] According to Hanks (2013), his Theory of Norms and Exploitations is strongly based on Pustejovsky's GL where words have central or core meaning(s) but infinite "combinatorial possibilities and meaning potential" (377).

[20] *"kochen"- 4.* sich auf Siedetemperatur befinden und dadurch seine Beschaffenheit verändern*, b)* umgangssprachlich, übertragen ⟨jmd., etw. kocht⟩, ⟨es kocht bei, in jmdm.⟩ sehr erregt, aufgebracht, wütend sein; einen hohen Grad an Erregung aufweisen (*„kochen", Wörterbuch der deutschen Gegenwartssprache (1964–1977)* retrieved on *03.07.2023, via DWDS ,*<https://www.dwds.de/wb/kochen>*).*

[21] "cook"- *informal*. to falsify (information) with the aim of hiding a certain type of activity ("Cook" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*

conventional logic-based semantic theories and projects such as FrameNet (https://framenet.icsi.berkeley.edu/fndrupal/) on one hand, and computational representations like word vectors (word embeddings) (e.g. *Word2Vec* Mikolov et al. 2013) on the other.[22] The number of word embeddings projects in NLP rises constantly; however, they mostly lack interpretability in that it is often impossible to trace back the semantic principles under the generated word vectors based on neural network approaches.[23] This aspect of linking conventional frame semantics to the word profile description, made possible by the latest developments in NLP and deep learning, is still seen as a research desideratum. For instance, despite giving credit to the endeavor of a comprehensive frame semantics model that attempts to include the semantic properties and syntactic "combinatory possibilities" of the word in all its senses in a frame, Hanks refers to the FrameNet project as:

> They [the frames] do not drive the theory. Frame semantics is not a lexically based or corpus-driven theory. FrameNet frames are based on speculation about frames in vacuo; corpus evidence is then adduced to support and modify the theoretical speculations. No attempt is made to analyze systematically the meanings or uses of any given lexical item. (Hanks 2013:388)

In the introduction of *Semantic fields and lexical structure*, Lehrer provides a preliminary definition of what a semantic field is, describing it as a "a group of words closely related in meaning, often subsumed under a general term" and notes that "the object of the semantic field is to collect all the words that belong to a field and show the relationship of each of them to one another and to the general term" (Lehrer 1974:1). In the case of the logic-based traditional description of food items, the objective of the work would have been the collection of all the words related to the food field from different sources of communication and show the relations in between them. Different types of relations would come to fill in the field of *food*, viz. hyponymy, meronymy, contiguity, and analogy (cf. Lipka 2002).[24] Lipka states that contiguity is by far the most important relation to analyze the distinctive features, or components, of meaning, "since it represents the relations between closely related meanings occupying a well-

[22] There is still a growing demand in future works both for word sense disambiguation (cf. Camacho-Collados and Pilevar 2018, Sun et al. 2021, Alian and Awajan 2023) as well as reducing vector space (Mikolov, Chen, et al. 2013).
[23] (Cf. Heřman 2021) for word embedding models preserving many of the semantic properties of words appearing in similar contexts tested on vector space models created for 15+ languages based on *TenTen* family of corpora.
[24] Hierarchies such as hyponymy "lead to the postulation of one type of lexical field" (Lipka 2002:153). According to Lipka "lexical gaps exist only with regard to word-field and not lexical fields because the latter has complex lexemes which can easily fill in the gaps", e.g. derivations (ibid.172).

defined, restricted semantic domain and exhibiting certain well-marked contrasts. That's to say, each meaning is distinctly set off from other related meanings by at least one important feature" (Lipka 2002:18). Similarly, Blank (1999) identify contiguity as the first and foremost semantic relation for describing a frame. By giving a more precise classification of contiguity types, Blank believes that all types of contiguities come from two big domains: *co-presence* and *succession*, which are based on the 'diachronic' and 'synchronic' models of human conceptualization, respectively. A frame activation is firstly rooted in the 'co-present' domain (Blank 2002:179). As Koch states, "there is contiguity between the elements of a frame, but also between the frame as a whole and each one of its elements" (Koch 2012: 261). What is proposed here is to reiterate the potential of describing food items with the help of the contiguity relation, both for the separate elements within one frame, and for the elements and the frame as a whole. Still, the question of authentic evidence derived from corpora remains open in this case. While contiguity is relevant not only for field semantics but also for frame semantics, it is particularly suitable for describing lexical items (primarily nouns) on a conceptual level and less suitable for verbs. That is also one of the reasons why this work does not look at the frames in the sense of Fillmore's frame semantics, as we do not strive to activate the different conceptual frames that culinary verbs might evoke, but rather to reveal the collocational behavior the words have in the respective language corpora, leaving frames on a more abstract level. The interlinguistic analysis in this work, which aims to identify possible translation equivalents based on the separate intralinguistic investigation of the languages involved, was carried out on the conceptual level to ensure comparability between German, Armenian and English. In addition, this work is also an attempt to bridge field and frame semantics through corpus-based analysis of word meaning by identifying collocational relationships with the lexical field that then activate the semantic frames, i.e. word profiles.

1.3 Reconsidering Lehrer's Classification of Culinary Verbs from the Componential Analysis Viewpoint of Field Representation

The first grains of the componential analysis are observed in the works of Hjelmslev (1953, 1958), who argued that linguistic structure could be described by pure linguistic relations. The analysis of meaning was then carried out in the form of oppositions, in what he called *content figurae*, e.g. *boy* as a *he-child* and *girl* as a *she-child*, *ram* as a *he-sheep* and *ewe* as a *she-sheep*

(Hjelmslev 1953:79). Componential semantics in Europe was further developed and fully articulated in the works of Pottier (1964, 1965), Coseriu (1964, 1967)[25] and Greimas (1966). The term *lexical encapsulation* was introduced by John Lyons to denote "the lexicalization of [a] syntagmatic modifying component" such as *kick with a foot* (Lyons 1977:262).

In criticizing Fillmore for underestimating semantic field theory, Lehrer (1993) states that both theories are applicable depending on the type of data to be investigated. Moreover, the drawbacks of one theory seem to be compensated by the other. Drawing parallels between the two theories, Lehrer argues that both frame and field theorists do the same thing, that is filling up the frame or the field with relevant items (the former more conceptual oriented, the latter – lexical) but using methodologically different approaches. In order to analyze a frame, the frame theorists first pick up "a conceptual domain" and then examine how a given language lexicalizes/verbalizes ideas within that domain. The field theorists do the contrary: they first compile "an inventory of expressions in the field," then analyze it by trying 1) either to link them to the conceptual domain, 2) or to combine the lexical items by means of semantic formal relations, such as hyponymy, synonymy, etc. While frame theorists do not see a sharp line "between knowledge of language and knowledge of the world - in other words - a dictionary and an encyclopedia," field theory aims at finding "the list of lexemes as a resource for expressing meaning" (Lehrer 1993:149-154). Frames usually allow multi-word expressions, whereas fields rather tend to focus on single words. The table below summarizes the comparison of the frame and field theory according to Lehrer (1993).

Frame theory	**Field theory**
Might include 'cooking on the table' in the *conceptual* domain of cooking since such way of cooking exists (outside the language).	Using the example of a 'meat fondue', the field theory would not include 'cooking on the table' (as a type of cooking, as opposed to 'cooking inside on the stove, in the oven', 'cooking outside' like 'barbecue', because

[25] Like Porzig (1934), Palmer ([1938] 1968), and Firth (1957), Coseriu as well believed in the meaning of the word in its co-occurrence with other words in the same context, however, Coseriu and Porzig were interested in the meaning of separate lexemes. Comparing the Romance languages, Coseriu singled out marked and unmarked lexical items (semantic features/terms), i.e. of color terms inherited from Latin but having lost their markedness in modern Romance languages ('niger'-shiny black vs. ater 'just black') (Coseriu 1964). However, neither of them were interested in the meaning of the constituents of collocations. This gap is attempted to fill in by Bartsch (2004) in pursuit of describing the "functional and structural properties of collocations".

	the English language does not have such a way of cooking.
First, a conceptual field which then is filled up with lexical items.	First, lexical items (inventory) making up a semantic field.
	Field theorists select fields that are "heavily and densely lexicalized" (Barsalou 1992 qtd. in Lehrer ibid. 154), containing a large number of single words.
	"Knowing the meaning of an item involves knowing a whole set of semantic entailments" (ibid. 153).
	Field theorists look for the range of meaning of the word and are often influenced by the contrasting words. (E.g. to analyze "n. *scratch* - a mark or small wound made by scratching," one needs to know *gash*, *slash* for large wounds and *puncture* for deeper wounds, since, in the opinion of field theorists these words (gash, slash, puncture) play a role in the range of meaning of *scratch*.

Thus, *componential analysis* focuses on the description of the lexical field, emphasizing the relations within the fields based on mutual oppositions, but the oppositions should have been determined and analyzed in detail.[26] In introducing the componential analysis method of describing the lexical fields and the studies in word meaning, Geeraerts (2010) summarizes the whole essence of this method:

> The fundamental notion of these studies [componential semantics] is precisely the idea that lexical elements in a field are distinguished by functional oppositions. (Geeraerts 2010:75)

[26] For instance, Weisgerber's (1927) description of the lexical field of kinship relationship oppositions was based on maternal or paternal lineage, gender, and/or different generations.

In her *Cooking vocabularies and the culinary triangle of Lévi-Strauss*, published in Anthropological Linguistics, Vol. 14, No. 5 (1972:155-171), following her previous work *Semantic Cuisine* (1969), Lehrer presents the cooking terms (cooking arts, techniques) of some Indo-European and non-Indo-European languages, e.g. English, German, and French, as well as Mandarin Chinese, Japanese, and Yokuba (Nigerian dialect). The componential analysis applied further by Lehrer in the same article presents seven binary and three non-binary components that seem to cover the overall culinary vocabulary in terms of cooking techniques in the described languages. The components are introduced below to a) discuss Lehrer's classification of the German and English verbs within the interests of this work, and to b) illustrate how they served as the initial basis of the manual semantic annotation. These components were selected as a starting point for this work, however, some were excluded, while others were added based on the corpus evidence. Below are the components of the analysis of culinary verbs according to Lehrer's classification (1974: 157-160).

1. [±Water], to indicate that *water* (or any type of *liquid*) is or is not used during a certain type of cooking.
2. [±Fat], to indicate whether or not any type of *fat*, be it *butter*, *oil*, or *margarine*, etc. is used in the cooking process. *Fry*[27] with all its hyponyms is stated as [+Fat]. The rest is [-Fat].
3. [±Direct heat] is the next component in Lehrer's list, which comes to indicate whether *direct* or *indirect heat source* is used for this or that cooking technique. This component does not apply to all the culinary words but is rather limited to certain lexical items. [+Direct heat] may be applied to "direct or radiated heat source," e.g. cooking under the *broiler* or on *open fire*, while [-Direct heat] applies to "indirect or conducted heat source," e.g. dishes cooked in an *oven*. Examples for direct heat cooking would be the English *fry*, *grill*, for indirect heat the English *bake*. Even though Lehrer suggests that for some languages, for instance, German, it would be more efficient to use only two non-binary components [+Broiler] for *fry and grill* and [+Oven] for *bake*. It is obvious that Lehrer's suggestion is grounded on her classification of German cooking terms into *kochen*, *braten* and *backen*. For this part, only *braten* and *backen* are of interest here. While *rösten*, *grillen* and *toasten* fall somehow – though not very clear-cut – into the category of *braten*, with *backen* standing all alone. And since *backen* can be done in

[27] The italics are mine to single out the verbs and the components.

any kind *oven*, Lehrer's suggestion to leave only [+Broiler] for *braten* and its hyponyms, and [+ Oven] for *backen* makes sense. Another question is whether we would then use the verb *backen* for the chicken prepared in an *oven* or rather *rösten* would be more appropriate.

4. [±Vigorous cooking action]. According to Lehrer, this component only applies to *boil* with its hyponyms. "Boil and steam are marked [+Vigorous] while *simmer, poach, stew*, and *braise* are [-Vigorous]. This component does not apply to other words," states Lehrer (1972:158). However, there is no such verb as *steam* in Lehrer's classification of English cooking words.
5. [±Long cooking time]: Lehrer herself argues that long and short cooking are rather relative distinctions; what matters here is actually how slow and how quick the dish can be done. *Parboil* is [-Long time]; *stew* is [+Long time]. Other terms are unmarked.
6. [±Large amount of cooking substance] to indicate that for instance something is cooked in a *large amount of fat* (*deep fry*) and in *small amount of fat* (*sauté*).
7. [±Submerged] to indicate whether the food is submerged in the substance in which it is being cooked. Nonetheless, Lehrer avoids calling this as a real component since it does not have a counterpart, such as *not submerged in fat*. According to her, [±Submerged] is rather "a presupposition" stating that "*steam*[28] is [-Submerged] … while boil and its hyponyms are [+Submerged], *fry* and its hyponyms are also [±Submerged]. I am not sure here that *sauté* is [+Submerged] since there is neither a lot of *fat* nor *water*. It may be instead [-Submerged], as the term submerge usually means: 1. Cause (something) to be under water (OED), 1.1. Descend below the surface of an area of water (OED). However, if *submerged* refers to the idea that the food is cooked directly in the substance no matter whether fully (*boil* with its hyponyms) or partially (*fry, sauté*), then Lehrer's distinction is clear. For reference, Antidote gives the definition of *sauté* as "to fry (something) for a small time — to sauté onions and mushrooms".

Lehrer introduces three additional non-binary components. These are:

8. [Special kind of utensil]. Lehrer identifies some possible utensils that mark this or that way of cooking; for instance, *braise* presupposes [+Covered pot], while the component [+Hot coal] could be used for *charcoal-broil* and *barbecue* (the examples are mine).

[28] It is rather interesting to observe that Lehrer uses the cooking technique of *steaming* so many times in her comparisons, but the verb does not appear in her classification table of English cooking vocabulary. Or perhaps *to simmer* was initially intended to be *to steam* and somehow failed to be so.

9. [Special ingredient used or food]. This component is not clear-cut. Lehrer only mentions *barbecue* [+Barbecue sauce] as being an optional component. Similarly, the French *gratiner* could be considered [+Cheese].
10. [Special purpose intended by the cooking process]: "*Poach* has the component [+To preserve shape (of food cooked)] and *stew* the component [+To soften]" (Lehrer 1972:157-159).

In summarizing her analysis results, Lehrer noticed among other things that there are more significant and less significant parameters to distinguish culinary verbs. For instance, *Special kind of utensil* would not be so important to differentiate cooking verbs, while *Heat source* and *the presence* or *absence of water* or *fat* would be decisive:

> …the primary parameters for distinguishing among the main culinary terms are the use of water or fat and whether the heat is direct or radiated. (Lehrer 1972:166)

However, later in the respective chapters of this work, corpus evidence shows that for each verb certain parameters are decisive and distinguishing, even though *Substance* plays a major role in differentiating the culinary verbs from one another.

In addition, Lehrer claimed that her data did not fully support all points of the *Culinary triangle* and states that *baking*, instead of *smoking*, could have been more appropriate based on the languages she investigated.

In order to get a deeper understanding of the components defining the cooking procedures suggested by Lehrer, a brief introduction of the *Culinary Triangle* (originally: *Le triangle culinaire*) of Lévi-Strauss (Lévi-Strauss 1965) (cf. Fig. 1 and Fig. 2) should be given here, however, not with the intention to argue for or support it.[29] Lévi-Strauss developed the culinary triangle in analogy with the phonological triangle of Jakobson and Halle (Fig. 1), labelled as "patterning of oral resonance features" and representing the vowels and consonants on the basis of two opposition pairs, i.e. "the grave/acute consonantal axis and the compact/diffuse vocalic axis," originally derived from the initial triangle comprising only the sounds *p*, *a*, *t* (Jakobson and Halle 1971:53).

[29] The *Culinary Triangle of Lévi-Strauss* was probably mostly elaborated in the works of Lehrer (1969, 1972), but it also served as the basis for a number of works both in the field of cooking vocabulary studies (Vilgis 2013), as well as anthropological, ethnical and sociological investigations (Eberle 1984).

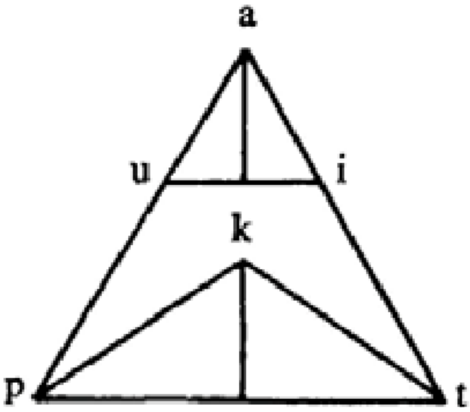


Fig. 1. The phonological triangle
by Jakobson and Halle (1971:53)

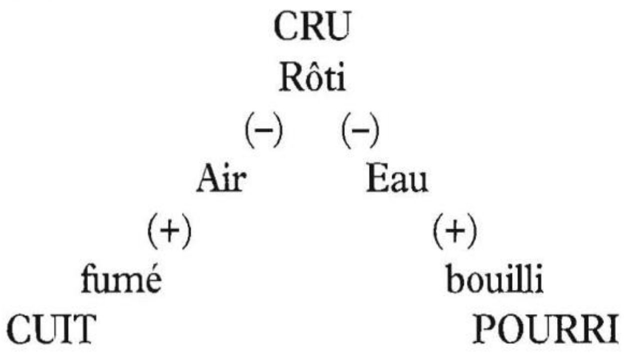


Fig. 2. The Culinary Triangle (Le Triangle Culinaire) by Lévi-Strauss (1965:17)

According to Lévi-Strauss, the *Culinary Triangle* divides such a universal domain as cuisine into *unmarked raw* (fr. *cru*) at the top and *cooked* (fr. *cuit*) and *rotten* (fr. *pourri*) at the left-hand and right-hand angles, respectively, where *cooked* is the consequence of *cultural* impact and *rotten* is the result of *natural transformation*, e.g. fermented food derived from the raw that exists in *nature*. Lévi-Strauss himself emphasizes, however, that the definition of being *cooked* or *rotten* is not clear-cut at all, stating that Italian cuisine eats *raw* more than French cuisine, yet the latter has more verbs to denote the process of cooking (1965: 17). The triangle also comprised derivational adjectives, i.e. *fumé* (smoked), *rôti* (roasted), and *bouilli* (boiled) with components such as *air* (air) and *eau* (*water*) to represent the state of the *raw* and its transformation.

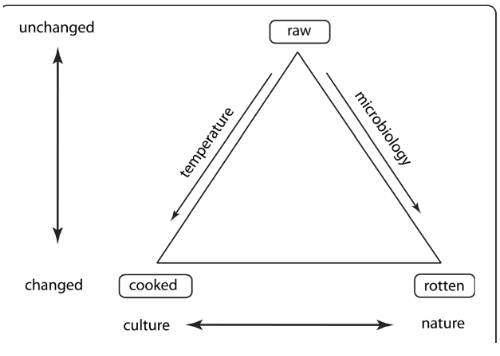


Fig.3. Vilgis (2013:3)

In reiterating that cooking and eating remain primarily cultural, Vilgis revisits the Culinary Triangle of Lévi-Strauss in his article *Texture, taste and aroma: Multi-scale materials and the gastrophysics of food*, discussing the common components for cooked and raw foods, i.e.

proteins, carbohydrates, fat and water, which define the structure and texture of food depending on the different degrees of solubility of these components, followed by a hierarchical structure of *food constituents*, where the aforementioned components determine the texture/structure, while the aroma defines the taste of the cooked food, with the perception of the two ensuring the *flavor* (Vilgis 2013:1–2). Fig. 3 illustrates the interpretation of the *Culinary Triangle of Lévi-Strauss* by Vilgis, where the "transformation from *raw* to *cook* is temperature-driven, from *raw* to *rotten* by microorganisms (ibid. 3).

Lehrer suggests substituting *baking* for *smoking* in order for the *Culinary Triangle of Lévi-Strauss* to be more or less representative of different cuisines.[30] By inserting the verb *frire* (fry), Lévi-Strauss then extends the triangle to tetrahedron in his endeavors "to fit" the French culinary verbs determined by the interactions of the substances – or the absence of any substance involved in the process. Even though it does not appear in this geometrical form in the original article by Lévi-Strauss, he describes where the verbs could be situated in the tetrahedron, highlighting the need to change it in case a new cooking substance needs to be inserted (Lévi-Strauss 1965:17). This adaption allowed for further research, especially among French ethnographers, to interpret the tetrahedron (Dournes 1969:44). Within the scope of this work, the illustration of the culinary tetrahedron by Lehrer, presented through the translations of the original French verbs, might shed light on the classification of the German, French, and English culinary verbs she presented in *Cooking vocabularies*. This also served as the preliminary basis for the annotation parameters underlying this work (Lehrer 1972:169) (cf. Fig. 4 and Fig. 5).[31]

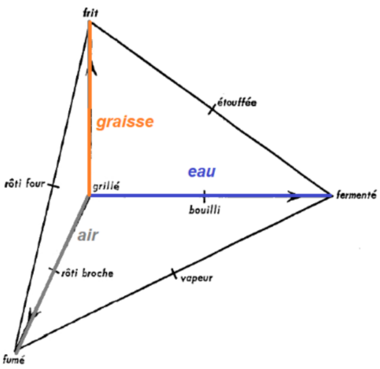


Fig.4. The Culinary Tetrahedron (Dournes 1969:44)

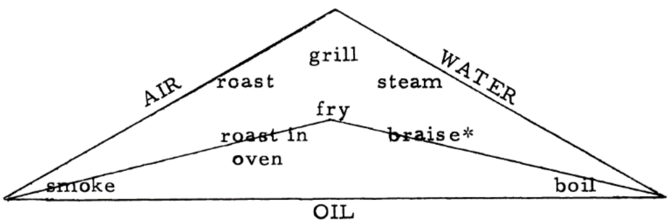


Fig.5. The Culinary Tetrahedron (Lehrer 1972:169)

[30] However, since there is no separate verb as *bake* in French, the triangle might be suitable for the French culinary field but not for the universal field of cuisine. *Faire cuir* (en. *make cook*) comes to denote the process of *baking* which could be investigated in further corpus-based studies yielding evidence to contribute to the *Culinary Triangle*.

[31] Lehrer notes that the original *étouffée* (*sauté**) is then changed in her interpretation of the tetrahedron (Lehrer 1972:169).

Among more recent studies on the research of cooking vocabulary, Coste's work on the comparison of English and Romanian culinary verbs should be noted. The objective of this study was to reveal the translation gaps, the *lacunae,* and as a result, providing translation techniques, e.g. hyponymy relations, primarily based on Lehrer's classification and Baker's (2011) translations techniques (Coste 2013). As stated in the *Introduction* of the present research, intralinguistic investigations and contrastive studies are of great importance for translation purposes, and this is one of the aspects emphasized in the present work, which aims at identifying translation models to achieve possible equivalence in the languages being compared. This statement is obviously not new; however, the proposed methodology of bridging the conventional logic-based and distributional semantics will contribute to further advancements in the field.

Based on the 4th and 5th century old German and French cookbooks, Wurm (2007) focuses in her dissertation on the historical aspect of the translation problems in culinary texts, highlighting the importance of cultural transfer in this type of specialized texts. Recent interest in corpus-based translation has resulted in a number of successfully published works, e.g. Kruger (2013), which comprises contributions to both corpus studies in general as well as methods of qualitative analysis for contrastive purposes based on parallel and comparable corpora with the objective of achieving equivalence in particular. Sánchez Cárdenas (2016) emphasizes the importance of comparable corpora for translation and teaching purposes in her work devoted to the translation of adverbs in medical specialized text. Her experiment, carried out with undergraduate students, showed that after compiling comparable medical corpora in English and Spanish, followed by training sessions of analyzing the corpora with tools such as *Sketch Engine,* the participants, having familiarized themselves with corpus analysis, are no longer inclined to simply select the first equivalent offered by the dictionary. Instead, they now tend to reflect on the context and consider more nuanced translation options and the "contextual and pragmatic properties of adverbs" (Sánchez Cárdenas 2016:195-217).

Drawing from a small corpus of German original narrative texts (282,739 words) and their translations into Spanish and Catalan (297,922 and 311,047 words respectively), both part of the COVALT (Valencian Corpus of Translated Literature) corpus, Oster and Molés-Cases (2016) identify the translation techniques employed for a) "food-related culturemes"; b) "actions carried out while eating or drinking, i.e. sipping"; c) "metaphorical expressions in which food is the source domain, i.e. bitterness" (Oster and Molés-Cases 2016:54). The classification of translation techniques in this study singles out “foreignization” and “domestication” as the two prevailing techniques applied by the translators (ibid. 70).

The same COVALT corpus served for another study taking into consideration the translation of food names between English and Catalan. Using the parallel sub-corpus of English (36 source text, around 1,2 million words) and their translations in Catalan (around 1,3 million words), Marco (2019) extracts equivalents of meal names, e.g. *breakfast*, *lunch*, *dinner*, *supper*, *tea* in British English and their counterparts/correspondences in Catalan, identifying considerable mismatches between the source and target languages. For instance, the British *breakfast* (very early in the morning) was translated as *desdejuni/desdejunar* three times more frequently than *esmorzar*, even though this solution would have been different, i.e. *esmorzar* prevailing over *desdejuni/desdejunar*, if the translations had been published in Barcelona and not in Valencia, reflecting regional differences in the timing and nature of breakfast. The experiment then singled out translation techniques based on the extracted equivalents of meal names in English and Catalan, highlighting that culturally saturated terms often necessitate adaptation. In particular, the transference of lexical items in the food domain in such culturally saturated texts often relies on “cultural adaptation, i.e. domesticating, for instance, translating the English meal name *tea* as *berenar* (a light afternoon meal typically consisting of a sandwich and some kind of cake or snack) as the counterpart does not exist in the target culture” (ibid. 48).

Wine-tasting metaphors and the challenges of their transference into target languages are discussed from the cognitive translation approach in Demaecker (2017). While proving that the application of the existing traditional theories for translating metaphors is not appropriate for wine-tasting metaphors, Demaecker analyzed wine metaphors through Mandelblit’s (1995) hypothesis, which involved examining the source and target language domains separately to identify potential equivalences in French and English wine-tasting metaphors.

The transference of food items treated as culture-specific, being a part of material culture (Newmark 1998), can often be challenging itself; however, when it comes to achieving equivalence in the collocations in culinary texts, the translator is obliged to treat the collocations as an inseparable unit, otherwise the word-for-word translation of the constituents would often result in unusual/atypical word combinations in the target language. In her book, *Fachsprachliche Kollokationen: ein übersetzungsorientiertes Datenbankmodell Deutsch-*

Spanisch, Cedillo (2004) argues that collocations have to be considered as units of translation depending on the type of the text, thus treating them partially as terms,[32] stating:

> „Sucht man nach äquivalenten Ausdrücken in anderen Sprachen, z.B. Spanischen, so kann man beobachten, dass keine von diesen Beispielen (‘Kaffee trinken’, ‘den Tisch decken’, etc.) wortwörtlich übersetzt werden kann, da im Spanischen dies zu unüblichen, untypischen Wortverbindungen führen würde, die die spanische Norm verletzen, wie z. B. *beber café für tomar café, *cubrir la mesa für poner la mesa. (Cedillo 2004:32)
> [In searching equivalent expressions in other languages, e.g. Spanish, one could observe that none of these examples could be translated word-for-word as it could lead to such unconventional, atypical word combinations in Spanish that would break the norm of the Spanish language, for instance *beber café für tomar café, *cubrir la mesa für poner la mesa].

Therefore, the translation of collocations into the target language does not necessarily need to carry the same meaning but should instead convey “the same communicative value” (“den gleichen kommunikativen Wert”) as the original does (ibid. 28).
As far as the question is concerned whether it is more difficult to translate the collocation from the mother tongue into the foreign language or vice-versa, opinions diverge among the linguists working in the field. For instance, Hausmann (1988) argues that, due to the unpredictable character of collocations in the target language, the translation of collocations is generally more problematic when equivalence is sought from the mother tongue into the foreign language. The other way round is less problematic, as collocations are “transparent for the foreign-language expert and are easy to encode”:

> Kollokationen bereiten vor allem Probleme bei der Hinübersetzung, da „sie bei der Enkodierung in einer Fremdsprache unvorhersehbar sind.“ Bei der Herübersetzung – weniger problematisch, weil „sie für einen Fremdsprachler durchsichtig und einfach zu kodifizieren sind”. (F.J. Hausmann 1988:139)

[32] Similarly, Friedrich addresses collocations as terms in specialized texts in his corpus-based study on German-American double taxation, highlighting the importance of the “learning and understanding” phases to be able to translate (Friedrich 2020).

Bergenholz and Tarp (1994), on the other hand, highlight the importance of *identifying* the collocations in the foreign language as the first step; only after that can translators suggest typical collocations in their mother tongue (1994:411-412).[33]

Thus, the need for a consistent methodology to identify possible translation equivalents of collocations in general, and in specialized texts such as the cooking recipes in particular, remains a gap irrespective of the current growing interest in corpus-based studies. Field and frame descriptions are often carried out on the conceptual and not on the word level, often accompanied by invented examples to fit into the frame (e.g.FrameNet). Vector spaces identify juxtaposition of words; however, they lack interpretability as the co-occurrence of these words might be merely accidental due to high frequency without necessarily being part of the same word profile (cf. Desagulier 2017). This work comes to fill in the gap by describing and interpreting the meaning of words while creating word profile comparable to vector space semantics, yet with a distinctive feature. The combination of computational data analysis based on statistical significance and manual semantic annotation ensures the interpretability of the identified word profiles and also contributes to identifying possible translation equivalents of food collocations in the languages being compared.

1.4 Conventional Collocation (F.J. Hausmann) vs. Collocation in Corpus Linguistics (J.F.Firth)

The shift from the paradigmatic description of lexical fields, without underestimating its importance, brought about the rise of syntagmatic analysis, which in turn highlights the significance of distributional semantics, introduced by the London school of Linguistics with John Firth's famous quote, which can be practically encountered in almost every work devoted to the distributional approach of analyzing lexical fields: "You shall know a word by the company it keeps" (Firth 1957b:11). Influenced by the works of early twentieth-century anthropologist Bronisław Malinowski, who himself was a follower of de Saussure, Firth introduces the term *collocation* and *meaning by collocation*, referring to combinability that functions also outside the sentence structure, i.e. *to smoke a cigarette, a pipe, weed*, etc. (Firth 1957a, 1957b). Even though the terms *collocation* and *collocability* are generally associated

[33] In her dissertation Katrin Berty (2017) illustrates the translation of collocations from the perspective of the knowledge of collocations in the target language and the importance of integrating it to the translation methods as collocations present a much bigger problem when translated from the foreign language into the mother tongue than the other way round.

with Firth's name, Harold Palmer was likely the first to have mentioned the term *collocation* in the introduction of his work *A Grammar of English Words*, defining it as "a succession of two or more words that may best be learnt as if it were a single word" ([1938] 1968:x). [34]

> I propose to bring forward as a technical term, meaning by *collocation*, and to apply the test of *collocability*. (Firth 1957a:194)

The notion of *collocability* comes from Katzian *selection restriction* (1963) and Whorfian *cryptotypes* (*covert grammatical categories*) (1956); however, Firthian collocation theorists were interested in describing the syntagmatic relations between linguistic items within the same contextual environment, also shaped by social contexts, going beyond the purely linguistic facts. Parallel to the Firthian term *collocation*, the study of words situated near each other has given different names to the same notion, each emphasizing different aspects of *collocability*. Katz and Fodor (1963) spoke of "selection restriction and projection rules," influenced by Porzig (1934), who named the syntagmatic relations between words *wesenhafte Beziehungen*. Coseriu (1967:296) coined the term *lexical solidarities* (*lexikalische Solidaritäten*), focusing on the "intrinsic relations between lexemes"[35], while Weinreich (1966) mentioned the 'transfer features'[36] and criticized Katz/Fodor's *selection restriction*.[37] However, in his paper *A Synopsis of Linguistic Theory*, Firth remarks that collocation should not be confused with *context*, "by which the whole conceptual meaning is implied," nor should it be taken as *citation* from the lexicographers, who only provide the meanings of words that they have found and support their definitions with citations for the arbitrarily selected *dictionary meaning* (Firth 1957b). He then goes on to say that the "habitual collocations […] are quite simply the mere word accompaniment, the other word-material in which they are most commonly or most

[34] Palmer viewed collocations more expanded as F.J. Hausmann by accepting the possibility of the collocation to be consisting of more than two words. F.J. Hausmann's concept of collocation is elaborated upon later in the text.
[35] Coseriu distinguished among three types of lexical solidarities, "depending on whether a classeme, an archilexeme, or a specific lexeme function as a distinctive feature", i.e. 1. Affinity, 2. Selection, and 3. Implication (Coseriu 1967:296, cf. Kastrovsky 1982:87).
[36] Cf. Tournier's (1985:229) "sème de transfert" and Cruse's (1986:105) "semantic traits." In criticizing the "selection restriction" of Katz and Fodor, Uriel Weinreich (1966) brings the example of the *He was drinking carrot* and marks it as "carrot" [+solid], <-solid> as the "transfer feature" <-solid> is transferred from the verb *drink* to its grammatical object *carrot,* speaking no "selection restriction" in the combinability of *drink and carrots* is present. As a result of transfer process, the object of carrots is reinterpreted as "carrot juice" (qtd. in Lipka 2002:129). These are already syntagmatic relations and not purely syntactic.
[37] Cf. Kastrovsky (1980, 1982) for a detailed comparison of Katzian selection restriction and Coseriu's lexical solidarities.

characteristically embedded" (Firth 1957b:12).[38] By comparing the words cows, lionesses and tigresses, Firth asserts with high certainty that the meaning of "giving milk," "somebody milks it" differentiates cows from the other two animals, as the "meaning of cows is indicated by such collocations as *Cows give milk* or *They are milking the cows*" (Firth 1957b:12).

The Sinclairian tradition of collocation analysis brings about the well-known COBUILD project, which is referred to in almost every work dealing with lexicography in general, and collocation analysis in particular, until the recent flourishing movement of corpus-based linguistic analysis making use of statistical significance. With a deep acknowledgement for the COBUILD project that resulted in the *Collins Cobuild English Language Dictionary* (1987), it is worth mentioning that Sinclair himself continuously accentuated the importance of "statistically significant collocation analysis," which at the time could not be carried out due to the lack of advanced software programs as well as hardware capacities (Sinclair 1987, 1991, 1998, 2004).

Based on the Sinclairian tradition, Hunston and Francis (2000) promoted the pattern grammar, which nonetheless indicates a pattern in its rather conventional sense (cf. Hornby 1954). It consists of word classes and functions words; however, collocations within the framework of pattern grammar were not grounded in statistical significance. As a result, pattern grammar did not allow for the distinction between two different senses of one of the same word, e.g. *fire an employee* and *fire a gun*, since they both share the same grammatical pattern of a transitive verb (*fire*) followed by a noun, i.e. V+N.

1.4.1 On the Structure of Collocations

Already in the works following Hausmann (1984, 1988, 1989, 2003, 2007), attempts are made to define the structure of collocations by determining the number of their constituent parts as well as differentiating between lexical and grammatical collocation. According to Hausmann, collocations are in general *binäre Einheiten* [binary units]; however, two collocations might collide and give rise to what he calls *Tripelkollokationen* [triple collocations], such as the separate collocations 'Kritik üben' [exercise criticism] and 'massive Kritik' [massive criticism], which would give rise to the *Tripelkollokation* 'massive Kritik üben'. This is seen in opposition to *clusters*, which are "nicht-zufällige Zwei-, Drei, Vierer-Gruppierungen" [non-random one,

[38] Already back in 1957, Firth spoke of word-embeddings which are the basis of the current vector space meaning (cf. Desagulier 2017).

two, or tree, word groupings or (clusters)], such as ‘at the hotel’, ‘to the hotel’, ‘went to my hotel’) (Hausmann 2003:318-320; cf. Mitchell (1975) defining *colligation* in the sense of syntactic collocability based on grammatical juxtaposition).

Thus, Hausmann views collocations as more or less fixed binary word combinations which, however, could be expanded as a result of the collision of two separate collocations, whereas clusters are mere sequences of words that are grammatically and syntactically, but not semantically, constrained. Throughout the present work, especially in the description of the German culinary verbs (Chapter 2), triple-structure collocations in the sense of Hausmann’s *Tripelkollokationen* are observed, e.g. ‘auf kleiner Flamme glasig dünsten’ resulting from combining the collocations ‘auf kleiner Flamme dünsten’ and ‘glasig dünsten’.

In his earlier definitions of what is a collocation, Hausmann defines them *as Zweierkombinationen von Wörtern* [two-word combinations] (1988:148), consisting of a *base* and a *collocate*, where the *base* is the dominant constituent, as it could be defined, learned and translated outside the context, while the *collocate* is always subordinate and does not possess the same properties of the *base* (Hausmann 2007:218). In *Le Dictionnaire de Collocations* (1989:1010), Hausmann introduces the typical structures of (mostly) binary collocations.[39]

a) noun + adjective (epithet)
b) noun + verb
c) verb + noun (object)
d) verb + adverb
e) adjective + adverb
f) noun + (preposition) + noun

In this work, we are primarily interested in the *c) verb + noun (object)* type of collocations.

According to Hausmann only – or rather mostly – nouns can serve as a *base* of a collocation. Verbs as a *base* are very rare, like ‘bellen’ [bark], which can be defined out of the context, meaning that one can understand it without the collocate ‘Hund’ [dog]. Even though Hausmann then recalls Coseriu’s *lexical solidarities* by stating that ‘bellen + Hund’ (verb + noun) types of collocations are very rare and can only occur if first we determine what ‘bellen’ is– namely the sound dogs make. A *base* is a banal word (‘Aufzug’, ‘Passagier’), while the *collocate* may vary from being a less idiomatic word (‘Aufzug holen’ vs. ‘Taxi rufen’) to a very idiomatic one (‘blind’ in ‘blinder Passagier’). A *collocate* may also be an idiomatic expression (*Redewendung*). Hausmann provides the example of ‘einen Streit vom Zaun brechen’, which

[39] Hausmann, F.J., however, mentions that a *base* can have several collocates at the same time (2007:218).

actually means ‘to provoke, to pick a quarrel’ and has nothing to do with a ‘Zaun’ [fence] (Hausmann 2007:218).

In *Was sind eigentlich Kollokationen?*, Hausmann (2003) criticizes Duden 2 (*Stilduden*), stating that it fails to list proper collocations under the lemma entry of the *base*. For example, in the collocation ‘Tisch abräumen’, where *Tisch* is the *base* and *abräumen* is the *collocate*. the collocation is rather hidden under the collocate *abräumen* (2003:310). The *Oxford Collocations Dictionary for Students of English* (2003), on the contrary, is highly appreciated by Hausmann, as here one can directly find a vast variety of collocations under the *base* itself. In other words, one finds possible collocations of *table* under the lemma *table* and not under *clear* when one looks for ‘to clear the table’. However, while not excluding triple structure collocations resulting from the *collision* of two collocations, such as ‘massive Kritik üben’ mentioned earlier in this chapter, Hausmann remains loyal to the idea of the binary character of collocation constituents, where the *base* (noun) is dominant while the *collocate* is subordinate. This highlights the independence of the *base*, which can be defined, learned and translated outside the context.

> Kollokationen (Beispiele: ‘confirmed bachelor’, to lay the table, ‘célibataire endurci‘, ‘mettre la table’, ‘eingefleischter Junggeselle’, ‘Tisch decken’) sind normtypische phraseologische Wortverbindungen, die aus einer Basis und einem Kollokator bestehen. Die Basis ist ein Wort, das ohne Kotext definiert, gelernt und übersetzt werden kann (*bachelor, table*). Der Kollokator ist ein Wort, das beim Formulieren in Abhängigkeit von der Basis gewählt wird und das folglich nicht ohne die Basis definiert, gelernt und übersetzt werden kann. (Hausmann 2007: 218)

According to Hausmann, attempts to describe collocations starting from a collocate and working towards the base are not effective unless carried out with the purpose of collecting classes of bases (Basisklassen) for the collocate. For instance, in cases of distinctive synonymy, e.g. *rufen* vs. *holen* (‘Taxi rufen’ vs. ‘Aufzug holen’), classes of bases (nouns) are gathered to show that not all synonymous verbs can be collocated in the same meaning (ibid. 2018).

In the framework of this research, however, the verb + noun (object) type of collocations, where the verb is the *base* (with the term *node* being equal to the *base* of collocations in our corpus-based approach) and the noun is the *collocate*, prove that verbs too could be dominant determining their collocates. This has been tested on a specialized corpus of culinary texts with more or less ambiguous words, i.e. the culinary verbs.

More recent definitions of collocation are nevertheless based for the most part on Hausmann's conventional *base + collocate* constituents of collocation. However, with the rise of corpus-based studies, collocations are now viewed from different perspectives and identified, for example, based on statistical significance. The difference is that statistical significance scores are derived from word counts in a specific corpus and are represented on a continuum scale, whereas the linguistic definition of collocates relies on linguistic intuition criteria and would be more or less categorical. For the statistical scores, the challenge would be to define some thresholds of word association strength (for different measures), where a collocate candidate also has a high chance of being intuitively classified as a true collocation in the linguistic sense. Yet this raises a philosophical question about whether linguistic intuition is the ground truth and mathematical association measures approximate it, or whether linguistic intuition itself is shaped by our own 'corpora' of words heard or read over our lifetime, causing us to approach borderline cases exactly in the same vague way we interpret the gray areas with lower values of association measures. This philosophical question is related to the distinction between corpus-based and corpus-driven approaches, discussed later in this chapter.

Stefanowitsch and Gries present the term *collostruction*, coined from *collocation* and *structure*, in questioning whether there are "significant associations between words and grammatical structure at all levels of abstractness," thus broadening the definition of collocation (2003:211). Bartsch defines collocations as "frequently recurrent, relatively fixed syntagmatic combinations of two or more words" (2004:11). Here the key word is *relatively*, as collocations are not as fixed as idioms but are not completely free word combinations since this co-occurrence of lexical items undergoes certain "combinational rules of syntax and semantics" (ibid. 11). Based on the *British National Corpus of contemporary British English*, Bartsch argues against the previous vague definitions of what is considered a collocation, suggesting her own, which comprises *quantitative, positional, structural* and *pragmatic* criteria, and proposes means of "modeling co-occurrence behavior, distribution and structural as well as functional properties of collocations" (ibid. 14). Drawing on frequency and statistical significance supported by mathematical algorithms and in compliances with the *operation working definition* developed by herself, Bartsch identifies collocations from authentic texts (BNC) and conducts both structural and functional property analysis of sample collocations. Observing the existing gap of systematic theory around collocations, Bartsch notices that, focusing on the widespread structures of recurrent word combinations, much of 20th century linguistics limited itself to the

"syntactic governing rules," while the study of collocations based on distributional data to uncover lexical co-selection constraints has largely been neglected (ibid. 15).
The present work does not have the objective to identify either existing or new structures of collocations; instead it does aim at singling out lexical co-selection constraints/semantic restriction based on distributional data. The hypothesis of the distributional approach, as Geeraerts states, is that "words that occur in the same context tend to have similar meanings" (2006:59).[40] The methodology of collocation analysis adopted in this work seeks among other things to provide a comprehensive description of lexical (word) fields for selected culinary verbs in Armenian, German, and English. In this framework, co-occurrences identified through statistical significance are then manually semantically annotated according to chosen parameters specific to culinary contexts viewed as specialized texts. This is the reason why the conventional definitions of collocations, e.g. Hausmann, discussed above, are not suitable for this work. Instead, the collocations of Armenian, English, and German culinary verbs in this study are derived from corpora using statistical significance,[41] followed by manual semantic annotation of the occurrences. Thus, a more up-to-date definition of what a collocation is seems more suitable for the identified collocations in this work, while not disregarding various definitions used for other purposes, such as investigating the structure of collocations, describing the properties of their constituents, etc.:

a) Collocations are, quite simply, co-occurrences of words in text. A statistically significant collocation is one that occurs more often than by chance. Such co-occurrences are measured by computational analysis of large corpora. (cf. Hanks 2013:1)

b) The name *collocation* implies, we're here dealing with a phenomenon that describes words which tend to occur in proximity (co + location) to one another because they have some kind of 'affinity' to, or 'affiliation' with, one another. (Weisser 2016:198)

Viewing collocations as the co-occurrences of two words based on frequency and statistical significance, comprising at least binary constituents and representing the *conventionalized patter of language use* is observed in Steyer (2008), who refers to collocations as *usuelle Wortverbindungen* (UWV) (also in Steyer 2000, 2004).

[40] Cf. Harris (1954), who viewed collocations in the sense of grammatical juxtaposition of words in a sentence, now referred to as *colligation*, and Apresjan (1966) who considered collocations for lexicographic purposes.
[41] Chapter 2 of this work presents different measurements of statistically significant collocations and singles out logDice (cf. Killgarriff et al. 2010) as the metric underlying the extraction of collocations in this work.

> Unter UWV verstehe ich konventionalisierte – zumindest binäre und minimal lexikalisch ausgefüllte – Muster des Sprachgebrauchts. Ihre Komponenten weisen eine auffällige statistische Affinität auf und sind in rekurrente, syntagmatische Strukturen eingebettet (usualisierte Syntagmen). (Steyer 2008:188)
>
> [By usual(conventional) word combinations, I understand conventionalized – at the very least binary and minimally lexically filled – patterns of language use. Their components exhibit a noticeable statistical affinity and are embedded in recurrent, syntagmatic structures (conventionalized syntagmata).]

However, UWV also include grammatical colligations of the type ‘auf Grund’, ‘Grund genug’, ‘in Grund und Boden’, etc. (ibid.), which shall not be considered as collocations within the framework of this work.

Thus, the definition of collocation is based here on frequency and recurrent word combinations relying on statistical significance measures, i.e. the *logDice* association score (cf. Evert 2005, Killgarriff et al. 2010), followed by manual semantic annotation of co-occurrences to ensure the interpretability of the word profiles as one of the outcomes of this work, as well as the comparison between the Armenian, German, and English culinary word fields. Different association scores – *T-score*, *MI-score*, *MI³-score*, *Minimum sensitivity*, *MI log Freq*, and *Dice* (cf. Evert 2005) – are not taken into consideration here, as they are elaborated upon in Chapter 2 of this work, as well as throughout the description of the German, Armenian, and English corpora where the *logDice* (Killgarriff et al. 2010) serves as the basis of measuring the strength of two words co-occurring in the same contextual environment.

1.5 Interpretable Semantic Word Profiles vs. Vector Space Meaning

We encounter various definitions of corpus from the early 90s until recent years, observing certain corrections in parallel with the development and integration of computational linguistic methods into corpus linguistics. The purpose the corpus is used for is one of the key aspects in defining what a corpus is, for instance, for linguistic analysis (a). The definition of a corpus by Atkins et al. (1992) (b) does not assume that it should necessarily be used for linguistic purposes, but still highlights the specificity of the purpose it has been compiled for, e.g. ‘electronic texts’. The idea of the compilation of a corpus, or more precisely what type of collected language material should or might be considered a ‘corpus’, is illustrated in definition c), where a corpus is viewed as a carefully collected 'textual material' compiled for a specific purpose, in order to be accessible for computers as opposed to a paper corpus. Moreover, this definition underlines that a corpus should be representative of a language as well as some text type, e.g. general vs. specialized corpora.[42] In later definitions, for instance (d), a corpus is described as a collection of carefully and *systematically* gathered texts used for linguistic analysis purposes. This definition excludes spoken corpora, i.e. audio corpora used for specific purposes, e.g. investigating dialects; however, if transcribed and compiled as a text collection, oral corpora then also fall within this definition of a corpus.

a) "[…] a corpus is a collection of naturally-occurring language texts, chosen to characterize a state or variety of a language." (Sinclair, 1991:171)
b) "[a corpus is] a subset of an ETL (Electronic Text Library) built according to explicit design criteria for a specific purpose." (Atkins, Clear and Ostler 1992:1)
c) "It should be added that computer corpora are rarely haphazard collections of textual material: They are generally assembled with particular purposes in mind, and are often assembled to be (informally speaking) representative of some language or text type." (Leech 1992:116)
d) "[…] any collection of texts that has been *systematically* assembled in order to investigate one or more linguistic phenomena." (Weisser 2016:13)

In the framework of this study, no particular definition of a *corpus* is given preference, but rather a generalized summary of all the definitions mentioned is adopted. On the one hand, much larger corpora such as German Web 2013 (*deTenTen13*) and English Web 2015 (*enTenTen15*), available in *Sketch Engine* (www.sketchengine.eu), are representative of the respective languages; on the other hand, the manually compiled English (*EN_REZ*), German (*REZ_DE*) and Armenian *(ARM)* culinary corpora represent a specialized text type, i.e. cookbooks, all *systematically collected* and used for *investigating linguistic phenomena*.[43]
Generally, corpus linguistic theory makes a binary distinction between corpus-driven and corpus-based approaches to linguistic analysis, first mentioned by Tognini and Bonelli (2001). In the framework of a corpus-based methodology, corpus data serves as evidence to support the pre-existing linguistic theories, bridging language and data.

[42] For more details on the "representativeness in corpus" (cf. Biber 1994).
[43] The German Web 2013 and English Web 2015 as well as the manually compiled corpora serving as the basis for the linguistic analysis within the scope of this work are described later in this as well as in the chapters following.

> […] the term corpus-based is used to refer to a methodology that avails itself of the corpus mainly to expound, test or exemplify theories and descriptions that were formulated before large corpora became available to inform language study. (Tognini-Bonelli 2001:65)

In traditional linguistic paradigms, researchers are interested in corpus linguistics as a method of restating their theories through corpus-based evidence, e.g. by extracting examples (Atkins, Osler and Clear 1992).

The corpus-driven approach gives the corpus an even higher value than a mere restatement or adjustment of existing linguistic descriptions by viewing the corpus itself as the carrier of the theory of language. The theories and descriptions of linguistic phenomena *are driven* from corpus.

> In a corpus-driven approach the commitment of the linguist is to the integrity of the data as a whole, and descriptions aim to be comprehensive with respect to corpus evidence. The corpus, therefore, is seen as more than a repository of examples to back pre-existing theories or a probabilistic extension to an already well defined system. The theoretical statements are fully consistent with, and reflect directly, the evidence provided by the corpus. (Tognini-Bonelli 2001:84)

However, the latest works of corpus linguistic do not make a clear-cut distinction between the *corpus-driven* and *corpus-based* approaches. In introducing the neo-Firthian school of corpus linguists, McEnery and Hardie reject the binary dichotomy of *corpus-based* and *corpus-driven* linguistic approaches by stating that "*all* corpus linguistics can justly be described as *corpus-based*" (McEnery and Hardie 2012:5).

In the framework of the present research, the non-binary character of corpus linguistics was adopted, with the linguistic analysis being both the outcome of the theories embodied in the corpus itself and a means of providing examples to back up previous knowledge-based descriptions. In other words, previous descriptions of culinary verbs are tested for validation, additions or amendments using the available corpus data, whereas new theoretical models driven from the corpus itself are also suggested. This work is therefore a mixture of corpus-driven and corpus-based approaches.

Cross-linguistic lexical studies for contrastive linguisitcs and translation purposes have attracted more attention with the rise of corpus-based lexical analysis. Differentiations have been formulated not only between corpus-driven and corpus-based methodologies but also between comparable and parallel types of multilingual corpora (cf. Altenberg and Granger 2002, and also Baker 1993).[44] In distinguishing between translation (parallel) and comparable corpora, Alentenber and Granger define the two as follows:

> Comparable corpora consist of original texts in each language, matched as far as possible in terms of text type, subject matter and communicative function… Translation corpora consist of original texts in one language and their translations into one or several other languages. (Altenberg and Granger 2002:4)[45]

Having the objective to conduct a corpus-based contrastive study of culinary vocabularies, e.g. culinary verbs in Armenian, German, and English, as well as addressing translation issues in finding possible equivalents for such specialized texts, comparable corpora in the aforementioned languages have served as the basis for reaching the goals of this study by facilitating intralinguistic analysis and the identification of possible translation equivalents based on interlinguistic comparison. As Babych et al. (2007) argue, using distributional similarity to identify possible translation equivalents as a means of harnessing equivalents "outperforms" those established for parallel corpora. The corpora used in this study are described and elaborated upon in the upcoming chapters for each language (see Chapter 2-4); however, it is worth mentioning here briefly that the following study was made possible by both manually compiled culinary corpora in Armenian, German, and English, i.e. cookbooks both in paper and digital format, as well as German Web 2013 (*deTenTen13*) and English Web 2015 (*enTenTen15*) available in *Sketch Engine* (www.sketchengine.eu). Unfortunately, there is no comparable Armenian corpus with regard to size and NLP requirements, i.e. tokenization, lemmatization, and POS-tagging. Therefore, only the manually compiled culinary corpus was used for the Armenian culinary verb analysis in this study. Nevertheless, the aforementioned manually compiled corpora, as well as the corpora available in *Sketch Engine*, all meet the requirements of *comparable corpora*, "matching as far as possible in terms of text type, subject matter and communicative function" as defined above (Altenberg and Granger 2002:4).

[44] Distinctions are made between parallel and comparable parallel corpora, where the two corpora "are matched as much as possible in terms of sampling frame", possessing not only the properties of comparable corpora but also being aligned in "time-frame and the language variety sampled" (Hareide 2019:21).
[45] Altenberg and Granger also differentiate between *unidirectional* and *bidirectional* translation corpora. While the former presupposes translation from only A language to B language, the latter also envisages the opposite direction as well, i.e. from B language to A (Altenberg and Granger 2002:4)

In the early 1990s, attempts were made to draw correlations based on the contextual similarity of words.

> The contextual representation of a word is knowledge of how that word is used…A word's contextual representation is assumed to include syntactic, semantic, pragmatic, and stylistic conditions governing the use of that word." (Miller and Charles 1991:4)

Following the conventions of the "distributional hypothesis" (Harris 1954, Firth 1957), which asserts that "words occurring in the same context have similar meanings," Rubenstein and Goodenough's study, focusing on nouns, was based on *co-occurrence* and *substitutability*, the former singling out the "overlaps" of the investigated nouns with other words in two different sentence contexts, the latter showing how substitutable a certain pair of nouns was in the sentence contexts. Miller and Charles (1991) repeated the experiment on the same 65 pairs of nouns and identified 30 pairs of different degrees of similarity based on the Rubenstein and Goodenough's findings, while focusing solely on *substitutability* and on "the relation of semantic similarity to contextual similarity at low levels of contextual similarity," where the co-occurrence "had been least satisfactory" (Miller and Charles 1991:12). In this study, native English-speaking students grouped nouns based on how they were used in sentences. Their groupings were a little different but mostly matched the results from Rubenstein and Goodenough. In the end, they found that "words from the same syntactic and semantic categories" tend to have similar meanings, especially when they are often used in the same contexts (ibid. 15-17).

The aforementioned experiment, especially the revisited one by Miller and Charles (1991), was more semantically oriented, though still within the scope of the distributional approach. Two decades later, with the rise of machine-supported translation, there was still a growing interest in extracting information from large corpora – now not merely comprising millions of tokens, but billions. The vector space model (VSM) comes into the limelight with the objective to calculate and visualize the similarity of words in their contexts based on a mathematically defined distance between words, i.e. the *Skip-gram* model (Mikolov, Sutskever, et al. 2013), which we will discuss later in this chapter.

Without aiming to delve into the specifics of the generation of VSMs, especially their underlying mathematical details, a simplified vector space of only seven words across three contexts is presented below for an intuitive understanding of how it works, as introduced by Desagulier in his article *Word embeddings: the (very) basics* (2018) (Fig. 1).

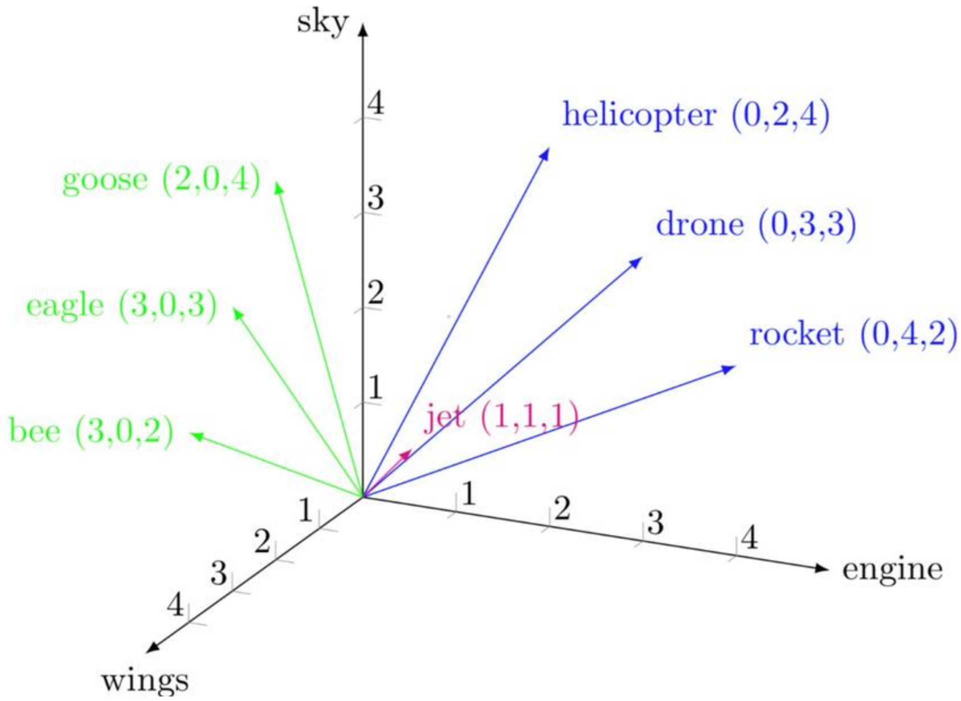


Fig. 1. A vector space of seven words in three contexts (Desagulier 2018, https://corpling.hypotheses.org/495)

Viewing "word embedding as computational implementation of distributional hypothesis," Desagulier uses a very small corpus of seven words (Fig.1) across three contexts, i.e. 'engine', 'wings', and 'sky'. Each word is assigned coordinates in a three-dimensional space, i.e. the contexts. According to how many times each of the seven words occurs in the three contexts, each of them is placed on a specific spot in the vector space. The word vector is then defined as "the arrow from the point where all three axes intersect to the end point defined by the coordinates" (Desagulier 2018). Drawing on cosine similarity, i.e. "the cosine of the angle between two word vectors," the similarity of two words can be measured. However, in the linguistic analysis of real data (unlike in the example above in Fig. 1), such co-occurrence matrix would include vectors for thousands of words with thousands of dimensions, depending on the size of the lexicon. For each word, the dimensions would represent co-occurrence counts with each other word in the corpus. To process these vectors efficiently, the number of dimensions is reduced applying mathematical techniques, such as the *Principal Component Analysis* (PCA) (Fig. 2). For instance, if the lexicon derived from a large corpus consists of ca.

500.000 words – resulting in the same number of vectors and dimensions, PCA can reduce the number of dimensions to about 400 dimensions, usually predetermined by the user based on trial-and-error method. However, such reduced dimensions would no longer be suitable for linguistically meaningful interpretation, since we are merely dealing with numbers used for positioning words in an abstract 400-dimensional space, and these positions capture their semantic properties only to some limit.[46] Although this abstract representation is useful for Machine Translation (MT) or natural language generation, since synonyms and paraphrases are located together, it does not improve our linguistic understanding about the structure of the semantic fields: new dimensions do not directly generalize any linguistic features; instead, they are used for calculating similarity between different words (cf. Fig 1 and Fig. 2). However, such generalizations are needed for Frame Semantics and lexicography (e.g., for finding definitions for words and differences with related terms).

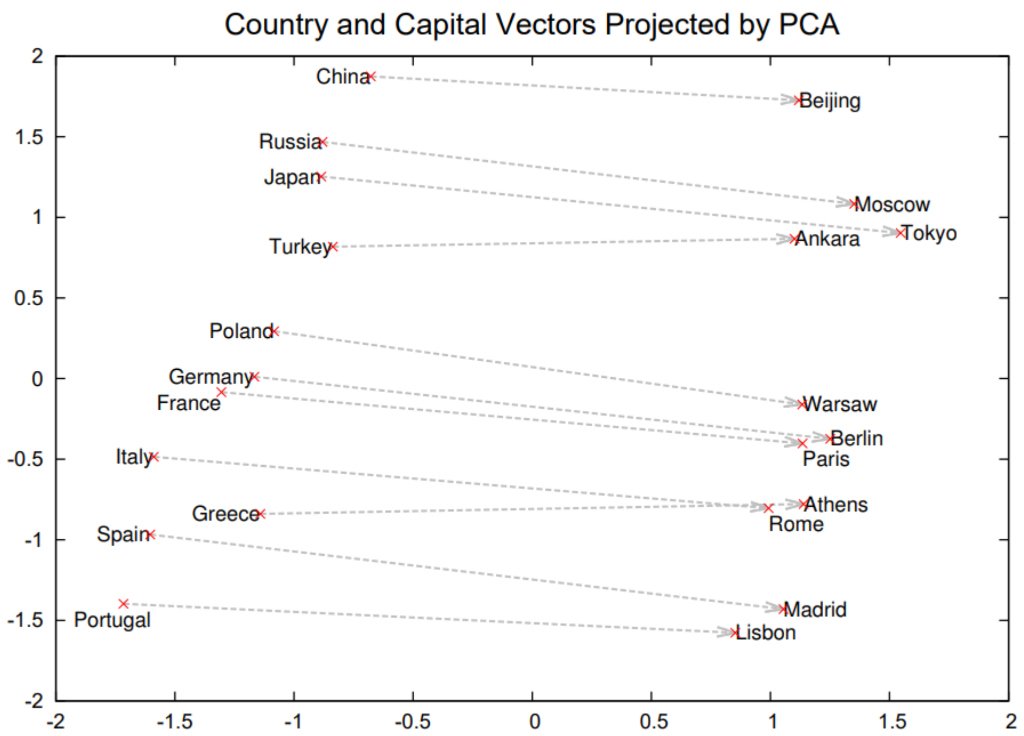


Fig. 2. “The 2-dimensional PCA projection of the 1000-dimensional Skip-gram vectors of countries and their capital cities”

[46] The skip-gram model implemented in *Word2Vec* learning algorithm generates the required number of non-interpretable dimensions (usually between 100 and 400) even without explicitly producing the co-occurrence matrix.

Even though the capitals are *correctly* correlated with the respective countries, the figure just shows the success of the Skip-gram model and PCA, which reduces its dimensions to two. As the authors state, "during the training [they] did not provide any supervised information about what a capital city means," emphasizing that the figure merely "illustrates the ability of the model to automatically organize concepts and learn implicitly the relationships between them" (Mikolov, Sutskever et al. 2013:4).[47]

The *Skip-gram* model and *PCA*, as well as word embedding implementations like *word2vec* in general, have been successfully used in projects interested in identifying correlations without providing a linguistic description of the semantic relations between them.

Such models, therefore, are not directly interpretable and less suitable for linguistic analysis. For instance, Gennaro, Buonanno, and Palmieri (2021) highlight the limitations of word embeddings, among other things also the incapability of generating word vectors based on syntactic relations. Wadud, Mridha, and Rahman (2022) discuss the shortcomings of a pre-trained word embedding (fastText) vector, by comparing it with their own locally trained word embedding model for the Bengali language (developed in Python), with maximum accuracy reaching 87.84%, as opposed to 86.75%. Mikolov, Chen, et al. (2013) reiterate the importance of generating word vectors at a lower computational cost. Nonetheless, word vectors, when combined with models such as the *Conditional Inference Tree* (CIT), may contribute to targeting the question of word sense disambiguation, thus enhancing machine translation. One of such corpus-based projects (Sun et al. 2021) generates word vectors with *word2vec* (available in python) for detecting ambiguous words and appropriate contexts to then identify "the high mean similarity" as the correct meaning based on the calculation of "the cross similarity of the word vector."

Our approach bridges distributional and conventional logic-based lexical semantics by extracting collocations from corpora based on statistical significance. The manual semantic annotation of the occurrences in this bottom-up study preserves the interpretability of contextual features. This approach can be used for creating more accurate lexicographic resources (cf. e.g. Apresjan 1966, 1974, 1992, 2000, 2002), as well as frame descriptions for semantic fields (e.g. FrameNet, WordNet). The adopted methodology underlying this work contributes to filling the

[47] By adding the feature *young* to the pre-trained model of vector representation of four animals, *duck, dog, cow,* and *cat*, carried out by Mikolov, Grave, Bojanowski, Puhrsch, & Joulin (2017) Sassenhagen and Fiebach demonstrate the resulting positions of the baby animals within the vector space. However, the word *calf* is notably absent from a position near *cow* (2019:5–6).

existing gap in the field of combining distributional and traditional semantics by taking the interpretability of contextual features as a starting point; it uses semantic classes rather than individual words as features; moreover, it allows not only to calculate similarity between words, or to list examples of words belonging to a certain class, but also to exactly show the nature of semantic distinctions between them. This can be used to generate clearer and better definitions or frame semantic representations for these groups of words as well as identify possible translation equivalents based on the contrastive analysis of original, authentic corpora in the languages being compared. The outcome, in terms of visualizations of the analyzed data, may resemble (in particular in the case of ordination techniques like *CA* and *MCA*) the vector space-based descriptions of words; however, this similarity is only superficial, as the visualizations are supported by manual semantic annotation, retaining their interpretability. The methodology has been successfully applied to address different linguistic objectives, thus proving its effectiveness (cf. Atayan et al. 2019; Gast and Atayan 2019; Atayan and Gast 2022, Straube 2023).

Chapter 2. Description of the German Web 2013 (deTenTen13) and REZ_DE Corpora

Two types of German culinary corpora were used to extract examples for later annotation. On one hand, a web corpus, viz. the *deTenTen13* with 19,808,173,163 tokens available in *Sketch Engine* corpus manager and text analyzer developed by Lexical Computing Ltd and accessed through the university license, and on the other hand, a much smaller, self-compiled corpus of cookbooks in German, viz. *REZ_DE* available in hard copy or digital format. Initially the choice of the two types of corpora was aimed at finding out whether they differ, and if so, to what extent. The underlying assumption was that the published cookbooks, which are mostly written by cooks and/or food critics or, in other words, food experts, would differ in their complexity and precision, both in text structure and lexical choices. The German web corpus *(deTenTen13)*, on the other hand, was not filtered and included a wide range of text types within culinary contexts, for instance cooking forums, blogs, internet recipes written by non-experts, menus, travel guides, etc.

Both corpora were analyzed with the *Sketch Engine* corpus manager and text analyzer. Among other functions, such as *Keyword* extraction, *N-gramm*, *thesaurus*, the *Sketch Engine* corpus query system also offers *Word sketch*[47] to represent the grammatical and the collocational behavior of a given word (Killgariff et al. 2004). *Word sketch* is based on the *Mutual Information* measurement of the salient collocates calculation in a given corpus, first introduced by Church and Hanks (1990). This method of processing the association between two words relies on the salience calculation of a word with its neighboring words, "collocates" with a "span" of four words (from left and right) in relation to the "node word" (Stubbs 1996:172 ff.). *Word sketch*, however, in addition to presenting the salient collocates of a word, took into account the shortcomings of the Church and Hanks method, e.g. the arbitrariness of the word span around the headword, and organized the words in grammatical relations, viz. the "grammatical pattern the word participates in" (Kilgarriff et al. 2004:5). Since we were interested in the actual language usage and had adopted collocation analysis as our methodology, *Word sketch* was used as the main tool for extracting collocations.[48] Several scores were previously developed to determine the association between two words seen as

[47] Based on the function *Word sketch*, *Sketch Engine* language analysis tool also offers *Word sketch difference* which compares the collocations of two separate word sketches. For instance, the collocations of the word *'steam'* with that of the word *'cook'*.
[48] On collocations, see Giacomini 2011.

collocations, among them *T-score*, *MI-score*, *MI³-score*, *Minimum sensitivity*, *MI logFreq*, and *Dice* (cf. Evert 2005). The latter yielded rather successful results, but "due to the very small numbers of *dice* values, a new association score" – *logDice* – was developed (Rychlý 2008:9).

$$logDice = 14 + \log_2 D = 14 + \log_2 \frac{2f_{xy}}{f_x + f_y}$$

In explaining the values of the *logDice* features, Rychlý states that a score of 14 is theoretically the highest, showing that all occurrences of *X* co-occur with *Y* and vice versa. However, in reality the highest score is typically around 10. *X* and *Y* here are two potentially collocated words. *F* is the relative frequency of the word in the corpus; therefore the size of the corpus does not determine the score (ibid. 9).

With the help of a filter option in the *Word sketch* function, one can view collocates with the highest association score. Besides the *logDice* score, *Word sketch* also reveals the number of occurrences (frequency) of a specific entry, viz. the number of concordances. For instance, the *logDice* score for the word *Süppchen* as one of the collocations of the verb *kochen* as its direct object in accusative case is rather high (10.4) with 6762 occurrences in *deTenTen13* corpus (see Fig.1). The tool also allows to see the concordances of the corpus with all examples of the collocations in question. The *Word sketch* in the *Sketch Engine Corpus-Query-System* (CQS) then presents lists of collocates of a word, categorized by the different grammatical relations in which the word takes part. For instance, for a verb, separate lists of collocates are displayed for "the subject, the objects, the conjoined verbs, modifying adverbs, prepositions and prepositional objects," corresponding to the grammatical relations the verb might have (Killgariff 2004:5). A (partial) *Word sketch* of the German culinary verb *kochen* is illustrated below (Fig. 1), with the nominal phrases in the dative and accusative cases being singled out.

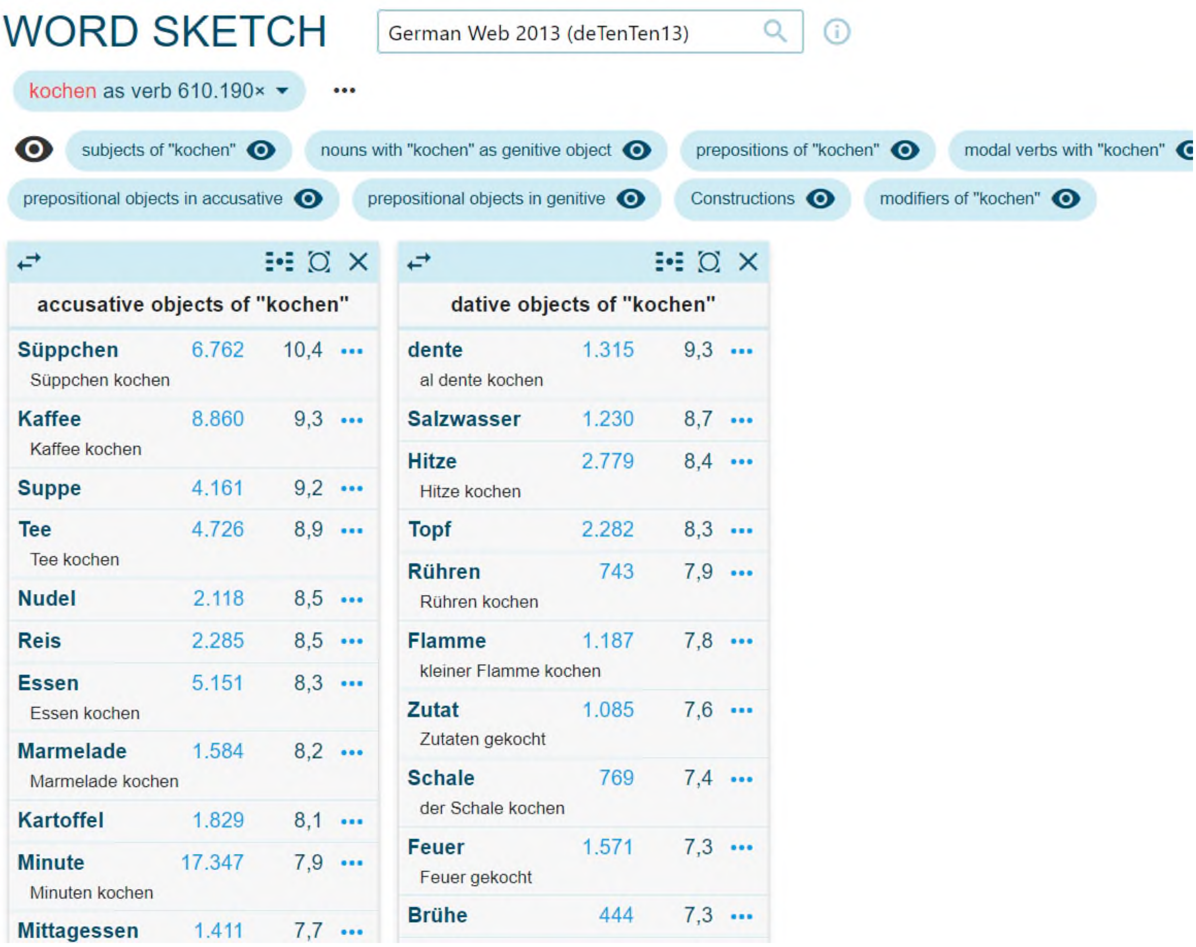


Fig. 1. Partial *Word sketch* of the German culinary verb *kochen* in the *deTenTen13* corpus

In this study, the column of collocates in the accusative case was into consideration since the nature of the culinary texts assumes that usually something is prepared, cooked by somebody, e.g. *Essen kochen*, *Zwiebeln dünsten*, even though verb modifiers and sometimes indirect objects in the dative case were also taken into account as a result of the annotation parameters, which is elaborated in detail below.

Since it would not be practically possible to include all collocations of eight German culinary verbs, namely *backen*, *braten*, *dünsten*, *grillen*, *kochen*, *rösten*, *schmoren*, and *toasten*, it was decided to include all collocations above 5.0 *logDice*, as they portray the strongest ones. The lower the *logDice* score, the weaker the collocation. Several tests were carried out in the framework of this research to reveal that none of the collocations with a *logDice* score below 5.0 had a relevant effect on the results of the analysis, as they were mostly either repetitions or hyponyms of previous collocations with much higher scores, i.e. above 5.0. For instance, *Espresso kochen* had a *logDice* score of 4.7 and was therefore excluded in the analysis of this work; however, its hypernym *Kaffee kochen*, with a *logDice* score of 9.3, had already been included in the collocation extraction.

Even with collocations scoring above *logDice* 5.0, it was still a challenge to determine the appropriate distribution of the extracted occurrences in relation to the identified collocation candidates in corpus for further manual semantic annotation. After identifying the desired number of occurrences to be annotated, a model was developed in which the number of examples selected for analysis with a given collocate was weighted proportionally to its score (see below).[49] Consequently, 200 occurrences have been annotated for hypernym and 100 for hyponym culinary verbs.[50] In the case of German culinary verbs, *kochen*, *braten*, and *backen* are hypernyms while *dünsten*, *grillen*, *rösten*, *schmoren*, and *toasten* are hyponyms.[51] However, if the number of entries in the given corpus was smaller than the necessary quantity of occurrences available in corpus, two strategies were used to close the gap. First, a more general food item came to substitute the more specific one (hyponymy relation); for instance, *Kaffeebohnen* instead of *Espressobohnen*. Second, if the substitution through hyponymy was not possible, only as many occurrences as the corpus evidence allowed were annotated. This explains why some of the verbs did not have exactly 200 or 100 annotated occurrences.

Nonetheless, it was problematic to determine how many examples with each collocate and the given node (the culinary verb) should be annotated to provide a representative image of the given corpus. In other words, how many examples of the overall 738 occurrences of e.g. *Spiegelei braten* from the *deTenTen13* sub-corpus (Fig. 2) should be annotated to represent the distribution of the collocate in relation to the context of the verb *braten*. Since our objective is to describe our corpora not based on the general frequency of occurrences but rather on the most typical (the strongest) collocations, collocates with the highest association score should have a higher representation in our sub-corpus (in this case for the verb *braten*) than those with considerably lower scores. For this purpose, an annotation quantity determination model was developed based on a mathematical representation of the typical collocates' distribution (Fig.2).

[49] Since our manually compiled *REZ_DE* sub-corpus comprised only specialized culinary texts, viz. cookbooks, and is much smaller than *deTenTen13,* all extracted examples were annotated irrespective of their number and the general or the specific character of the verb. For instance, all 48 occurrences of the verb *backen* in the *REZ_DE* sub-corpus were annotated. No occurrences of the verb *toasten* could be observed in the same sub-corpus; therefore, no example was annotated here.

[50] The same criteria have also been implemented for the manual semantic annotation of the occurrences of the English and Armenian culinary verbs and their collocations.

[51] The division of the given culinary verbs into hypernyms and hyponyms in this study is based on the classification of the German culinary verbs provided by Lehrer (1972).

1	Word	Freq.	logDice*1000	Number	Dice	Model	Designation in 200	Rounded	Zero→1
2	Spiegelei	738	9.630	1	0,0483614	0,0584	30,512672	31	31
3	Steak	630	8.830	2	0,0277763	0,0339867	17,757254	18	18
4	Zwiebel	647	8.300	3	0,0192366	0,0247617	12,937446	13	13
5	Schnitzel	387	8.240	4	0,018453	0,019779	10,33407	10	10
6	Speck	394	8.120	5	0,0169802	0,0166156	8,681298	9	9
7	Würstchen	264	7.730	6	0,0129581	0,0144104	7,5291182	8	8
8	Bratwurst	249	7.700	7	0,0126914	0,0127759	6,6751138	7	7
9	Wurst	425	7.640	8	0,0121744	0,0115106	6,014049	6	6
10	Fleisch	1468	7.530	9	0,0112807	0,010499	5,4855078	5	5
11	Filet	173	7.410	10	0,0103804	0,0096697	5,0521966	5	5
12	Hackfleisch	150	7.250	11	0,0092907	0,008976	4,6897809	5	5
13	Frikadelle	127	7.090	12	0,0083154	0,0083863	4,3816702	4	4
14	Gans	150	6.850	13	0,007041	0,0078781	4,1161433	4	4
15	Knoblauch	168	6.830	14	0,0069441	0,0074351	3,884671	4	4
16	Tofu	102	6.680	15	0,0062584	0,0070451	3,6808911	4	4
17	Kartoffel	278	6.660	16	0,0061722	0,0066988	3,4999556	3	3
18	Hähnchen	98	6.560	17	0,0057589	0,006389	3,3381022	3	3
19	Fisch	701	6.410	18	0,0051902	0,0061101	3,1923641	3	3
20	Kotelett	74	6.380	19	0,0050834	0,0058574	3,0603682	3	3
21	Pfannkuchen	90	6.350	20	0,0049788	0,0056274	2,9401928	3	3
22	Olivenöl	140	6.320	21	0,0048763	0,005417	2,8302642	3	3
23	Fischfilet	68	6.250	22	0,0046453	0,0052237	2,7292802	3	3
24	Lachs	94	6.220	23	0,0045497	0,0050455	2,6361541	3	3
25	Garnele	78	6.150	24	0,0043343	0,0048805	2,5499711	3	3
26	Champignon	68	6.120	25	0,0042451	0,0047274	2,4699553	2	2
27	Spargel	106	6.110	26	0,0042157	0,0045848	2,3954442	2	2
28	Ei	488	6.100	27	0,0041866	0,0044516	2,3258684	2	2
29	Schalotte	64	6.070	28	0,0041005	0,004327	2,2607358	2	2
30	Bratkartoffel	61	6.040	29	0,0040161	0,00421	2,1996187	2	2
31	Rührei	61	6.030	30	0,0039883	0,0041	2,1421434	2	2
32	Bacon	58	6.020	31	0,0039608	0,0039963	2,087982	2	2
33	Gemüse	313	5.920	32	0,0036955	0,0038984	2,0368456	2	2
34	Ochse	60	5.920	33	0,0036955	0,0038059	1,9884783	2	2
35	Omelett	51	5.830	34	0,003472	0,0037182	1,9426529	2	2
36	Forelle	63	5.760	35	0,0033076	0,0036349	1,8991667	2	2
37	Speckwürfel	45	5.720	36	0,0032172	0,0035558	1,8578386	2	2
38	Pilz	136	5.710	37	0,0031949	0,0034805	1,8185058	2	2
39	Burger	57	5.710	38	0,0031949	0,0034088	1,7810219	2	2
40	Butter	159	5.690	39	0,0031509	0,0033403	1,7452546	2	2
41	Ente	85	5.660	40	0,0030861	0,0032749	1,7110842	2	2
42	Aubergine	47	5.660	41	0,0030861	0,0032124	1,6784022	2	2
43	Spieß	86	5.650	42	0,0030648	0,0031525	1,6471098	2	2
44	Roulade	46	5.630	43	0,0030226	0,0030951	1,6171168	2	2
45	Zucchini	50	5.620	44	0,0030017	0,00304	1,5883408	2	2
46	Fett	180	5.610	45	0,002981	0,0029871	1,5607066	2	2
47	Schinken	62	5.600	46	0,0029604	0,0029363	1,5341449	2	2
48	Fleischstücke	42	5.580	47	0,0029196	0,0028874	1,508592	2	2
49	Fischstäbchen	42	5.580	48	0,0029196	0,0028403	1,4839895	1	1
50	Stockbrot	44	5.570	49	0,0028995	0,0027949	1,4602832	1	1
51	Zwiebelring	41	5.570	50	0,0028995	0,0027512	1,4374233	1	1
52	Zwiebelwürfel	41	5.550	51	0,0028595	0,0027089	1,4153633	1	1
53	Rösti	40	5.520	52	0,0028007	0,0026682	1,3940605	1	1
54	Medaillon	40	5.420	53	0,0026131	0,0026288	1,3734751	1	1
55	Bulette	37	5.400	54	0,0025772	0,0025907	1,35357	1	1
56	Spanferkel	37	5.370	55	0,0025241	0,0025538	1,3343107	1	1
57	Braten	50	5.310	56	0,0024213	0,0025181	1,3156652	1	1
58	Bällchen	37	5.290	57	0,002388	0,0024836	1,2976034	1	1
59	Huhn	75	5.250	58	0,0023227	0,0024501	1,2800973	1	1
60	Keule	41	5.230	59	0,0022907	0,0024176	1,2631206	1	1
61	Kartoffelscheib	32	5.200	60	0,0022436	0,002386	1,2466488	1	1
62	Hühnchen	36	5.190	61	0,0022281	0,0023554	1,2306587	1	1
63	Hamburger	53	5.180	62	0,0022127	0,0023257	1,2151288	1	1
64	Truthahn	32	5.150	63	0,0021671	0,0022968	1,2000388	1	1
65	Paprika	40	5.100	64	0,0020933	0,0022687	1,1853693	1	1
66	Brotwürfel	29	5.090	65	0,0020788	0,0022414	1,1711025	1	1
67	Puffer	39	5.060	66	0,0020361	0,0022149	1,1572213	1	1
68	Leber	75	5.040	67	0,002008	0,002189	1,1437097	1	1
69	Storch	38	5.010	68	0,0019667	0,0021638	1,1305526	1	1
70	Kartoffelpuffer	28	5.010	69	0,0019667	0,0021393	1,1177356	1	1
71	Total								228

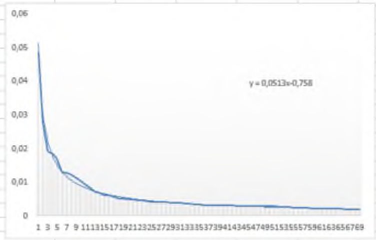


Fig.2. Annotation quantity determination model for the verb *braten* in the *deTenTen13* corpus

First of all, the association score *logDice*, with which the collocates of a certain node (verb) had previously been listed, were converted into *Dice*[52] to see the real differences between the collocates of the same node, followed by a 2D-line diagram to visually represent the distribution of the collocates in relation to the given verb context across the whole *deTenTen13* corpus. Next, a trend line (e.g. exponential, potency, logarithmic) was generated to best match the 2D-line. Based on the formula of the chosen trend line, a model was developed to determine the distribution of the collocates in the given sub-corpus. In the case of the collocates related to the context of the verb *braten*, the model made it possible to annotate 31 occurrences of the collocate *Spiegelei* in the overall 200 annotated examples (see Fig.2), to proportionally

[52] Dice coefficient "like MI (*Mutual Information*), is a maximum-likelihood estimate of *association strength*" measurement used in collocation extractions (Evert 2005:200, cf. also Rychlý 2008).

represent the 738 occurrences in the *deTenTen13* sub-corpus. According to the same model, only one occurrence with the node *braten* and the collocate *Kartoffelpuffer* was annotated proportionally to represent 28 examples of the same *deTenTen13* corpus. It is easy to observe that *Spiegelei braten* is at the top of the trend line, while *Kartoffelpuffer braten* is at the furthest bottom. In the next step, the number of occurrences to be annotated were rounded, and, if necessary, all collocates with approximately zero values were automatically calculated as one. This is the reason why, for some verbs, the exact number of 200 or 100 examples were exceeded, as all zeros have been converted into one. As illustrated on the annotation quantity determination model (Fig.2), 228 occurrences instead of 200 were annotated due to rounding of all decimal numbers.

Word	Freq.	logDice*1000	Number	Dice	Model	Designation in 100	Rounded	Zero→1
Toastbrot	37	8.920	1	0,0295643	-0,0036	13,5668883	14	14
Toastscheibe	13	8.820	2	0,02758447	-0,00368585	13,8904165	14	14
Brotscheibe	26	8.630	3	0,0241807	-0,00373701	14,0832332	14	14
Weißbrotscheibe	6	8.220	4	0,01819896	-0,00377374	14,2216599	14	14
Toast	80	8.060	5	0,01628853	-0,00380248	14,3299682	14	14
Graubrot	5	7.700	6	0,01269144	-0,00382613	14,4190747	14	14
Weißbrot	22	7.680	7	0,01251672	-0,00384624	14,4948452	14	14
Toaster	5	5.420	8	0,00261314	-0,00386374	14,5608024	15	15
	194							**113**

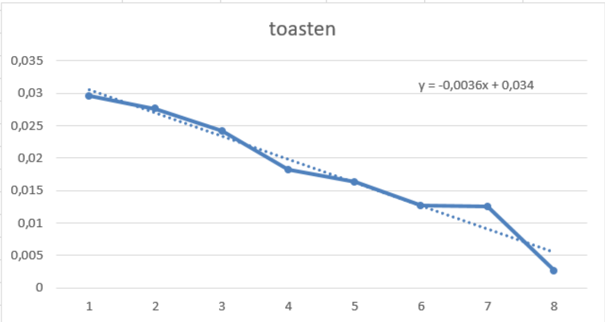


Fig. 3. Annotation quantity determination model for the verb *toasten* in the *deTenTen13* corpus

Fig. 3. illustrates another example of quantity determination model where the actual frequency of certain collocates is lower than what the model determines. *Graubrot toasten*, for instance, occurs only 5 times in the *deTenTen13* sub-corpus; however, according to the model, 14 occurrences should have been annotated. In such cases, as mentioned above, examples of a more general category were annotated instead, viz. *Brot toasten*. This had not affected the annotation and the final analysis in any way, since by the end of all annotations, the values of all parameters were neutralized (generalized) to allow the visualization tools to build up correlations while retaining as much information from the annotation as possible.

2.1 Description of the Manual Semantic Annotation Parameters

Before introducing the category parameters that served as the basis for the annotation of the exported occurrences from both the *Web 2013 (deTenTen13)* and *REZ_DE* sub corpora, it is worth noting that even within the predetermined model for the exact quantity of occurrences, the examples were not randomly chosen. On the contrary, due to the density of German recipe texts (examples 1-4), where even an entire recipe procedure might be condensed in a sentence, another strategy of annotation was also integrated.

1. Den Grünkohl zugedeckt etwa 1 Std. im Ofen bei 200 °C *schmoren (deTenTen13)*
   [Braise* the kale covered for about 1 hour in the oven at 200 °C].

2. Das Öl im Topf erhitzen, die Zwiebelringe mit dem Knoblauch kurz glasig *dünsten (deTenTen13).*
   [Heat the oil in a pot, shortly sauté* the onion rings with the garlic until translucent].

3. Mandelblättchen in eine Pfanne ohne Fett *rösten*. *(REZ_DE)*
   [Roast* the almond flakes in a pan without any fat].

4. 2-3 EL Öl in einer großen beschichteten Pfanne erhitze und den Fisch darin pro Seite ca. 3 Minuten goldbraun *braten*. *(REZ_DE)*
   [Heat 2-3 tbs. of oil in a large coated pan and fry* the fish for about 3 min. on each side until golden brown].

Such dense contexts gave rise to the parameters underlying the manual semantic annotation of this research. They revealed that certain parameters actually determine the choice of the culinary verb in the production of the original recipe on the one hand, and provide a possible basis for language comparison on the other. This led to the idea that if we could identify such dense culinary contexts in other languages and annotate them with the same parameters, it might be possible to develop equivalent models based on the analysis of comparable – rather than parallel – corpora. Thus, all occurrences from both the *deTenTen13* and *REZ_DE* sub-corpora annotated with the developed parameters were carefully selected dense contexts such as those mentioned above (examples 1-4).

German culinary contexts have initially resulted in eight parameters of annotation in addition to the verb (the node), namely *Ingredient*, *Substance*, *Dish*, *Utensil*, *Manner*, *Heat intensity*, *Resultative (adjectives)*, and *Heat source*.[53] The annotated table also included the actual occurrences exported from the respective corpus and the name of the corpus (Tab. 1).

Sentences	Verb	Ingredient	Substance	Dish	Utensil	Manner	Heat Intensity	Resultative	Heat source	Corpus
1. Butter im Topf erhitzen und die feingeschnittene Zwiebel darin glasig dünsten.	dünsten	feingeschnittene Zwieb	Butter	N_A	Topf	N_A	N_A	glasig	N_A	REZ_DE
2. In einem Topf Kokosöl erhitzen, die Zwiebeln glasig dünsten .	dünsten	Zwiebeln	Kokosöl	N_A	Topf	N_A	N_A	glasig	N_A	REZ_DE
2. Nach ca. 2 Minuten Knoblauchscheiben hinzufügen. Anschließend Zitronensaft und Butter dazugeben und alles weitere 2 Minuten dünsten.	dünsten	Knoblauchscheiben	Butter	N_A	N_A	N_A	N_A	N_A	N_A	REZ_DE
3. Nach ca. 2 Minuten Knoblauchscheiben hinzufügen. Anschließend Zitronensaft und Butter dazugeben und alles weitere 2 Minuten dünsten.	dünsten	Knoblauchscheiben	Zitronensaft	N_A	N_A	N_A	N_A	N_A	N_A	REZ_DE
7. Gemüse in kochendem Salzwasser ca. 10 Minuten dünsten.	dünsten	Gemüse	Salzwasser	N_A	N_A	N_A	N_A	N_A	N_A	REZ_DE
10. Zugedeckt bei schwacher Hitze ca. 10 Minuten dünsten.	dünsten	N_A	N_A	N_A	N_A	zugedeckt	bei schwacher Hitze	N_A	N_A	REZ_DE
12. Zwiebeln fein würfeln und in einem El Butter goldgelb dünsten .	dünsten	Zwiebeln	Butter	N_A	N_A	N_A	N_A	goldgelb	N_A	REZ_DE
13. Danach die Zwiebeln in etwas Öl langsam weich dünsten ohne das sie Farbe annehmen.	dünsten	Zwiebeln	Öl	N_A	N_A	N_A	N_A	N_A	N_A	REZ_DE
75. Zwiebel im Bratfett glasig dünsten.	dünsten	Zwiebel	Bratfett	N_A	N_A	N_A	N_A	glasig	N_A	REZ_DE
77. Öl in einem großen Topf erhitzen. Ingwer, Zwiebel und Knoblauch darin glasig dünsten.	dünsten	Ingwer	Öl	N_A	Topf	N_A	N_A	glasig	N_A	REZ_DE
89. 2 EL Butter, 2-3 EL Zucker und Zitronensaft im Topf erhitzen. Die Apfelstücke darin zugedeckt ca. 10 Minuten dünsten (sie sollen weich werden, aber nicht zerfallen), dabei zwischendurch öfter umrühren.	dünsten	Apfelstücke	Zitronensaft	N_A	Topf	zugedeckt	N_A	N_A	N_A	REZ_DE
90. Bohnen putzen, waschen. In wenig kochendem Salzwasser zugedeckt 8-10 Minuten dünsten.	dünsten	Bohnen	Salzwasser	N_A	N_A	zugedeckt	N_A	N_A	N_A	REZ_DE
91. Nach ca. 2 Minuten Knoblauchscheiben hinzufügen. Anschließend Zitronensaft und Butter dazugeben und alles weitere 2 Minuten dünsten.	dünsten	Knoblauchscheiben	Zitronensaft	N_A	N_A	N_A	N_A	N_A	N_A	REZ_DE
95. Gefrorenen Spinat und 4-5 EL Wasser zufügen. Zugedeckt ca. 15 Minuten dünsten.	dünsten	Spinat	Wasser	N_A	N_A	zugedeckt	N_A	N_A	N_A	REZ_DE
98. Butter in einem Topf erhitzen. Kohlrabi darin kurz andünsten. 5-6 EL Wasser zugießen. Mit Salz und Pfeffer würzen. Zugedeckt bei schwacher Hitze ca. 10 Minuten dünsten.	dünsten	Kohlrabi	Wasser	N_A	Topf	zugedeckt	bei schwacher Hitze	N_A	N_A	REZ_DE

Table 1. (Partial). Annotation of the *REZ_DE* corpus concordances (occurrences) of the verb *dünsten*

However, the parameters *Ingredient* and *Dish* have not been considered for the final analysis of the annotation due to their very high number of separate collocates, which practically makes meaningful generalization impossible. This is also partially a result of the absence of a general category. By "very high number" it is meant that the ingredients are so numerous and so ontologically different that they become statistically irrelevant and presumably have less effect on determining the choice of the verb. The parameter *Dish* has been excluded from further analysis for two reasons: first, some dishes are general categories themselves, having only a few hyponyms (e.g. *"Eintopf"* for *Gemüseeintopf*, *Fleischeintopf*, *Kartoffeleintopf*, etc. or *"Suppe"* for *Gemüsesuppe*, *Hühnersuppe*, *Süppchen*, *Kürbissuppe*). Second, the distribution of the *Dish* parameter is so vast that many dishes would at some point become *Ingredient* due to the generalization. For instance, *Salzkartoffeln kochen* and *Pellkartoffeln kochen* would then be generalized and neutralized as *Kartoffeln kochen*, being regarded as *Ingredient Kartoffel*. Similarly, all dishes with some sort of pasta, such as *Spaghetti Bolognaise* and *Pasta Carbonara*, would ultimately be generalized as *Nudeln* or *Pasta kochen*, which are already

[53] For all annotated occurrences of the German *deTenTen13* and *REZ_DE* sub-corpora see Appendix 1.

present in the parameter *Ingredient*. Consequently, the parameters *Ingredient* and *Dish* did not yield results that made a statistically relevant difference in the determination of the culinary verb choice in the given culinary context.[54] Thus, for both the intra- and interlinguistic analysis, six of the eight parameters were taken into consideration since their effect was strong enough to determine or explain the verb choice.

The table below illustrates the eight parameters serving as the basis for the manual semantic annotation of the occurrences extracted from the respective German, English, and Armenian sub-corpora(Table 2).

Parameter	**Explanation**	**Examples**
Ingredient	Mostly represented as a direct object (nouns in the accusative case) in recipes and culinary text in general. All ingredients are viewed as intermediary products and differ from their counterparts under the category *Dish* in that they are not ready to eat, at least in that specific recipe and/or culinary text.	<u>Geschälte Kartoffeln</u> in Salzwasser kochen. *(REZ_DE)* [Cook* <u>the peeled potatoes</u> in boiling salt water].
Substance	This parameter shows in what kind of substance the ingredient needs to be cooked or the absence of any cooking substance. *Cook* here is to be understood as any of the culinary verbs in the three languages in question and is synonymous with 'to be prepared'.	<u>Die Butter</u> in einer beschichteten Servier-Pfanne erhitzen. Die Zwiebelwürfel <u>darin</u> glasig dünsten. *(deTenTen13)* [Heat <u>the butter</u> in a coated serving pan and sauté* the diced onions in it until translucent].
Dish	This parameter includes exclusively dish names, e.g. *Eintopf*, *Fleischauflauf* or something edible that could be classified as an end-product at	<u>Suppe</u> abdecken, bei mittlerer Hitze ca. 40 min. kochen lassen. *(REZ_DE)*

[54] Some verbs were later found to be also determined by the parameter *Ingredient*, which will be discussed in detail in Chapter 2.4.

	the end of the recipe. For instance, *pasta* was annotated as a *Dish* if it was ready to eat at the end of the recipes and was not used in its raw form as an intermediary ingredient to prepare the dish.	[Cover and let the soup cook*/simmer* over medium heat for about 40 minutes]. In einer anderen Pfanne Butterschmalz zerlassen und ein Spiegelei braten. *(REZ_DE)* [In another pan, melt the ghee and fry* the egg sunny-side-up].
Utensil	This parameter represents all types of containers the ingredient/s is/are cooked in. A rather informative category playing an important role in the choice of the verb in the text production. For example, *Pfanne* (pan) would be used more with the verb *braten* as it is rather shallow, while *Topf* (pot) would be used with *kochen* because the latter is often done in water, which requires a deeper container. However, the following explanation is not absolute and variations of utensils have been observed throughout the corpora annotation, where e.g. *kochen* has also been possible in a *Pfanne*.	In einem Topf jeweils 2 EL Butter und Olivenöl erhitzen, die Schalotte darin bei mittlerer Hitze dünsten. *(deTenTen13)* [Heat 2 tbs. of butter and olive oil in a pot and sauté* the shallots over medium heat]. Braten Sie den Speck in einer Pfanne kross. *(deTenTen13)* [Fry* the bacon in a pan until crispy]. Bratentopf erhitzen, Fleisch hingeben und scharf anbraten, ca. 1 Liter Wasser dazugeben und eine Stunde schmoren lassen, bis das Fleisch gar ist. *(REZ_DE)* [Heat the Dutch oven, sear* the meat, add around 1l water and braise*/stew*/simmer it for 1 hour until the meat is ready].

Manner	This parameter shows how the dish or the ingredient has been cooked*.	Lassen Sie das Ganze dann <u>zugedeckt</u> ca. 15 Minuten schmoren und gießen dann soviel Wasser dazu, dass das Gulasch knapp bedeckt ist. *(deTenTen13)* [<u>Cover</u> and let everything braise*/simmer* for 15 minutes and pour so much water to hardly cover the goulash]. Rösten Sie dafür den Sesam <u>leicht</u> in der Pfanne und stellen Sie ihn beiseite. *(deTenTen13)* [<u>Lightly</u> roast* the sesame seeds in a pan and put them aside]. Butter in einer Pfanne zerlassen, darin die Brotscheiben <u>von beiden Seiten</u> rösten. *(deTenTen13)* [Melt the butter in a pan and roast* the slices of bread in it <u>on both sides</u>].
Heat intensity	This parameter illustrates different heat intensities observed throughout the annotation, further generalized to 3 main values of the category: low, medium, and high.	Zwiebelringe hineingeben und <u>bei mittlerer Hitze</u> 3 Min. glasig dünsten. *(deTenTen13)* [Add the onion rings and sauté* over medium heat for 3 minutes until translucent].

		Traditionell lässt man diese Eintöpfe sehr lange <u>bei niedriger Hitze</u> schmoren, sodass sich der Geschmack der einzelnen Zutaten richtig entfalten kann. *(deTenTen13)* [Traditionally one let's these kind of stews braise* over low heat for a long time so that the flavors of separate ingredients could be unfolded].
Resultative (adjectives)	Initially this parameter was considered as part of the *Manner* parameter. However, a clear-cut difference between the two is that the parameter *Manner* shows the process and the way of cooking something, while *Resultative* (adjectives) — the final result, the "appearance" of the ingredient or the dish. Hödl discussed the structure verb + resultative adjectives in her attempts to compare German verb compounds with their French and Spanish equivalents on examples such as 'fein schneiden', 'gar kochen', 'goldbraun braten', etc. (Hödl 2004:1–3).	Öl in einem großen Topf erhitzen. Ingwer, Zwiebel und Knoblauch darin <u>glasig</u> dünsten. *(REZ_DE)* [Heat the oil in a large pot and sauté* the ginger, onion and garlic <u>until translucent</u>]. Eine Pfanne mit Butter erhitzen und die Brotscheiben darin <u>goldbraun</u> rösten. *(deTenTen13)* [Heat the butter in a pan and roast* the bread sliced <u>until golden brown</u>]. Die Kartoffel <u>bissfest</u> garen, pellen und als einzeln Scheibe in einer großen Pfanne mit Bratöl von beiden Seiten <u>kross</u> braten. *(REZ_DE)* [Cook*/boil* the potatoes <u>al dente</u>, peel and fry* them in

		separate slices in frying oil on both sides until crispy].
Heat source	This parameter differentiates the source of heat used to cook* a given dish, as mentioned in the recipe or any culinary text. For some verbs, e.g. *grillen* or *backen*, this was a determining parameter and had a strong effect on their distribution.	Den Kuchen auf dem Blech im heißen Ofen 15 Minuten *backen*, bis der Teig zu bräunen beginnt. *(REZ_DE)* [Bake the cake*/pie* in the oven (on a baking tray) for 15 minutes until the dough starts to get brown]. Rösten Sie ein paar Marshmallows über dem Lagerfeuer und… *(deTenTen13)* [Roast* a couple of marshmallows over the campfire and …].

Table 2. Annotation parameters based on the German corpora examples

The initial analysis of both of the German sub-corpora, annotated separately, raised the necessity to see if, in reality, the two differed and, if they did, to what extent. The underlying assumption was that the differences should have been minor since the structure of both recipes produced by professionals and culinary texts in general were almost the same. Therefore, the two corpora were compared with regard to the above-mentioned six parameters (without *Ingredient* and *Dish*), viz. *Substance*, *Utensil*, *Manner*, *Heat intensity*, *Resultative (adjectives)*, and *Heat source*. However, in order to ensure an objective comparison between the *deTenTen13* and *REZ_DE* sub-corpora, the annotated occurrences were reduced according to the lowest number of examples for each German culinary verb, irrespective of the sub-corpus, resulting in the same amount of annotation for both of them. For instance, for the German culinary verb *backen*, 204 and 48 occurrences– respectively from the *deTenTen13* and *REZ_DE* sub-corpora – were initially annotated. For the comparison of the two sub-corpora, 48 random examples of *backen* from both *deTenTen13* and the *REZ_DE* were taken into consideration, as no more than overall 48 examples of *backen* in the *REZ_DE* could be initially annotated. The verb *toasten* was excluded from the comparison due to the same reason: the absence of any annotated examples in the *REZ_DE* corpus. Consequently, the initially annotated 1392 and 607 occurrences respectively[55] for the *deTenTen13* and *REZ_DE* sub-corpora were reduced to 387 for each corpus (see the Tab. 3 below).

Verb	**backen**	**braten**	**dünsten**	**grillen**	**kochen**	**rösten**	**schmoren**	**toasten**
REZ_red	48	101	90	33	115	61	62	0
TT_red	48	101	90	33	115	61	62	0

Table 3. Reduced number of annotated examples for the comparison of the *deTenTen13* and *REZ_DE* sub-corpora

In all of the aforementioned annotation parameters the differences between the two annotated corpora were statistically insignificant so that, in the end, the annotated examples were compiled together. Further analyses were carried out based on a single joint corpus for every language instead of two. The following mosaic plot (Fig. 4) illustrates the comparison between the annotated examples of all the verbs of the *REZ_DE* (marked REZ_red) and in *deTenTen13* (marked TT_red) in relation to the parameter *Resultative* (adjectives). The grey columns show that there are no statistically significant over- or underrepresentations of the *Resultative* (adjectives) parameter values in the two sub-corpora. The slightly different width of the columns represents merely the difference in the distribution of each of the *Resultative* (adjectives) parameter values, which, however, is of no statistical significance, with all values below the 2.0-fold of the standard deviation.

[55] All annotated occurrences of the German *deTenTen13* and *REZ_DE* sub-corpora in Appendix 1.

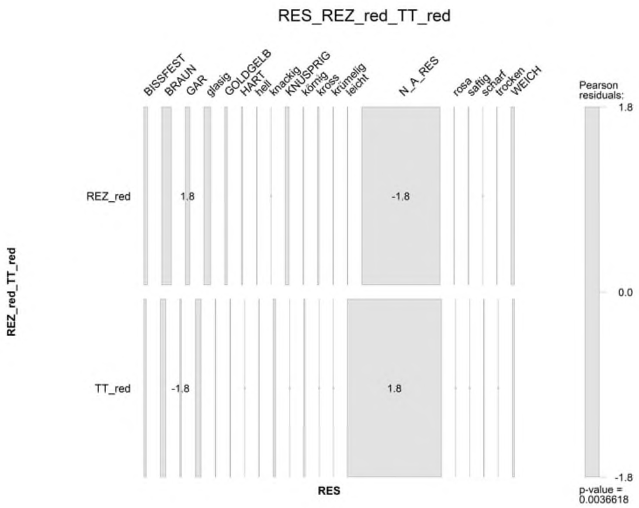


Fig. 4. The *Mosaic-Plot* of the *REZ_DE* and *the deTenTen13* sub-corpora in relation with the parameter *Resultative* (adjectives)

The distribution of all *Resultative* (adjective) parameter values were nearly the same for both corpora, as illustrated in the *Mosaic-Plot* (Fig. 4). The vertical columns here denote the number of annotated examples for the *REZ_red* and *TT_red* corpora being of the same amount and consequently illustrated of the same height. The width of the columns marked in color grey denotes that there were statistically no significant over- or underrepresentations between the separate values in the aforementioned *Resultative (adjectives)* parameter. Even though the distribution of values such as {gar} (cooked through) and {N_A_SUB} (denoting no mention or absence of any cooking substance) differed slightly in the annotated examples of the two corpora, the difference was still below the statistically significant deviation.

No statistically significant differences were observed in most of the six parameters being compared, except for the *Substance*, *Manner*, and *Utensil*.[56] Nonetheless, the observed statistically significant differences in the separate values of the aforementioned parameters were minor; merely slightly above the standard deviation and occurring in no more than two values of one and the same parameter (cf. Fig. 4 and Fig. 5).

The distribution of the annotated examples between the *REZ_red* and *TT_red* sub-copora just above the 2.0-times the standard deviation was observed, for instance, in the parameter *Utensil* (Fig. 5), viz. with the value {Spieß} (marked light blue) being slightly overrepresented in the *deTenTen13* (TT_red) sub-corpus and consequently slightly underrepresented in the *REZ_DE*

[56] For all *Mosaic-Plot* visualization of the comparison between both the German *deTenTen13* and *REZ_DE* as well as the English *enTenTen15* and *EN_REZ* in all the relevant annotated parameters see Appendix 6.

(REZ_red) (marked light red). The other *Utensil* parameter values marked grey show statistically no significant over- or underrepresentations.

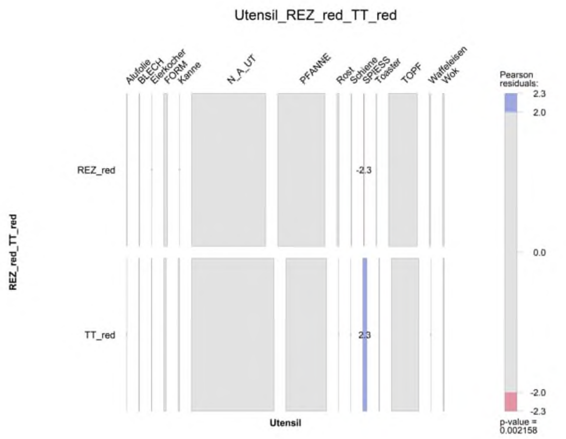


Fig. 5. The *Mosaic-Plot* of the *REZ_DE* and *deTenTen13* sub-corpora in relation to the parameter *Utensil*

2.2 Neutralization (Generalization) of the Annotation Parameters and Reduction of the Annotated Occurrences

It was already stated throughout the description of both the *deTenTen13* and the *REZ_DE* sub-corpora that certain parameter values were generalized (neutralized) to yield more visible results and to identify the highest effects of parameter values on the choice of a given culinary verb. The aforementioned annotated examples too, for instance, in the description of the annotation parameters (Table 1), revealed that for the verb *dünsten* several types of different cooking oils were annotated as *Substance* parameter values, e.g. {Kokosöl, Sonnenblumenöl}, and {Butter, Bratfett}, just to name a few. However, for further analysis {Kokosöl, Sonnenblumenöl} was generalized as {ÖL} while {Butter, Bratfett} as {FETT}. The same way, for instance, the *Heat intensity* parameter values {auf kleiner Flamme, bei niedriger Hitze, bei schwacher Hitze, bei geringer Hitze, bei/auf niedriger Temperatur, auf kleiner Stufe, auf kleiner Hitze} were neutralized (generalized) as {niedriege Hitze} (low flame) to allow for a more consistent analysis of the annotated data. For neutralization (generalization), either a more general category value, a hypernym, was selected (the case of {Öl} and {Fett}), or semantically synonymous values were grouped together (as in {niedriege Hitze}).[57] However, even after

[57] All neutralized (generalized) parameter values appear capitalized (e.g. {FETT}) in all visualizations techniques (e.g. M(CA), CCCG, and CIT) throughout this work to avoid double generalizations caused by the case-sensitiveness of *R*. Not generalized values remained not capitalized, e.g. {Tomatenmark} of the Substance parameter.

comprehensive and detailed generalization, the statistical analysis generated too much noise due to the high number of annotated parameters on one side and many different values of certain parameters on the other side. Therefore, in order to generate clearer visualizations as well as to have a more representative selection of information, annotated examples were reduced according to the frequency of the parameter values and their effect/relevance on the description of the given culinary verb. This is the reason why different parameter values were reduced based on the frequency of individual values. The reduction was carried out for the above-mentioned six parameters, without the parameters *Dish* and *Ingredient*.[58]

The data set reduction procedure resulted in an overall reduction of the annotated occurrences of both the *deTenTen13* and the *REZ_DE* sub-corpora (combined as one corpus) from 1999 to 1809 examples.

The neutralization as well as the dataset reduction was also carried out for both the English and the Armenian annotated sub-corpora, which will be elaborated on in Chapters 3.1 and 41 of this work.

2.3 Introduction to *Correspondence Analysis* (CA) and *Multiple-Correspondence Analysis* (MCA) of the German culinary verbs

Correspondence Analysis (CA) contributed to the visualization of the distribution of the German culinary verbs with regard to relevant parameters. Overall, the extracted occurrences were annotated according to nine parameters, namely *Ingredient*, *Substance*, *Dish*, *Utensil*, *Heat intensity*, *Manner*, *Results*, and *Heat source*. However, since one of the major objectives of the research was to suggest possible equivalents to the German culinary verbs, the *CA* was carried out with regard to the parameter *Verb* as a *dependent variable* in relation to the relevant parameters considered as *independent variables*, using the technique of stacked tables to integrate together multiple independent variables in a single contingency table (cf. Greenacre

[58] Two criteria served as the basis for reducing the annotated parameter values: the considerably huge difference between the frequencies of the occurrences annotated in that specific parameter (between the preceding and succeeding collocates), and the maximal coverage for all the frequencies the specific value provides insuring minimal loss for the further analysis and visualization of the corpora. For instance, the *Heat source* parameter was reduced to the value {Glut} (ambers) as the difference in frequencies between {Glut} and the succeeding collocate {Waffeleisen} (waffle maker) was considerably large, viz. eight annotated occurrences. The parameter *Substance*, for examples, was reduced to the value {Wein}(wine) as a minimum of 97% of the annotated occurrences were covered. However, both criteria of parameter reduction insure at least around 97%-98% coverage of the initial annotated occurrences, in some cases even up to 99%, e.g. for the *Heat intensity* parameter being reduced to the value {hohe_Hitze} (high heat).

and Blasius 2006:21). For a more detailed description, different sets of relevant parameters were analyzed in relation to the eight German culinary verbs, i.e. *backen, braten, dünsten, kochen, grillen, rösten, schmoren,* and *toasten*.

For the overall *M(CA)* 1999 contextualized examples from the original German texts were previously manually annotated according to the aforementioned eight parameters. However, due to the neutralization and reduction of the parameters as described in Chapter 2.2, the number of final annotated examples was reduced to 1809 occurrences. As mentioned before, these examples were extracted from both the *German Web 2013 (detenten13)* available in *Sketch Engine*, and the *REZ_DE* corpora compiled on our own from cookbooks available either in print or digital format.

But what is a *CA* (Fig. 1)? How is this two-dimensional plot interpreted? CA is described as a graphical visualization of the analysis of contingency tables displaying the relations among rows and columns. The graphs are generated with the help of the programming language *R* (see Gries 2013, 2021), displaying the synthesized data of the rows and columns on "two-dimensional scatterplots" in form of dots. The distance between dots of the same type is related to the degree of similarity of the "profiles" the rows and columns have (Alberti 2013:27). Here, the frequencies of the values in respective rows and columns play a large role, meaning that a large number of dots on the same spot have a higher probability of appearing on the graph than those with fewer dots.[59] It is important to note that initially, the profiles of rows and columns are thought to be neutral, with no difference between the profiles and all values positioned near the intersection[60] of the vertical and horizontal axes of this two-dimensional visualization graph. More similar dots, thus similar data information, result in similar profiles and near-to-each-other dots. Different-from-each-other profiles "are pulled" from the intersection and are scattered across the four quadrants of this two-dimensional graph. The same is true for the independent as well as the dependent variables; the former representing our annotation parameters while the latter the culinary verbs in their respective languages. Two criteria visualize strong correlations among independent and dependent variables: the distance of the dots from the intersection and the angle formed by drawing imaginary (not present on the graph)

[59] In discussing the advantages of the *Correspondence Analysis* highlighting its visualization effects, Jenset and McGillivray called the projection of the row and the column point into the same space as a 'biplot' and the data interpretation of each dimension (X and Y axis) as 'percentage explained inertia', thus resulting in a "full representation of the X and Y axis together" (Jenset and McGillivray 2013:317). Thus, the sum of the 'percentages of explained inertia' in *Correspondence Analysis* reflects the overall data analysis. In the same article, Jenset and McGillivray argued that "principle component analysis and correspondence analysis are the techniques best suited for corpus linguistics" (ibid. 301).

[60] Alberti named it "centroid" (2013:27).

lines between two dots (from independent and dependent parameter values each) and the intersection. The further the dots are from the intersection and the sharper the angle is, the stronger the correlation.

Nonetheless, a strong correlation between certain independent and dependent variables might generate an "outlier" (Alberti 2013:28, cf. Di Franco 2016) due to its significantly large difference in profiles compared to the other strong correlations. For instance, the German culinary verb *schmoren* (a dependent variable) creates a strong correlation among others with the *Substance* parameter values (independent variables) {Tomatenmark} (tomato paste) and {Wein} (wine). However, the latter two values are illustrated on the CA graph as outliers placed at the top right-hand corner of the graph (Fig. 1a, cf. Table 1) as there is a significantly large difference in the distribution of the *Substance* parameter values {Tomatenmark} and {Wein}: almost all occurrences are with *schmoren*, except for just one example with *kochen*. No other verb shares a similar profile, except for *kochen* which is the reason why *kochen* is also situated near *schmoren*, due to the additional correlation with {Brühe, Wasser, N_A_SUB}.[61] Table 1 below illustrates how outliers are generated in *M(CA).* In our analysis, we represent dependent variables (verbs) as qualitative supplementary variables, which are not taken into consideration for the analysis as such, but rather placed on the plot a posteriori.

	BRUEHE	FETT	N_A_SUB	OEL	**Tom_Mark**	WASSER	WEIN
bac	0	6	232	4	0	0	0
bra	0	62	42	154	0	0	0
dün	14	87	31	113	0	21	0
gri	0	2	191	22	0	0	0
koc	13	0	177	2	1	90	1
rös	0	23	182	25	0	0	0
sch	39	7	102	17	**18**	29	**16**
toa	0	0	85	1	0	0	0

Table 1. Contingency table of the dependent and independent variables

Multiple Correspondence Analysis (MCA) visualizes all parameter combinations of the contingency table; thus "the convergences and divergences of the independent variables are also taken into consideration" (Atayan and Gast 2024). Clusters identified on the *MCA* visualization graph might denote either a) strong correlations between independent variables of

[61] *MCA* is in general more affected from outliers than *CA*. "An important characteristic of MCA is its marked sensitivity towards small variations and perturbations present in the data" (Di Franco 2016:1304).

two different annotation parameters, e.g. how often the parameters *Substance* and *Utensil* cross among each other (e.g.{Öl} x {Pfanne}) (Fig.1b) or b) similar profiles, viz. different values of one and the same parameter having the same syntagmatic distribution, e.g.{Öl}(oil) and {Fett} (fat), both (separately) correlating with {Pfanne} (pan). The correlation of the dependent variables and the identified clusters of independent variables illustrated on the *MCA* graph (Fig.1b) is more of a side effect since their presence near the cluster/s is the result of strongly correlated independent variables (the parameter values in red) bringing the dependent ones (the verbs in green) with them into the cluster/s. In other words, the image of the *MCA* graph would remain unchanged if the dependent variables were left out of the analysis. Unlike *CA* (Fig.1a), *MCA* (Fig.1b) shows no direct correlation between the independent and dependent variables. Different values of one and the same parameter appear near each other on the *MCA* graph, as they have similar profiles, which could be traced in *R* by calculating the deviation in distribution based on Pearson residuals.[62] The aforementioned statements are true under the general condition that both the parameter and the verb in question are away from the intersection point of the two axes. The parameters gathered around or near the intersection are less significant for the total variance. As described in Atayan and Gast (2024), *CA* is more like a "diagnostic-inductive method" where, from certain contextual elements (markers) such as *Heat source* or cooking *Substance*, the correct culinary verb may be identified. Different independent variables appear together on the *CA* graph as they separately correlate with the dependent variables, the specific verb(s). *CA* does not generate correlations between the independent variables. The *MCA*, on the contrary, serves as "prognostic-deductive"/cause-effect analysis method where conclusions are made on "causal inferences". Whereas in *CA* it is the other way round: the cause is attempted to be found based on the effects. Atayan and Gast (2024) draw parallels between *CA* and medical diagnosis, where from different symptoms a certain disease is diagnosed or identified. *MCA* is then like a weather forecast, where different factors such as air humidity, pressure and temperature foresee the weather as a causal inference.

[62]Additionally, *MCA* displays strong correlation of parameters with the color intensity/saturation of the red triangles. Blurry, less intense red small triangles represent weaker while the more intense ones the stronger correlations. Indirectly, the redder the triangle, the better is the quality of representation of the dot on the two-dimensional map under the condition, as mentioned before, the sharper the angle between the parameter and the verb, the stronger the correlation. The aforementioned statements are true under the general condition that both the parameter and the verb in question are away from the intersection point of the two axes (see Atayan et al. 2019, Atayan and Gast 2024).

2.3.1 CA vs. MCA the German Culinary Verbs

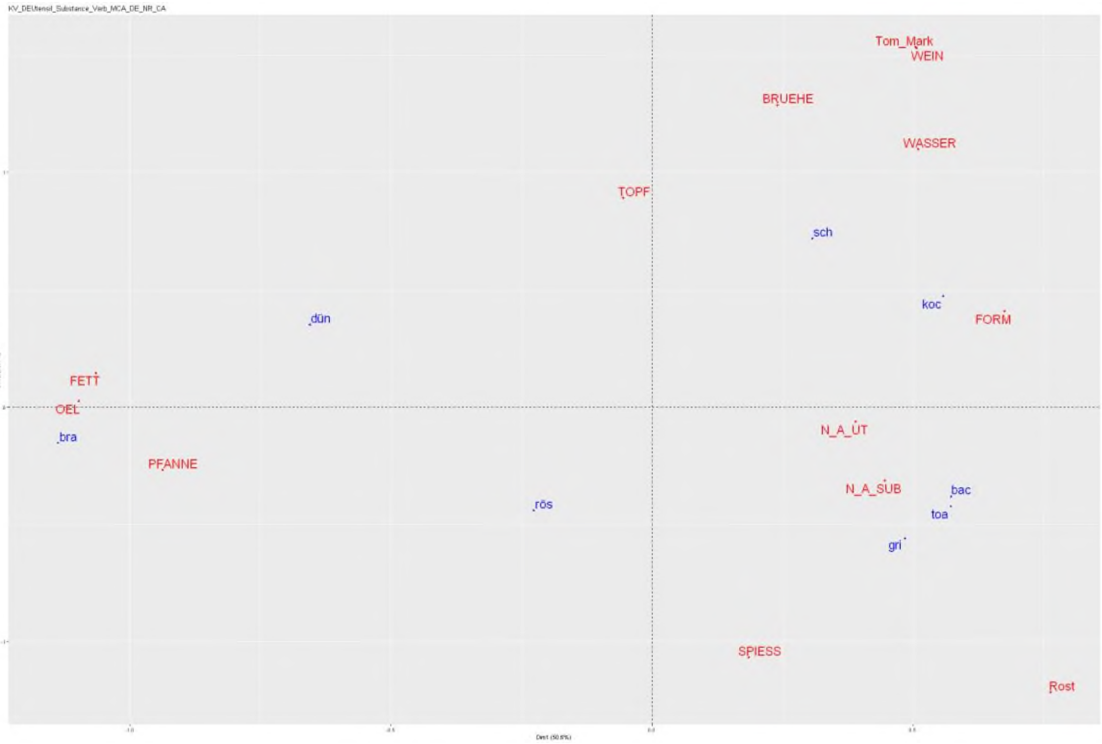


Fig. 1a. CA: German culinary verbs (*Substance* and *Utensil* parameters)

Both the *CA* and the *MCA* (Fig. 1a. and Fig.1b) illustrated nearly the same distribution of the *Substance* and *Utensil* parameter values correlating with the respective verbs. The *CA* (Fig.1a) as well as the *MCA* (Fig.1b) graphs are divided into three parts according to three large clusters. The upper-left and upper-right sides of the *CA* graph (Fig. 1a) show the correlations grouped around some type of *liquid* cooking substance and {Topf} (pot) or {Form} (baking form/dish). The lower-right part of the *CA* graph exhibits the strong correlations gathered around {N_A_Sub} and {N_A_UT}, while the lower-left one shows those with some type of *oil*-like or *fat*-like cooking substance (Fig.1a).

The verbs *backen*, *toasten*, and *grillen* correlate with the {N_A_SUB} value of the *Substance* as well as the {N_A_UT} of the *Utensil* parameters respectively (examples 1-3), thus identifying that these three verbs are very often used in culinary contexts where there is *no mention of any type of substance and utensil*.

1. Den Kuchen ungefähr 45 Minuten *backen*. *(REZ_DE)*
   [Bake the cake*/pie* for 45 minutes].

2. Am schönsten ist es, Steaks auf offener Flamme zu *grillen*. *(detenten13)*
   [The best is to grill the steaks over open fire].

3. Das Toastbrot toasten und abkühlen lassen. *(detenten13)*
   [Toast* the toast bread and let it cool down].

The verb *rösten* correlates with both {Fett, Öl} (fat, oil) values of the *Substance* and the {Pfanne} (pan) value of the *Utensil* parameter (examples 4-5).[63]

4. Sie *rösten* die Kaffeebohnen in 1-2 Min. in einer *Pfanne*. *(detenten13)*
   [One roasts* the coffee beans in a pan in 1-2 minutes].

5. In einer *Pfanne* mit *Öl*, 1 Knoblauchzehe und die Brotscheiben rösten. *(detenten13)*
   [Roast* 1 garlic clove and the bread slices in a pan with oil].

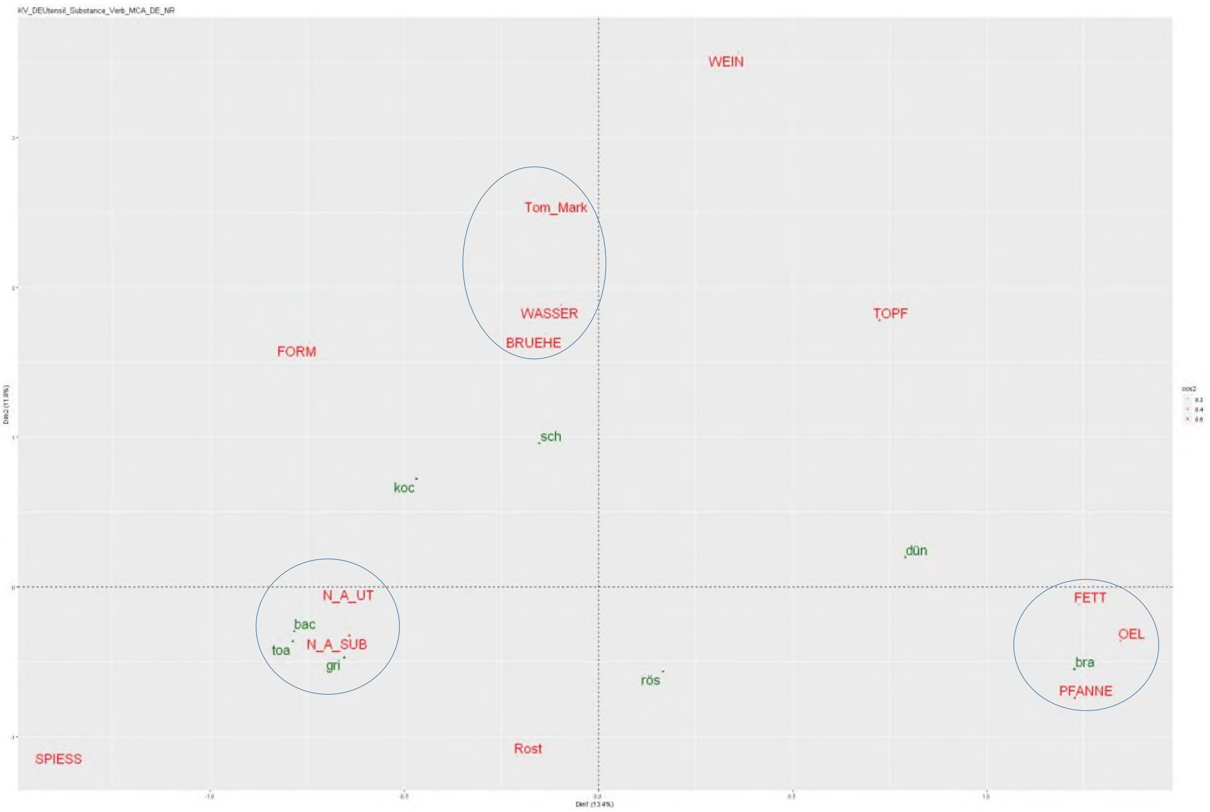


Fig. 1b. MCA: German culinary verbs (parameters *Utensil* and *Substance*)

[63] The peculiarities of the German verb *rösten* correlating with both verbalized and non-verbalized values of the *Substance* and *Utensil* parameters are thoroughly discussed in Chapter 2.4.2 of this work.

The MCA of the German culinary verbs (Fig. 1b) illustrates another cluster of independent parameter values having similar profiles, i.e. {Öl and Fett} both having a similar syntagmatic distribution, viz. correlating often with {Pfanne} (examples 6-8).

6. Das *Öl* in einer beschichteten *Pfanne* erhitzen und die Frikadellen darin rundum braun *braten*. *(detenten13)*
   [Heat the oil in a coated pan and fry* the patties on all sides until brown].

7. Eine Pfanne mit *Olivenöl* erhitzen und den Spargel darin 5 Minuten *braten*. *(REZ_DE)*
   [Heat the olive oil in a pan and fry* the asparagus for 5 minutes].

8. In einer großen *Pfanne* das *Fett* erhitzen, die Kartoffeln kross braten und den Zucker zugeben. (detenten13)
   [Heat the fat in a large pan, fry* the potatoes until crispy and then add the sugar].

Another cluster of independent variables, viz. the {Tomatenmark, Wasser, Brühe} (tomato paste, water, stock) values of the *Substance* parameter, is illustrated on Fig. 1b due to their profile similarities, which then "pulls" the verb *schmoren* as a qualitative supplementary variable near to them (example 9).

9. Zugedeckt bei kleiner Flamme unter Zugabe der *Fleischbrühe* und dem *Weißwein* ca. 45 Minuten *schmoren* und durch ein Sieb passieren. *(REZ_DE)*
   [Braise* it closed*/covered* for 45 minutes over low heat by adding the *stock* and the white *wine* and press through a strainer].

As a result of the overestimation of the most relevant parameter values to emphasize their statistical significance on the distribution, three outliers are observed on the MCA graph (Fig.1b). {Spiess} and {Rost} (spit and grid/rack) of the *Utensil* parameter, which also create a strong correlation with the {N_A_SUB} value of the *Substance* parameter, are illustrated here as outliers for the verb *grillen* (example 10-11).

10. Die Brötchen kurz auf dem Grill rösten, die Burger auf dem *Rost* von jeder Seite 5 Minuten *grillen*. *(detenten13)*
    [Toast*/roast* the buns briefly on the grill, grill the burgers on the grid for 5 minutes on each side.]

11. Das gewürzte Hackfleisch zu 4 *Spießen* formen und auf einem Grill oder Rost 5-7 Minuten von jeder Seite *grillen*. *(detenten13)*
    [Shape the seasoned minced meat into 4 skewers and grill on a barbecue or grill for 5-7 minutes on each side]

The *Substance* parameter {Wein} creates an outlier correlating with the verb *schmoren* and being positioned on the MCA graph (Fig.1b) separate from the cluster of other *liquid*-like substances having similar profiles correlating with the verb *schmoren* as well (see example 9 above).

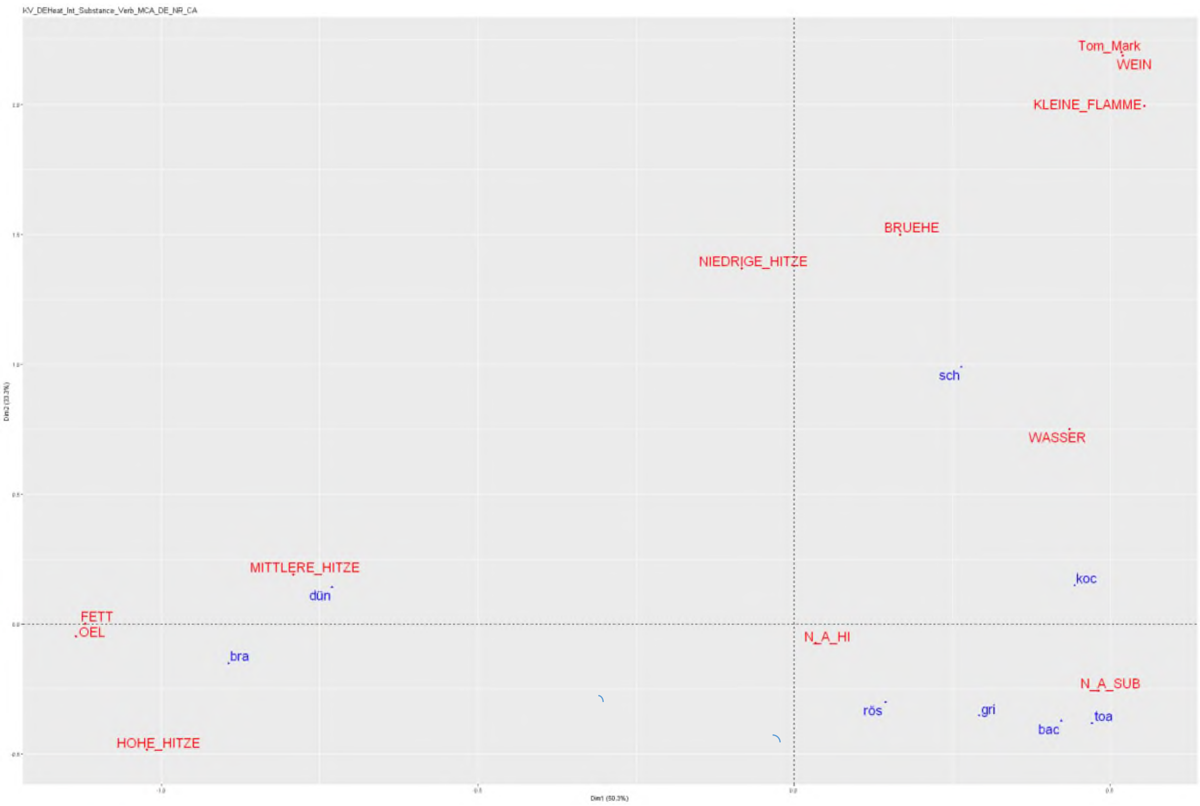


Fig. 2a. CA: German culinary verbs (parameters *Substance* and *Heat intensity*)

This *CA* graphical visualization (Fig. 2a) demonstrates the correlations between the independent and the dependent variables, viz. the German culinary verbs and the values of the *Substance* and *Heat intensity* parameters. The *CA* graph could again be visually divided into 3 parts according to *liquid*, *oil*, and *absence* or *non-verbalization* of any substance. {Fett} and {Öl} of the *Substance* parameter values correlate with the verbs *dünsten* and *braten*. Nevertheless, there is a clear distinction showing *braten* to be “pulled” from the intersection by

{Hohe_Hitze} (high heat)[64] (example 12), while *dünsten* by {Mittlere_Hitze}(medium heat) (example 13) of the parameter *Heat intensity*. *Schmoren* appears to be strongly correlated with the {kleine_Flamme} (low flame) and {niedrige_Hitze} (low heat) values of the *Heat intensity*, as well as {Brühe,Wasser, Tomatenmark, Wein} (stock, water, tomato paste, wine) of the *Substance* parameters, that is, with *liquid* substances (example 13).

12. Im Wok oder einer großen Pfanne portionsweise *Öl* erhitzen und das Fleisch *bei starker Hitze braten*. *(detenten13)*
    [Heat oil portionwise in a wok or a large pan and fry* the meat over high/strong heat].

13. Die *Brühe* angießen und den Kohl ca. 60 Minuten *bei mittlerer Hitze schmoren* lassen. *(detenten13)*
    [Pour the stock and let the cabbage braise*/simmer* for 60 minutes over medium heat].

[64] The German {Heat_intensity} parameter value {Hohe_Hitze} comprises also such values as {bei starker Hitze} and {bei große Hitze}.

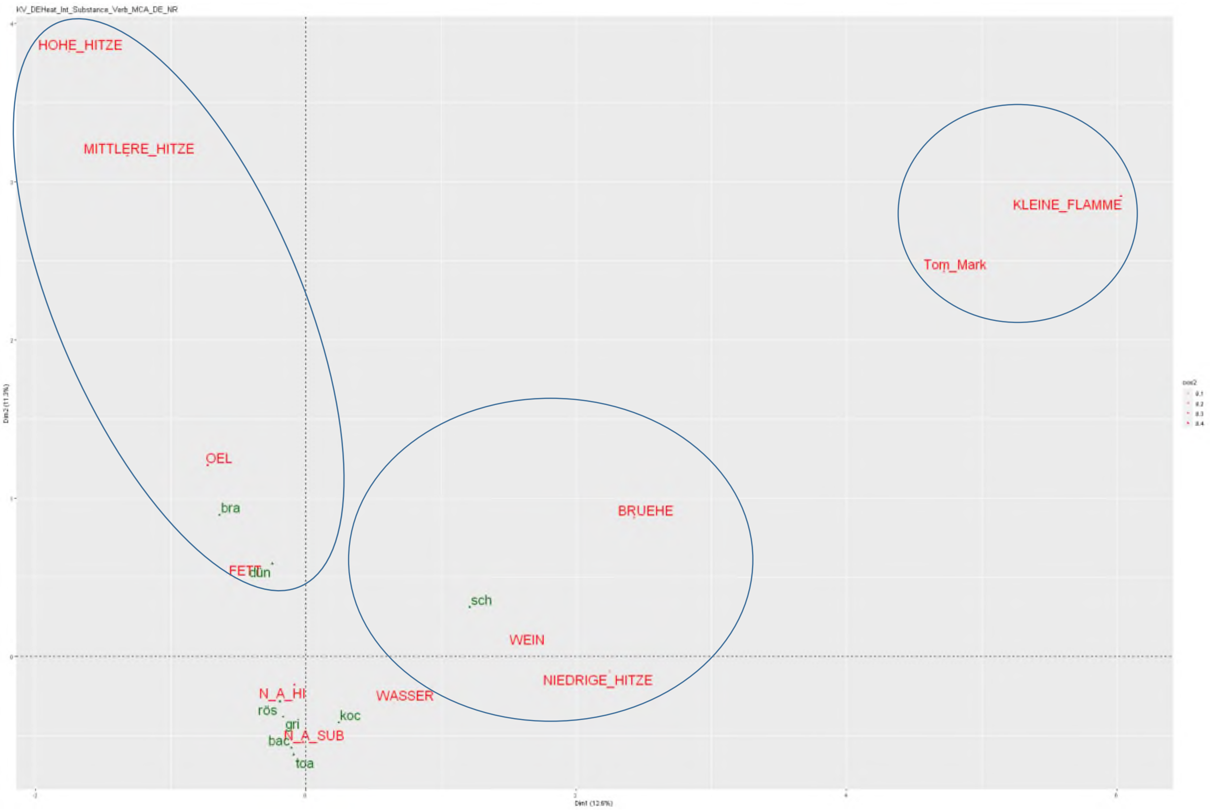


Fig. 2b. MCA: German culinary verbs (parameters *Substance* and *Heat intensity*)

The *Heat intensity* parameter values {hohe Hitze} (high heat) and {mittlere Hitze} (medium heat) appear on the MCA graph (Fig. 2b) to have similar profiles having similar syntagmatic distribution, viz. co-occurring with the *Substance* parameter values {Öl, Fett} (oil, fat). The verbs *braten* and *dünsten* are also included in this cluster as strongly correlated *Substance* and *Heat intensity* parameter values (at the left-hand side of Fig. 2 b), as qualitative supplementary variables with {hohe Hitze} (high heat) illustrated as an outlier for the verb *braten*. The *Substance* parameter value {Wein} (wine) forms a cluster with {Wasser} (water) and {Brühe} (stock) due to their profile similarities, as well as their strong correlation with the *Heat intensity* parameter value {niedrige_Hitze} (low heat) (Fig. 2b). On the upper left-hand corner of the MCA graph (Fig.2b.), a cluster of strongly correlated values of {Tomatenmark} (tomato paste) and {kleine_Flamme} (low/small* flame) of the *Substance* and *Heat intensity* parameters respectively is illustrated, both being overestimated as the most significant parameter values for the verb *schmoren*, thus generating an outlier.

Therefore, as mentioned at the beginning of this chapter, *CA* is more appropriate for identifying the correct verb choice as a diagnostic process depending on the contextual elements, the annotation parameters, than the *MCA*.

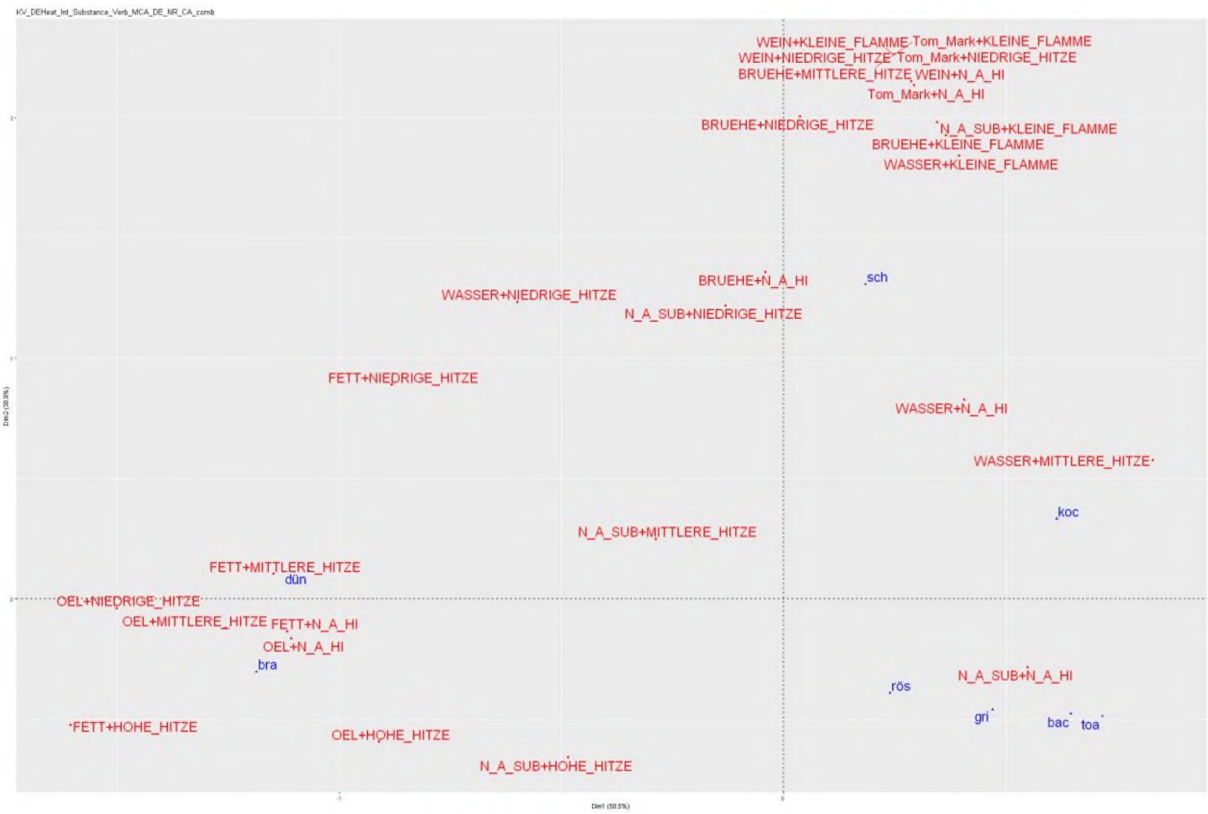


Fig. 3. CA with concatenated parameters (*Substance* and *Heat intensity*)

Another way of visualizing strong correlations between the independent and dependent variables is the *CA with concatenated parameters*, representing the cross product of both sets of values (Fig. 4). Here, too, strong correlations are identified by the sharp angle between the parameter pairs and the verb(s) as well as by the distance of the pairs from the intersection. For instance, Fig. 3 not only restates the strong correlations between the German culinary verb *braten* and {Öl}x{hohe_Hitze} (cf. Fig. 2a and Fig. 2b) but also reveals further parameter combinations strongly correlated with *braten*, i.e. those of {Fett + hohe Hitze}, {Öl + hohe_Hitze}, and {N_A_SUB + hohe_Hitze}. While it is apparent that *Substance* and *Heat intensity* parameter combinations such as {Fett + mittlere_Hitze}, {Öl + mittlere_Hitze} as well as {Öl + niedrige_Hitze} are strongly correlated with the German culinary verb *dünsten*, {Öl + N_A_HI} and {Fett + N_A_HI} appear to create strong correlation with both *dünsten* and *braten*.

Another cluster of *Substance* and *Heat intensity* concatenated parameters conveys strong correlations between *liquid* substance and {niedrige Hitze} (low heat)[65] (example 14) with the verb *schmoren.* The following cluster also illustrates that *schmoren* also strongly correlates with {N_A_SUB + kleine_Flamme} and {N_A_SUB + kleine_Flamme} as well as {Tomatenmark + N_A_HI} and {Wein + N_A_HI} (examples 15 and 16).

14. Die *Brühe* angießen und das Gulasch zugedeckt ca. 40 Min. *auf kleiner Flamme schmoren*. *(detenten13)*
    [Pour the stock, cover*/put the lid* and braise* the goulash for about 40 minutes over medium heat].

15. Die Hälfte der Marinade dazu gießen, das *Tomatenmark* ebenfalls dazu geben und das Fleisch dann für zwei bis drei Stunden schmoren lassen. *(detenten13)*
    [Add half of the marinade as well as the tomato paste and let the meat braise*/simmer for 2-3 hours].

16. Dann mit dem *Wein* auffüllen und den Braten zugedeckt ca. 3 Std schmoren lassen. *(detenten13)*
    [Then fill up with wine, cover and let the roast meat* braise*/simmer* for about 3 hours].

[65] The *Heat intensity* parameter value *{niedrige Hitze}* is generalized and might include such values as {bei schwacher Hitze}, {auf kleiner Feuer, {bei niedriger Hitze}, {auf kleiner Stufe}, {auf niedriger Stufe}. For the list of all generalizations see Appendix 2.

Fig. 4 a. CA: German culinary verbs (parameters *Utensil* and *Manner*)

The CA analysis (Fig. 4a) illustrates several strong correlations between different values of the *Utensil* and *Manner* (independent variables) parameters and the German culinary verbs (dependent variables). Among the more salient features, for instance, the verb *schmoren* strongly correlates with {zugedeckt}x{Topf, Form} (example 17), and the verb *braten* with {rundherum, wenden}x{Pfanne} (example 18) of the *Manner* and *Utensil* parameters respectively. However, the verb *rösten* is also “pulled” by the {rundherum, kurz, leicht, wenden, Seite} and {Pfanne} values of the *Manner* and *Utensil* parameter, respectively, illustrating a strong correlation here, too. The distribution of the verb *rösten* slightly differs from that of *braten* as it is also “pulled” by the *non-verbalized values* of the *Manner* and *Utensil* parameters annotated as {N_A_MAN} and {N_A_UT} respectively.

17. Die Tomate in *Topf* etwas zerkleinern, alle mit Salz, Pfeffer und Kraut würzen. *Zudecken* ca. 1 Stunde schmoren. *(REZ_DE)*
    [Hash the tomatoes a bit in a pot, season everything with salt, pepper and herbs. Cover and braise* for about 1 hour].

18. 1 EL Öl in einer großen *Pfanne* erhitzen und Garnelen darin *rundherum* ca. 6 Minuten braten. *(REZ_DE)*

[Heat1 tbs. oil in a pan and fry the shrimps in it on all sides*/all around* for about 6 minutes].

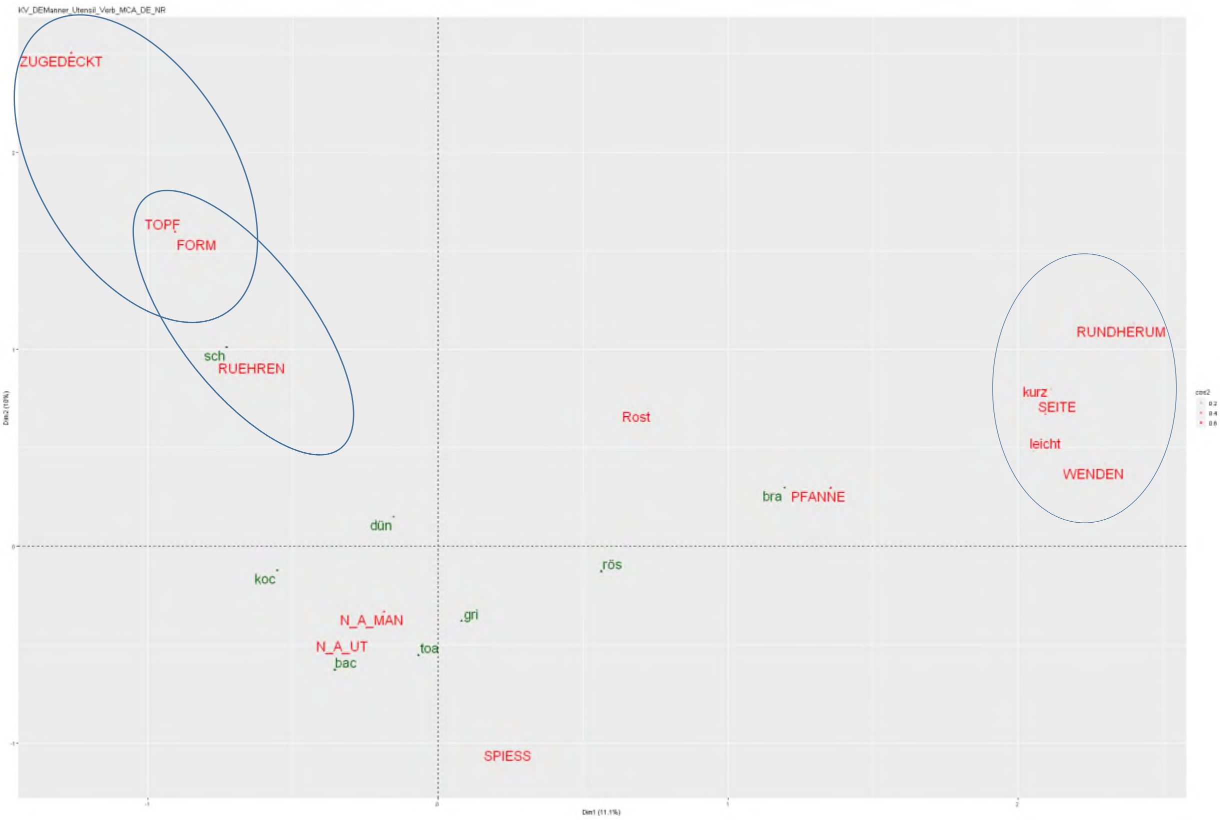


Fig .4b. MCA: German culinary verbs (parameters *Utensil* and *Manner*)

It is worth restating that in the *CA* (e.g. Fig. 4a), the two most informative dimensions of the common distribution of dependent variables (the culinary verbs in blue) and independent ones (in red) are represented. In the MCA (e.g. Fig. 4b), first, the independent variables, e.g. the *Manner* and *Utensil* parameter values, are grouped in clusters identifying their co-occurrence in the same contexts. Then the cluster "pulls" the verb(s) as qualitative supplementary variables near to it, also illustrating the correlation between the cluster and the dependent variable, the verb. Thus, for instance, on the *MCA* graph (Fig. 4b) the *Manner* parameter values {wenden, rundherum, kurz, leicht, Seite} (turn, all over, shortly, lightly, on…side) are grouped together due to their profile similarities. Moreover, the aforementioned independent variables form a cluster with the *Utensil* parameter value {Pfanne}(pan) identifying their co-occurrence in the same contexts and showing a strong correlation between them. The distribution of the *Utensil* parameter {Rost} (grid, rack) is particularly interesting. First, {Rost} is placed in the middle of two clusters of independent variables due to its profile similarity with the *Utensil* parameter values {Pfane} (pan) on one side, and {Topf} (pot) as well as {Form} (baking form, dish) on the other side. Second, {Rost} also co-occurs with the *Manner* parameter value {zugedeckt} (closed, with the lid on) (cf. examples 19-20 and 21).

19. Das Gemüse mit Öl bepinseln, pfeffern und salzen und *unter mehrmaligem Wenden* auf dem *Rost* grillen. *(detenten13)*
    [Brush the vegetables with oil, season with pepper and salt and grill* on the grid*/rack*, by turning them several times].

20. *Von beiden Seiten* auf dem *Rost grillen*, bis das Fleisch gar ist, erst dann salzen. *(REZ_DE)*
    [Grill* on both sides on the rack*/grid* until the meat is cooked*/ready* and only then season it with salt].

21. So können z.B. in einem 3 grillers joy zwei Hähnchen indirekt grillen (*bei geschlossenem Deckel*) und auf den verbleibenden Rosten dann direkt Grillgut auflegen.
    [For example, you can grill two chickens indirectly in a 3 grillers joy (with the lid closed) and then place food directly on the remaining grates.]

Thus, the positioning of {Rost} is influenced by the aforementioned clusters from the right and left sides. In addition, in the MCA (Fig. 4b) the *Utensil* parameter value {Spieß} generates an outlier for the verb *grillen* (cf. examples 21-23), while {Spieß} and {Rost} have similar profiles.

22. *Spieße* auf dem heißen Grill ca. 10 Min. grillen, bis das Fleisch schön gebräunt und durch ist, dabei einmal wenden. *(REZ_DE)*
    [Grill* the spits*/skewers* on the hot grill for 10 minutes until the meat is nicely brown and ready by turning it once].

23. Sie wollen nur ein Bier trinken und ein paar Marshmallows *auf Spieß rösten*. *(detenten13)*
    [You would like to only drink beer and roast* a couple of marshmallows on a spit].

Fig. 5a. CA: German culinary verbs (parameters *Manner* and *Resultative* (adjectives))

Both the *CA* (Fig. 5a) and the *MCA* (Fig. 5b) present more or less the same distribution of the dependent (the verbs marked in blue and green) and the independent variables (the parameter values marked in red) in relation to the parameters *Manner* and *Resultative* (adjectives). However, since *CA* illustrates the correlations between the independent and dependent variables, while *MCA* identifies clusters of independent variables either correlating with each other or having profile similarities which then "pull" the verb(s) as qualitative supplementary variables near to them, both visualization methods are worth taking into consideration.

The *CA* graph demonstrates strong correlations between {zugedeckt} of the *Manner* parameter as well as {gar} and {weich} of the *Resultative* (adjectives) parameter and the German culinary verbs *schmoren*, *kochen*, and *dünsten* (examples 24-25).

24. Bei schwacher Hitze, *zugedeckt*, ca. 2 1⁄2 Stunden den Braten *schmoren*, bis er *weich* ist. *(detenten13)*
    [Covered, braise*/simmer* the roast meat over (very) low heat for about 2,5 hours until its soft*/tender*].

25. *Einen Deckel auf den Topf legen*[66] und den Grünkohl auf mittlerer Stufe in ca. 90 Minuten gar schmoren. *(detenten13)*
[Put the lid on the pot and braise* the kale over medium heat for about 90 minutes until its done*/cooked*].

The German culinary verb *dünsten* shows a strong correlation with the *Resultative* (adjectives) parameter values {bissfest} (al dente) and {glasig} (translucent),[67] as well as the *Manner* parameter value {rühren} (to mix), as illustrated in the *CA* graph (Fig. 5a). The verb *braten* correlates strongly with the *Resultative* (adjectives) parameter values {knusprig} (crispy) and {braun} (brown), as well as with the *Manner* values {wenden, rundherum, Seite} (turn over, all around, on … side(s)).
Fig. 5a also illustrates a strong correlation between the verb *rösten* and {kurz, leicht} (shortly, lightly) as well as {goldgelb} (golden) of the *Manner* and *Resultative* (adjectives) parameter values respectively.

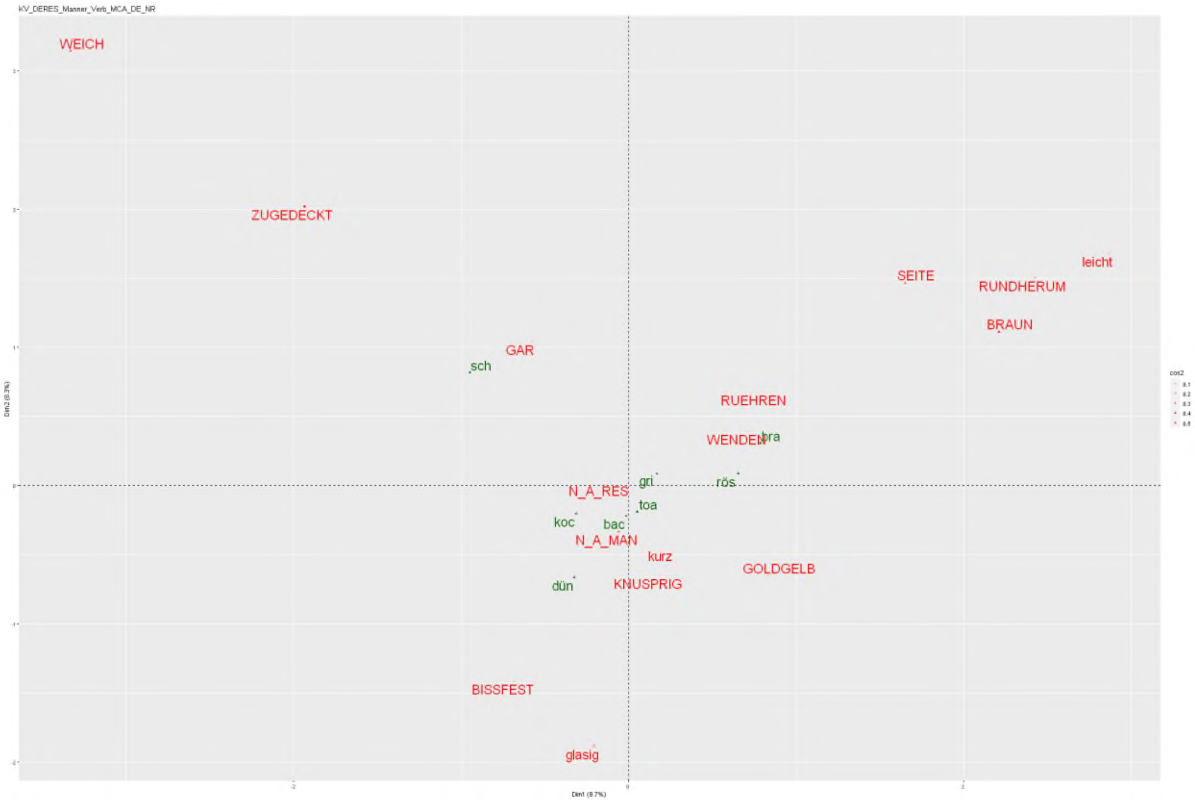


Fig. 5b. MCA: German culinary verbs (parameters *Manner* and *Resultative* (adjectives))

[66] The *Manner* parameter value *{zugedeckt}* is generalized and might include such values as {einen Deckel legen}, {mit geschlossenem Deckel}, {mit einem Deckel abdecken} and the like. For the list of all generalizations see Appendix 2.
[67] {Glasig} is positioned far at the bottom of the graph, forming almost a line with the verb *dünsten*. This suggests that the angle between the two is so sharp that it is difficult to distinguish, therefore it appears like a line.

The *MCA* on Fig. 5b sets {weich} (soft) as an outlier by placing it in the upper-left corner and overestimating it as a relevant parameter by positioning it in a linear cluster with {gar} (ready, done) and {zugedeckt} (covered, the lid on), showing their co-occurrence in the same contexts. In addition, this cluster includes the verb *schmoren* as a qualitative supplementary variable (see examples 24-25 above). Due to profile similarities, {rühren} (to mix) forms a small cluster with another *Manner* parameter value {wenden} (to turn) on MCA graph (Fig. 5b), as well as with the *Resultative* (adjectives) parameter value {goldgelb} (golden), correlating with these two values (example 26).

26. Die Butter in einem Topf erhitzen und das Mehl darin *unter Rühren goldgelb rösten*. (deTenTen13)
    [Heat the butter in a pan and roast*/toast* the flour in it, stirring until golden brown.]

Due to similar syntagmatic distribution and co-occurrence in the same contexts, another cluster of independent variables – the *Manner* parameter values {leicht, rundherum, Seite} (lightly, all over, on…side) and{braun}(brown) of the *Resultative* (adjectives) – is visualized on MCA graph (Fig.5b) (example 27).

27. In einer Pfanne das Butterschmalz erhitzen und die Brotwürfel darin *rundherum goldbraun* rösten. (deTenTen13)
    [Heat the clarified butter in a pan and toast*/roast* the bread cubes all over*/all around* until golden brown all over].

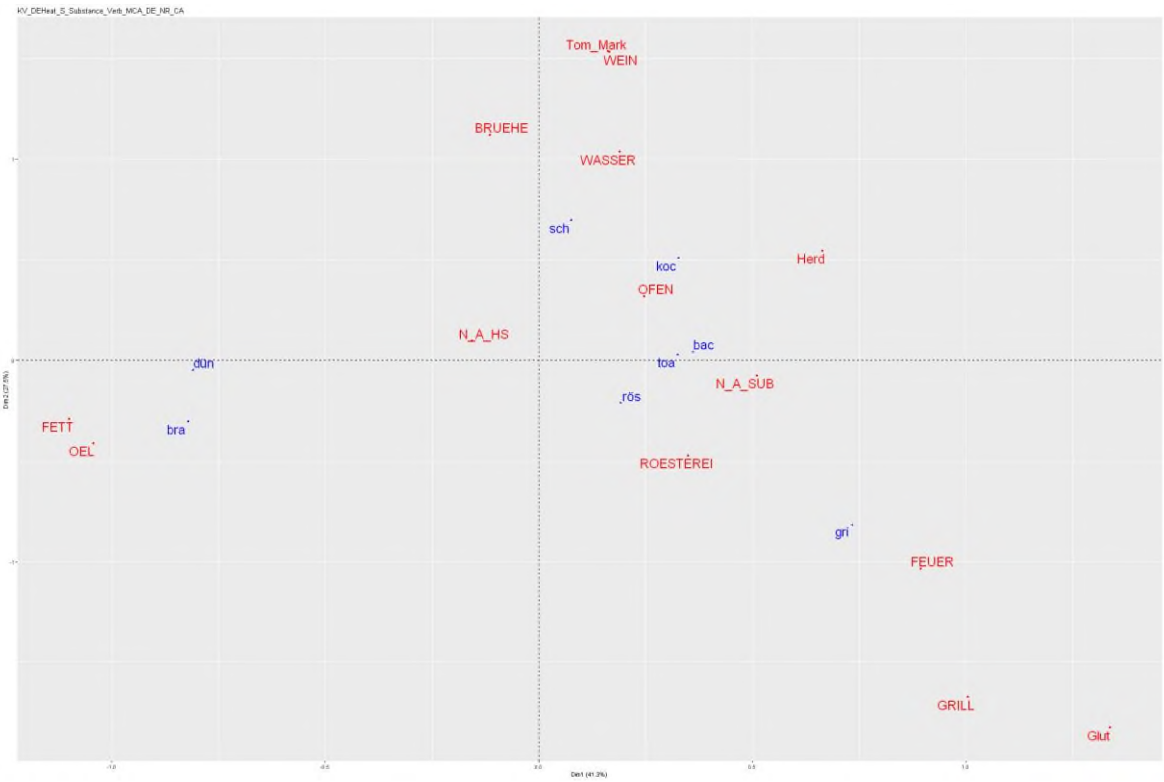


Fig 6a. CA: German culinary verbs (parameters *Heat source* and *Substance*)

The following CA graph (Fig. 6a) illustrates several strong correlations between independent and dependent variables. First, the graph is "divided" into three parts. In the upper part, strong correlations are shown between the verb *schmoren* and the *Substance* parameter values {Brühe}, {Wasser}, {Tomatenmark}, and {Wein} (all *liquid*), as well as the *Heat source* value {Ofen} (oven). The verb *kochen* is "pulled" by both the *Substance* parameter values {Wasser} (water) as well as the *Heat source* values {Herd} (stove) and {Ofen} (oven). Second, *grillen*, *rösten*, and *toasten*, and partially also *backen* are grouped on the lower left-hand side of the CA graph mostly due to {N_A_SUB} (non-verbalized *Substance* parameter values). The verb *rösten* correlates strongly with {Rösterei} (roasting plant) (example 28), while *grillen* with the *Heat source* parameter values {Grill} (grill), {Feuer} (fire), and {Glut} (embers) (examples 29).

28. Herkömmliche Röstereien rösten den Rohkaffee nur 3-4 Minuten in kontinuierlichen Verfahren. *(deTenTen13)*
    [Conventional roasting plants only roast the green coffee for 3-4 minutes in a continuous process.]

29. Nun das Fleischstück auf beiden Seiten über heißer Glut eineinhalb Minuten grillen. *(REZ_DE)*

[Now grill the piece of meat on both sides over hot embers for one and a half minutes.]

Third, the verb *dünsten* correlates with the *Substance* parameter value {Fett} (fat), while *braten* correlates with {Öl} (oil), thereby assigning the lower left-hand side of the CA (Fig. 6a) to the *oil*-like, *fat*-like cooking substances, even though *braten* and *dünsten* here have almost the same distribution.

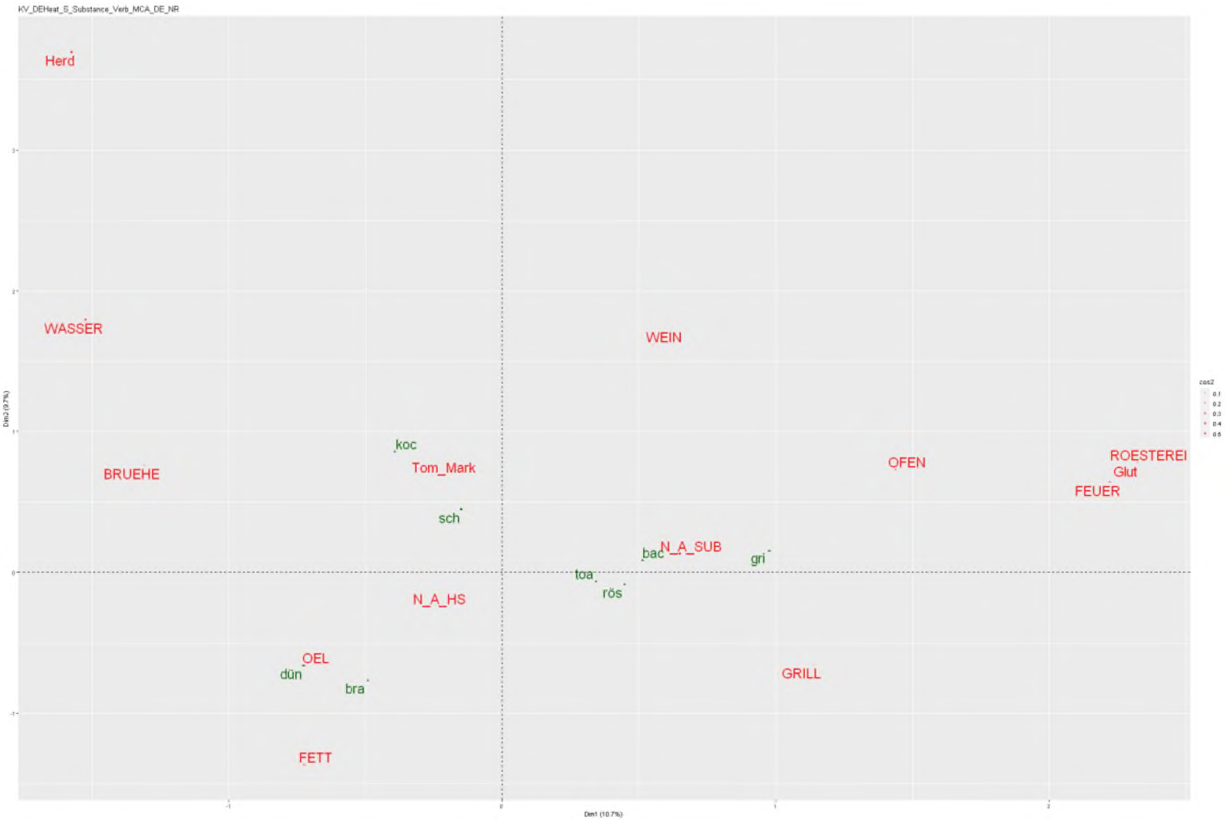


Fig. 6b. The MCA of the German culinary verbs (parameters *Heat source* and *Substance*)

The *MCA* graph (Fig. 6b) identifies a cluster on the upper right-hand side that includes the *Substance* parameter values {Wasser} (water) and {Brühe} (stock, fond) due to their profile similarities. This cluster also encompasses the *Heat source* parameter value {Herd} (stove), indicating its strong correlation with the aforementioned *Substance* parameter values (see Table 2). As a peculiarity of the *MCA* visualization method, {Herd} (stove) is overestimated here, thus generating an outlier (example 30).

30. Ich habe früher immer zuerst das Fleisch gekocht, in der gleichen *Brühe* dann den Rest auf dem *Herd* köcheln lassen. *(REZ_DE)*

[I used to always cook*/boil* the meat first, and then to simmer the remaining ingredients in the same stock on the stove.]

Due to its correlation with the *Heat source* parameter values {Ofen} (oven), as well as profile similarities with the {N_A_Sub} (no verbalized values) of the parameter *Substance*, the *Substance* parameter value {Wein} (wine) creates another cluster illustrated on this *MCA* graph (Fig. 6b). However, as shown in Table 2 below, another small cluster could be identified here, viz. the *Heat source* parameter values {Ofen} and {Grill} (grill), which share a similar syntagmatic distribution and correlate with the *Substance* parameter value {N_A_SUB}. Due to their profile similarities, the values {Öl} (oil) and {Fett} (fat) of the *Substance* are grouped together on the *MCA* graph (Fig. 6b), with the verbs *braten* and *dünsten* included here as qualitative supplementary variables (examples 31-32).[68]

31. In einer großen Pfanne das *Fett* erhitzen, die Kartoffeln kross *braten*. *(deTenTen13)*
    [Heat the fat in a large pan, fry the potatoes until crispy].

32. In einem Topf Kokosöl erhitzen, die Zwiebeln glasig dünsten. *(REZ_DE)*
    [Heat the coconut oil in a pan and fry the onions until translucent.]

	FEUER	Glut	GRILL	Herd	N_A_HS	OFEN	ROESTEREI
BRUEHE	-1.8908876	-0.5402536	-1.4036201	2.5300980	0.8013327	-1.6357697	-0.5730255
FETT	-2.8686532	-0.9093823	-0.6696265	-2.5111140	2.3130632	-2.7064789	-0.9645456
N_A_SUB	5.2641871	1.5801096	1.7743009	-1.7100556	-2.4938066	3.0689748	1.6759593
OEL	-4.0454026	-1.2225991	-0.3430142	-0.7101447	2.1072813	-2.2609982	-1.2967621
Tom_Mark	-1.0145432	-0.2898695	-0.7531030	0.4489028	-0.0774381	1.5182550	-0.3074530
WASSER	-2.7539609	-0.7868460	-2.0442858	6.5719311	0.5877861	-2.4218451	-0.8345762
WEIN	-0.9596618	-0.2741891	-0.7123642	0.5636489	-0.7275376	3.5989136	-0.2908214

Table 2. Contingency table of the independent variables (parameters *Heat source* and *Substance*)

[68] The cluster may also include the *Heat source* parameter value {N_A_HS} (no verbalized values) as it correlates with the {Öl} and {Fett} values of the *Substance* parameter, as traced in Table 2.

The *Heat source* parameter values {Rösterei} (roasting plant), {Feuer} (fire), and {Glut} (embers) are grouped together on the *MCA* graph (Fig. 6b) again due to their profile similarities. These values could also have been grouped with {Ofen} (oven); however, due to the overestimation of highly relevant parameter values, regardless of the number of annotated examples in the *MCA*, {Rösterei} (roasting plant) and {Glut} (embers) have generated outliers, while {Ofen} (oven) is also "pulled" by the {N_A_SUB} (no verbalized values) of the *Substance* parameter (see Table 2 above).

2.4 *Context-Conditional Correlation Graph (*CCCG) Visualization of the German Data Analysis

The *Context-Conditional Correlation Graph* (CCCG) contributed to the visualization of the contrasting the German culinary verbs *backen, braten, dünsten, grillen, kochen, rösten, schmoren* and *toasten,* in particular by highlighting the most important respective differences. This type of data visualization allowed to illustrate simultaneously all strong correlations between the verb and parameter values contrasting it to the other verbs at the same time and within a single graph. Each verb is represented in a specific color, together with its strongly correlated parameter values; for instance, the verb *kochen* and {wine} are marked in green (Fig.1). The comparison is always carried out in verb pairs, where the verb, used as reference for comparison, is displayed at the center of the graph, while the other verbs surround it and are linked to it with the edges of the graph. The difference between the two verbs being compared is determined by the number of edges connecting them, as well as the deviation value near each parameter value: more edges correspond to differences between the verbs across a greater number of parameters while high deviation values indicate stronger differences. Consequently, fewer edges mean that the two verbs being compared are semantically closer. The numerals near each parameter value show the deviation of the observed number of cases with a certain parameter value for a specific verb from the expected number with regard to the same parameter value of (an)other verb(s). It is important to note that the deviation values determine the position of the edge label. We consider as statistically significant deviation values higher than twice the standard deviation (calculated as the square root of the expected value for a given parameter). Thus, this form of visualization is appropriate when we are interested in the semantics of culinary verbs, by demonstrating how strongly they differ from one another and what semantic similarities, across which parameters they show the most important distributional differences.
For instance, if we consider the independent uniform distribution of *Manner* parameter value[69] {kurz} (shortly) (Fig. 1) when comparing two German culinary verbs such as *toasten* and *backen*, each verb would be expected to have the same proportion of examples. In general, deviation values between the expected and the observed distribution within the standard deviation are typically seen as statistically non-significant. For the sake of clarity and legibility of the graph representation, in this particular case, the distribution of the parameter *Manner* with respect to German culinary verbs is represented only for the cases with a deviation over

[69] For a more detailed information on annotation parameters see Chapter 2.1.

2.5 times the standard deviation. For the same reason, not all verbs are always compared using the context-conditional correlation graphical visualization. The *CCCGs* were generated using the *qgraph* package in the programming language *R* for statistical computation and exported along with accompanying excel-tables for each graph illustrating the differences more clearly in cases where the graph is too complicated.

In our German sub-corpus, there are 266 annotated examples with *dünsten* and 86 with *toasten.*[70] The expected value for the independent distribution of the *Substance* parameter value {Öl} for the verbs *dünsten* and *toasten* is ≈86 and ≈28 respectively[71], provided there are overall 114 annotated examples with {Öl} for the two verbs together. The observed distribution, however, in our German annotated sub-corpora (*deTenTen13* and *REZ_de* together) differed significantly from the expected distribution: after neutralization of parameter values, 113 co-occurrences with the *Substance* parameter value {Öl} were with the verb *dünsten,* while there was only 1 example with *toasten*. Thus, the difference in distribution between the observed and the expected value is 26,8522727[72] for the verb *dünsten* and -26,8522727 for *toasten.*[73] The observed deviation value (in this case 2.9[74] near {Öl}) shows the difference of the observed and expected distribution values in terms of the standard deviation. The graph plot (Fig. 1) illustrates that for the verbs *dünsten* and *toasten,* the number of co-occurrences of the *Substance* parameter values {Öl} with *dünsten* is 2.9 times the standard deviation, indicating that the distribution is not random but rather statistically significantly affected by the aforementioned parameter.

[70] For the list of all annotated examples see Appendix 1.
[71] 114*266/(266+86)= 86,1477273; 114*86/(266+86)= 27,8522727
[72] 113-86,1477273= 26,8522727
[73] 1-27,8522727= -26,8522727
[74] 26,8522727/√86,1477273=2,8930713≈2.9

2.4.1 CCCG of the German Culinary Verb *toasten*

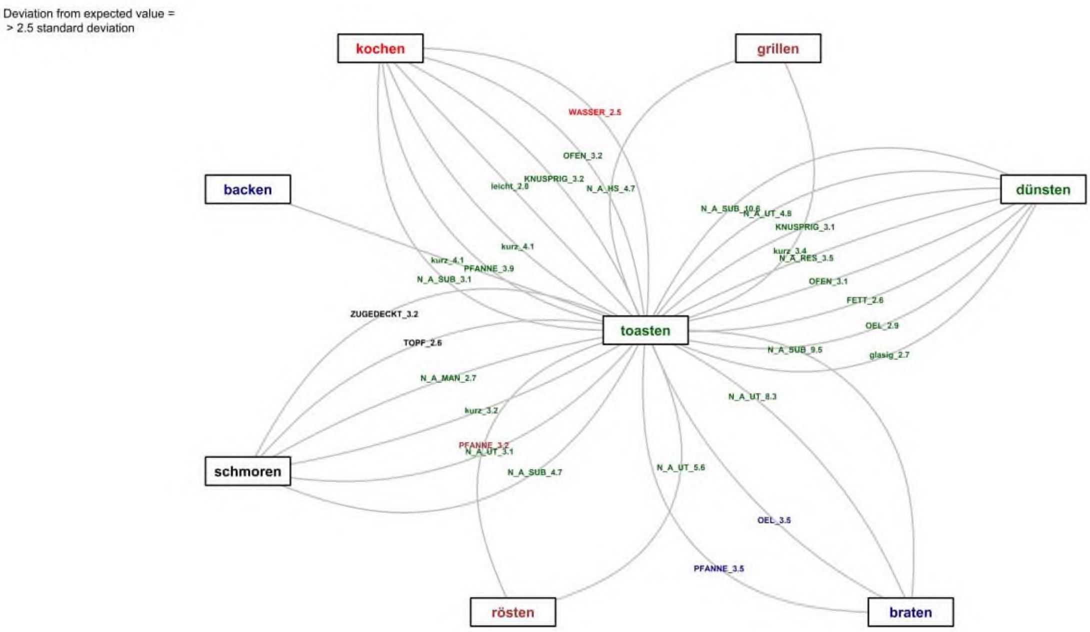


Fig. 1. CCCG analysis of the German culinary verb *toasten*

This *CCCG* (Fig. 1) illustrates the comparison of the German culinary verbs in pairs, with the verb *toasten* being compared to the other seven verbs, i.e. *braten*, *backen*, *dünsten*, *grillen*, *rösten*, and *schmoren*. Fewer edges connecting two verbs indicate greater semantic similarity between them, so the verbs *backen*, *grillen* as well as *rösten* are singled out as the verbs most semantically similar to *toasten* within their respective pairs (Fig. 1). If we consider *toasten vs. backen*, the difference between the two verbs lies only in the parameter *Manner* with the value {kurz} (shortly) (marked in green as the verb itself) overrepresented for the verb *toasten* by 4.1 time the standard deviation, based on the hypothesis of an independent uniform distribution. However, due to insufficient data in form of *non-verbalized parameter values* annotated as N_As for almost all other parameters, *backen* is identified as the semantically closest verb to *toasten*. Nevertheless, the parameter *Manner* has a statistically significant impact on the distribution of the occurrences for these two verbs, semantically differentiating them from one another. In the pair *toasten* vs. *grillen,* in addition to the *Manner* parameter value {kurz}, which is identified as overrepresented for the verb *toasten*, the latter correlates strongly with the *Heat source* parameter value {N_A_HS}, which display no verbalized values (example 1).

1. Den Toast *kurz toasten* und mit Senf bestreichen. *(detenten13)*
   [Briefly/shortly* toast the toast* and coat with masters].

In the distribution of the *Utensil* parameter values when comparing the verbs *toasten* and *rösten*, the co-occurrences of the {Pfanne} with *rösten* exceed the expected value by 3.2 times the standard deviation (example 2), while *toasten* correlates strongly with non-verbalized *Utensil* parameter value. However, also in this pair the semantic closeness of the verbs is partially due to the insufficiency of data, e.g. regarding the parameters *Substance*, *Heat intensity*, and *Resultative* (*adjectives*).

2. In einer Pfanne Kürbiskerne ohne Fett goldbraun *rösten*, dann herausnehmen. *(detenten13)*
   [In a pan roast* the pumpkin seeds without fat until golden brown then take out].

In the verb pair *toasten* vs. *braten*, the parameters *Substance* and *Utensil* significantly affected the distribution of the occurrences. For instance, in the distribution of the *Substance* parameter values when comparing the verbs *toasten* and *braten*, the number of occurrences of {Öl} with *braten* exceeds the expected value by 3.5 times the standard deviation (example 3). As mentioned above, the verb *toasten*, on the contrary, is identified as strongly correlated with no verbalized values in the *Substance* and *Utensil* parameter.

3. In einer Pfanne etwas Öl erhitzen, darin 4 Spiegeleier braten. *(REZ_DE)*
   [In a pan heat some oil and fry* 4 eggs sunny side up].

In the verb pairs *toasten* vs. *kochen* as well as *toasten* vs. *dünsten*, the semantic difference is observed in almost all of the six annotation parameters, except for *Heat intensity*. For instance, the co-occurrences of the *Substance* parameter value {Wasser} with *kochen* are overrepresented by 2.5 times the standard deviation, while *toasten* shows a strong correlation with *non-verbalized substance value* annotated as {N_A_SUB}. At this point, it is important to emphasize the significance of the *Substance* parameter in differentiating the culinary verbs across the languages analyzed in this work. For this purpose, a *Mosaic-plot* graphical visualization method was generated to illustrate the distribution of the German verbs with regard to the parameter *Substance* (Fig. 2).

2.4.2 *Mosaic-Plot* Visualization of the German Data Analysis

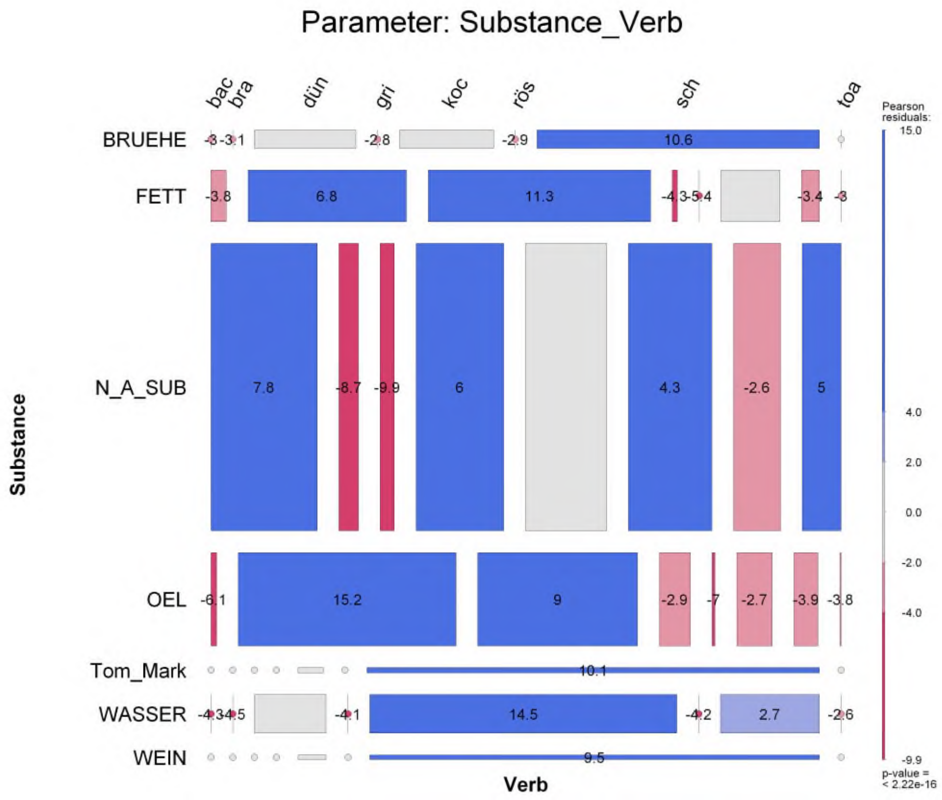


Fig. 2. Distribution of the German culinary verbs with regard to the parameter *Substance*

Before we go into the details of each verb and its correlations with separate parameter values, it is worth mentioning how the *Mosaic-Plot*[75] actually works. It was first introduced back in 1981 by Hartigan and Kleiner a method for plotting contingency table data. Moreover, it was successfully extended to multidimensional table data with more than two variables. Recalling Hartigan and Kleiner, Friendly explain that if "row and column variables are independent, then expected frequencies are products of the row and column totals (divided by the grand total)" (1994:190). If there was no correlation, then the distribution for all parameters would be the same. On the plot, the height of the vertical columns corresponds to the frequency of a given

[75] Behind the *Mosaic-Plot* visualization graph is the regression analysis of statistical methods for categorical data. "Regression analysis models the relationships between a response variable and one or more predictor variables (see Frost 2019,) https://statisticsbyjim.com/glossary/residuals/ . "Predictor variables", also called independent values, are in our case the parameters of annotation, e.g. *Substance*, *Manner*, *Heat intensity*, etc. "Response variables" are the culinary verbs under consideration. The visualization of *Mosaic-Plots* is based on Pearson residuals. "A residual is the difference between the observed value and the mean value that the model predicts for that observation" (ibid.).

parameter in the corpus.[76] In other words, for instance, the parameter value of {N_A_SUB} is the most frequent *Substance* parameter value annotated in the corpus. The rows, on the other hand, illustrate the distribution of the annotated occurrences with each parameter value (in this case, *Substance*) in relation to the eight German culinary verbs. The width of the columns displays the marginal frequency of a given parameter value in relation to the verbs. Besides the rectangle columns (or "tiles"), the mosaic method for visualizing categorical data analysis also uses blue, red, and grey[77] "colors and shadings" (Friendly 1994:191). Overrepresentations (stronger correlations) of each verb and the single parameter values are illustrated in blue or dark blue, while gray shows combinations of parameters without a statistically significant over- or underrepresentation. The red or dark red color represents a complementary distribution of a verb and a parameter values, i.e. the underrepresented combinations. In other words, combinations marked in blue represent correlations, those in red indicate anti-correlations.

Going back to the distribution of the German culinary verbs in relation to the parameter *Substance*, the value {N_A_SUB} is overrepresented for the verb *rösten* as well as for *backen*, *grillen*, and *toasten*. However, due to the manual semantic annotation of the occurrences for the verb *rösten*, the *Substance* parameter value {N_A_SUB} might either denote the actual absence of any cooking substance or just a missing verbalization. The non-verbalization of the cooking substance, mainly *fat* or *oil*, might also mean that its absence is assumed in general or merely implied for the given occurrences; however, in other examples the co-occurrence of *rösten*, {Fett} or {Öl} might be verbalized. Throughout this research, all instances of non-verbalized *Substance* parameter values, as well as {ohne Fett} (without fat) and {ohne Öl} (without oil) co-occurring with the verb *rösten*, were annotated as {N_A_SUB}. Therefore, as this annotation drawback concerns only the German verb *rösten*, a small qualitative analysis came to cast light on all instances of {ohne Fett} (without fat) and {ohne Öl} (without oil), initially annotated as {N_A_SUB}, as opposed to the rest of the non-verbalized *Substance* parameter values also annotated as {N_A_SUB}. The following qualitative analysis was based on a total of 268 annotated examples of *rösten* in the German sub-corpus, with 61 ones co-occurring with the *Substance* parameter value {ohne Fett, Öl} (without any fat or oil) and 207 ones co-occurring with {N_A_SUB}, denoting all occurrences where no *Substance* parameter value is

[76] Friendly calls them "rectangles" while Hartigan and Kleiner "tiles" (cf. Friendly 1994; Hartigan and Kleiner 1981).

[77] Earlier versions of the mosaic display used "colors and shading" by marking the positive deviation (overrepresentation) from the independent uniform distribution with solid-lined rectangles and black close-to-each-other linear shades, while the negative deviation was marked with dashed-outlined rectangles with relatively far-from-each-other lines (Friendly 1994:191).

verbalized.[78] As a result only the parameters *Ingredient* (Fig. 2.1) (initially excluded from the analysis due to the large number of different values, making generalization impossible), *Utensil* (Fig. 2.3), and *Manner* (Fig. 2.4) proved to have a statistically high impact on the distribution of the annotated examples co-occurring with {ohne Fett, Öl} (without any fat or oil) as opposed to those with {Rest}.[79] The parameters *Heat intensity*, *Resultative (adjectives)*, and *Heat source* showed no statistically significant over- or underrepresentations. The verbs *dünsten* and *braten*, on the contrary, typically presuppose a type of substance (Fig. 2). For instance, the latter two are strongly collocated with {Fett} (example 4), be it *Butter* (butter), *Butterschmalz* (ghee), *Margarine* (margarine) or any other fat, as well as {Öl} (examples 5–6), e.g. *Rapsöl* (canola oil), *Sonnenblumenöl* (sunflower oil), *Olivenöl* (olive oil), etc.

4. Für die Pilze die *Butter* schmelzen, die Schalotte mit dem Thymian darin glasig *dünsten*. *(detenten13)*
   [For the mushrooms melt the butter and sauté* the shallots with the thyme until translucent].

5. In einer Pfanne *Butterschmalz* erhitzen, die Steaks darin drei Minuten von jeder Seite *braten*. *(REZ_DE)*
   [Heat the ghee in a pan and fry* the steak for 3 minutes on each side].

6. Etwas *Margarine* in der Pfanne erhitzen, Spiegeleier *braten* und leicht mit Salz würzen. *(detenten13)*
   [Heat some margarine in a pan, fry* the eggs sunny side up and lightly season with salt].

[78] For all annotated examples see Appendix 1.
[79] {Rest} is not an initially annotated *Substance* parameter value; however, it has been represented as such only for the Mosaic-Plot graphs for better illustration purposes, denoting the rest of the {N_A_SUB} values.

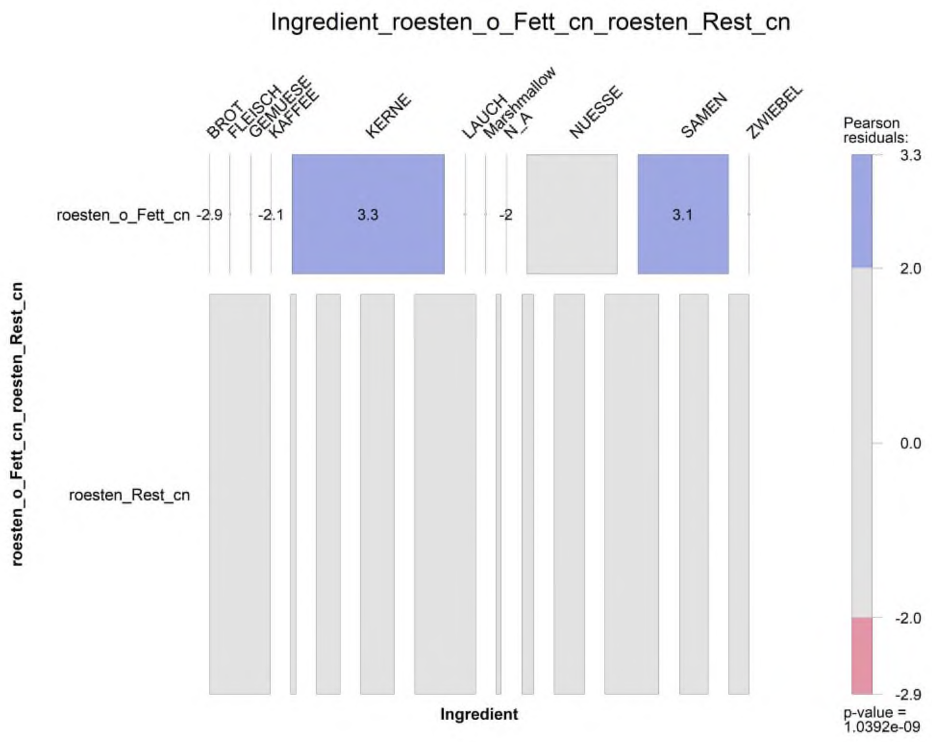


Fig. 2.1. Mosaic-Plot of {ohne Fett_Öl} vs. {Rest}) co-occurring *rösten*

This *Mosaic-Plot* displays the distribution of the separate values of the *Ingredient* parameter co-occurring with the German verb *rösten* and the *Substance* parameter value {ohne Fett_Öl}, as opposed to {Rest}. The *Ingredient* parameter values {Kerne} (nuts) and {Samen} (seeds) are overrepresented for the verb *rösten* in co-occurrences with {ohne Fett_Öl}, while {Brot} (bread) and {Kaffee} (coffee) (generalized from different sorts of coffee beans, such as *Espressobohnen*, *Kaffeebohnen*) are underrepresented, i.e. anti-correlated, which shows that *roasting coffee beans without any fat or oil* is already implied by encyclopedic knowledge.[80] Thus, all the other occurrences annotated as {N_A_SUB}, denoting a non-verbalized *Substance* parameter value (except for {ohne Fett_Öl}), might denote; a) some type of oil/fat may be used, b) the absence of any type of oil/fat is strongly implied, e.g. *Kaffeebohnen rösten (roast the coffee beans)* (example 7).

7. Sie *rösten* die *Kaffeebohnen* in 1-2 Min. *(detenten13)*

[80] Chapter 5.2.5 of this work illustrates in *Context-conditional correlation graphs* (CCCG) (Fig. 10) the translation possibilities between the German and English verbs, with the verb *rösten* compared in pairs to the six English culinary verbs in this work, i.e. *bake*, *braise*, *cook*, *fry*, *sauté*, and *roast*. Different dictionary definitions of the verb *rösten* are provided as a footnote, emphasizing the compliance of the corpus evidence with certain senses of the verb.

[They (will) roast* the coffee beans in 1-2 minutes].

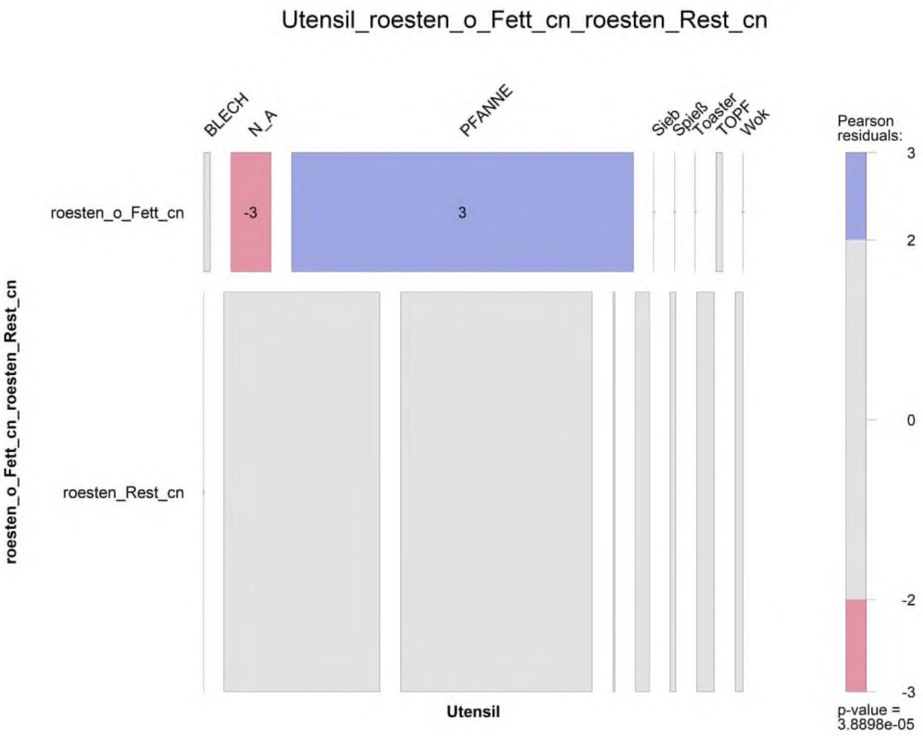


Fig. 2.2. Mosaic-Plot of {ohne Fett_Öl} vs. {Rest}) co-occurring *rösten*

As mentioned above, also the parameter *Utensil* impacted the distribution of the co-occurrences of *rösten* and the *Substance* parameter value {ohne Fett_Öl} (without any fat or oil) (Fig. 2.2). The *Utensil* parameter value {Pfanne} (pan) is overrepresented for the verb *rösten* co-occurring with {ohne Fett_Öl} (without any fat or oil) (marked blue), which denotes that a considerable amount of annotated examples with the verb *rösten* in co-occurrence with {ohne Fett_Öl} also co-occur with {Pfanne} (pan).[81] The *Utensil* parameter {N_A}, denoting non-verbalized values in this parameter, on the contrary, is underrepresented for the verb *rösten* co-occurring with {ohne Fett_Öl} (without any fat or oil) (marked red), suggesting the presence of any type of utensil in the annotated occurrences of *rösten* with {ohne Fett_Öl} (without any fat or oil). This Mosaic-Plot graph further illustrates; a) on one hand, the *Utensil* parameter values {N_A} and {Pfanne} (pan) as the most frequent ones in this sub-corpus of the co-occurrences of *rösten* with {ohne Fett_Öl} and {Rest}, b) on the other hand, the distribution of the aforementioned

[81] For the list of all parameter generalization (neutralization) of the German data see Appendix 2.

two *Utensil* parameters in relation to *rösten* co-occurring with {ohne Fett_Öl} and {Rest}. Thus, the number of occurrences of {Pfanne} (pan) and the verb *rösten* co-occurring with {ohne Fett_Öl} is considerably higher than those with {N_A}.

8. In einer *Pfanne* Kürbiskerne *ohne Fett* goldbraun *rösten*, dann herausnehmen. *(detenten13)*
   [Roast* the pumpkin seeds in a pan without (any) fat until golden brown, then take out].

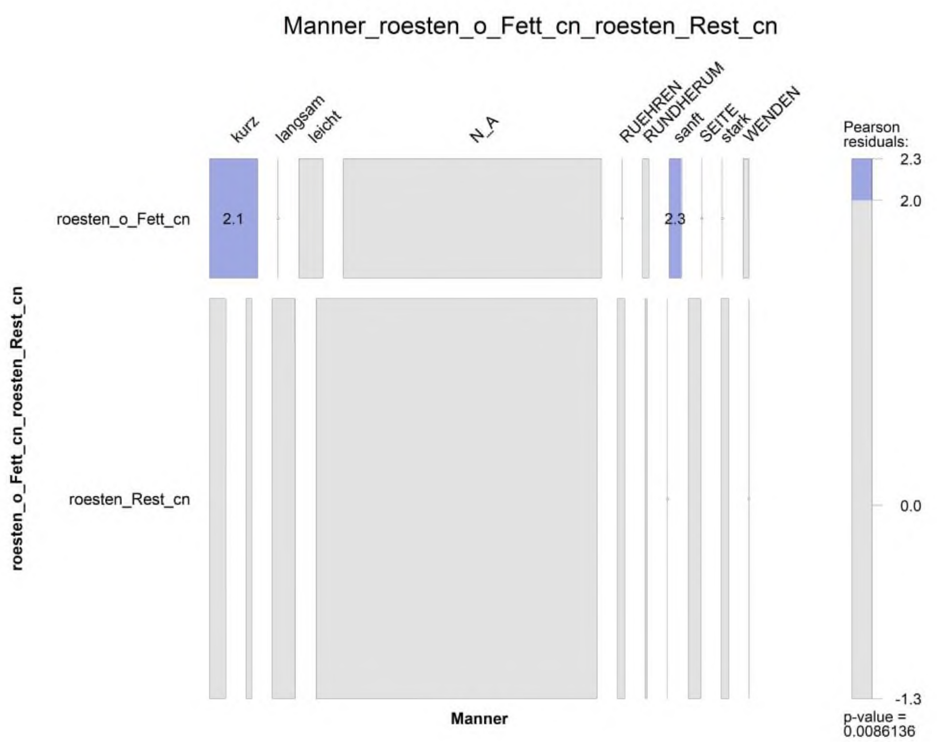


Fig. 2.3. Mosaic-Plot of {ohne Fett_Öl} vs. {Rest}) co-occurring *rösten*

The distribution of the co-occurrences of the verb *rösten* and the *Substance* parameter value {ohne Fett, Öl} (without fat or oil) in opposition to cases with {Rest} in the German sub-corpus is also impacted by the *Manner* parameter. Its values {kurz} (briefly, shortly) and {sanft} (gently) (example 9) are overrepresented for the verb *rösten* co-occurring with {ohne Fett, Öl} (without fat or oil) (Fig. 2.3).

9. Am einfachsten geht das, wenn man die Kastanien *kurz* in einer Pfanne ohne Fett röstet. *(detenten13)*

[The easiest way is to shortly roast* the chestnuts in a pan without any fat].

The grey columns in all three aforementioned *Mosaic-Plot* graphs (Fig. 2.1, 2.2 and 2.3) represent parameter values without any over- or underrepresentation in relation to the co-occurrences of the German culinary verb *rösten* and the *Substance* parameter value {ohne Fett, Öl} (without fat or oil). Those associated with *rösten* and {Rest} (the remaining examples annotated as {N_A_SUB} show that there is no statistically significant impact of the distribution of occurrences in this German sub-corpus by any of the *Ingredient*, *Manner*, and *Utensil* parameter values except the ones elaborated upon in detail.

2.4.3 CCCG of the German Culinary Verb *dünsten*

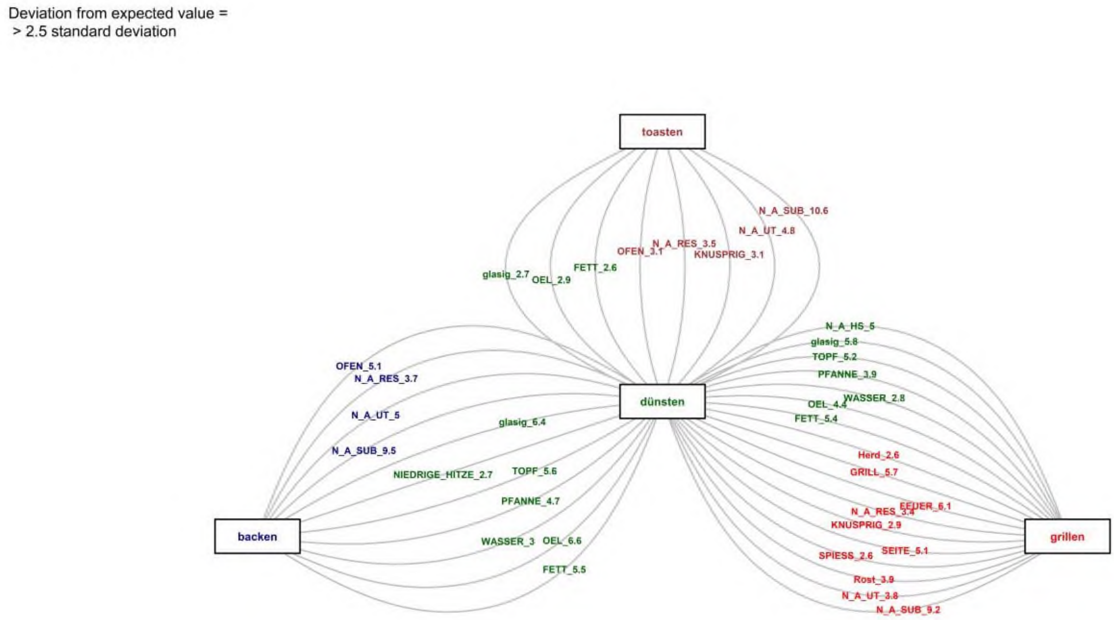


Fig. 3. *CCCG* analysis of the German culinary verb *dünsten*

The interpretation of this *CCCG* graph might seem somewhat problematic without carefully it zooming out, as considerably more correlations are displayed than in Fig.1, which illustrates the comparison of the verb *toasten* to the rest of the German culinary verbs. However, the graphs generated by *R* and are also accompanied by tables in *.xlsx* format containing all necessary information to observe the overrepresentations[83] (e.g. partial Table 3.1 below). In order to focus on the most important information, not all correlations are commented on in detail. In particular, the parameter values generating the most significant divergences between verbs are discussed here. For instance, in the distribution of the *Resultative (adjectives)* parameter values for *backen* and *dünsten*, the number of occurrences of {glasig} (translucent) with *dünsten* exceeds the expected value by 6.4 times the standard deviation under the hypothesis of an independent uniform distribution.

In other words, the graphical plot simultaneously displays all the blue rectangular fields in the *Mosaic-Plot* graphical visualization method described above (see Fig. 2), without the grey and the red ones.

[83] See full tables of the Context-Conditional Correlation Graph display in Appendix 3.

The *CCCG* (Fig. 3) of *dünsten* – accompanied by the list of correlations[84] as partial tables inserted throughout the analysis – as a reference verb in comparing to the rest of the German culinary verbs, illustrates the distinguishing parameters with the respective values between the verb pairs being compared. Since the number of edges denote the difference between the verb pairs, *dünsten* differs from the other German verbs in almost all parameters. Separate verb pairs are elaborated in detail below; however, in all pairs, the most differentiating parameter is *Substance*. For instance, in the verb pairs *dünsten-rösten*,[85] *dünsten-toasten*, and *dünsten-backen*, the verb *dünsten* differs from the ones it is compared to by correlating strongly with either the *Substance* parameter values denoting some type of *liquid* cooking substance, i.e. {Wasser} (water) or some *oil*, i.e. {Öl, Fett} (oil or fat), or both, as opposed to *no verbalization of any substance* annotated as {N_A_SUB}. As observed in the partial table below, the verb *dünsten* semantically differs from *backen* in all of our annotation parameters, except for *Manner*, as the overwhelming majority of the annotated examples for *dünsten* and *backen* have no verbalized values in the *Manner* parameter.

98	backen	duensten	OFEN_5.1	0.378598025661458
100	backen	duensten	N_A_RES_3.7	0.408383785026579
109	backen	duensten	N_A_UT_5	0.380408019417635
113	**backen**	duensten	**N_A_SUB_9.5**	0.34195637876834
101	**duensten**	backen	**glasig_6.4**	0.362669087767066
105	duensten	backen	NIEDRIGE_HITZE_2.7	0.447364487857646
106	duensten	backen	MITTLERE_HITZE_2.4	0.46812465420032
107	duensten	backen	TOPF_5.6	0.371182961542389
108	duensten	backen	PFANNE_4.7	0.38585757383173
111	**duensten**	backen	**WASSER_3**	0.432589048026184
112	**duensten**	backen	**OEL_6.6**	0.360515577999244

Table (partial) 3.1. Actual distribution of the annotated occurrences in *dünsten-backen* verb pair

[84] For all *CCCGs* in *.xlsx* format see Appendix 3.1.

[85] The peculiarities of the parameter *Substance*, especially its non-verbalized values initially all annotated as {N_A_SUB} as opposed to {ohne Fett} (without oil or fat), co-occurring with the verb *rösten*, were addressed and elaborated upon in Chapter 2.4.2.

In comparing the verb *dünsten* to *backen*, the co-occurrences of the *Substance* parameter values {Öl}[86] (oil) and {Fett} (fat) with *dünsten* surpass the expected value by 6.6 and 5.5 times the standard deviation, respectively, under the hypothesis of an independent uniform distribution, while *backen* correlates with {N_A_SUB} (cf. examples 10 and11).

10. In einem Topf *Kokosöl* erhitzen, die Zwiebeln glasig *dünsten*. *(REZ_DE)*
[Heat the cocoa oil in a pot and sauté* the onions until translucent].

11. Den Kuchen ungefähr 45 Minuten *backen*. *(REZ_DE)*
[Bake* the cake/pie about 45 minutes]. (Translation mine)

The semantic difference in the verb pair *dünsten-grillen* was observed in all annotation parameters, except for *Manner* and *Heat intensity*. However, here too, the similarity in the aforementioned two parameters was due to non-verbalized values, indicating that even if some values are verbalized, they did not give rise to statistically significant differences in the distribution of occurrences for the verbs under comparison with regard to this parameter. Besides the parameter *Substance*,[87] *Heat source* and *Utensil* were decisive in differentiating parameters in the comparison of *dünsten* and *grillen*. For instance, when comparing the verbs *grillen* and *dünsten*, the co-occurrences of the *Heat source* parameter value {Feuer} (fire) and {Grill} (grill) with *grillen* surpass the expected value by 6.1 and 5.7 times the standard deviation, respectively, under the hypothesis of an independent uniform distribution (examples 12 and 13) (Fig. 3, partial Table 3.2).

61	duensten	grillen	N_A_HS_5	0.380433199137938
68	duensten	grillen	glasig_5.8	0.368639002126439
71	duensten	grillen	TOPF_5.2	0.376335770191362
74	duensten	grillen	PFANNE_3.9	0.403638895275366
76	duensten	grillen	WASSER_2.8	0.445219601460266
77	duensten	grillen	OEL_4.4	0.390138155288058
79	duensten	grillen	FETT_5.4	0.374274839936645

[86] The *Substance* parameter {Öl} (oil) was generalized (neutralized) to include different types of oil, e.g. {Kokosöl, Olivenöl} (cocoa oil, olive oil). For the list of all generalizations/neutralizations of parameter values see Appendix 2.
[87] It is obvious that all parameters displaying a difference above a 2.0 standard deviation value are statistically significant for the distribution of the occurrences; however, it is important to emphasize that for some verbs certain parameters are more decisive than others in differentiating the verbs being compared.

60	grillen	duensten	OFEN_2.2	0.482776814002617
62	grillen	duensten	Herd_2.6	0.452922054053243
63	**grillen**	duensten	**GRILL_5.7**	0.370537683629998
65	**grillen**	duensten	**FEUER_6.1**	0.365807174290671
66	grillen	duensten	N_A_RES_3.4	0.417131502087489
67	grillen	duensten	KNUSPRIG_2.9	0.439598097574976
69	grillen	duensten	SEITE_5.1	0.378447358507511
72	grillen	duensten	SPIESS_2.6	0.452922054053243
73	grillen	duensten	Rost_3.9	0.403100027810293
75	grillen	duensten	N_A_UT_3.8	0.404560347514746
78	grillen	duensten	N_A_SUB_9.2	0.343419626067204

Table (partial) 3.2. Actual distribution of the annotated occurrences in *dünsten-grillen* verb pair

12. Beim *Lagerfeuer grillen* wir Würstchen – oh JA! *(detenten13)*
[We will grill* sausages by the bonfire- oh, yes!]

13. Lammhaxen *auf dem heißen Grill* rundum ca. 8 Minuten *grillen. (detenten13)*
[Grill the lamb shanks on the hot grill for about 8 minutes all around].

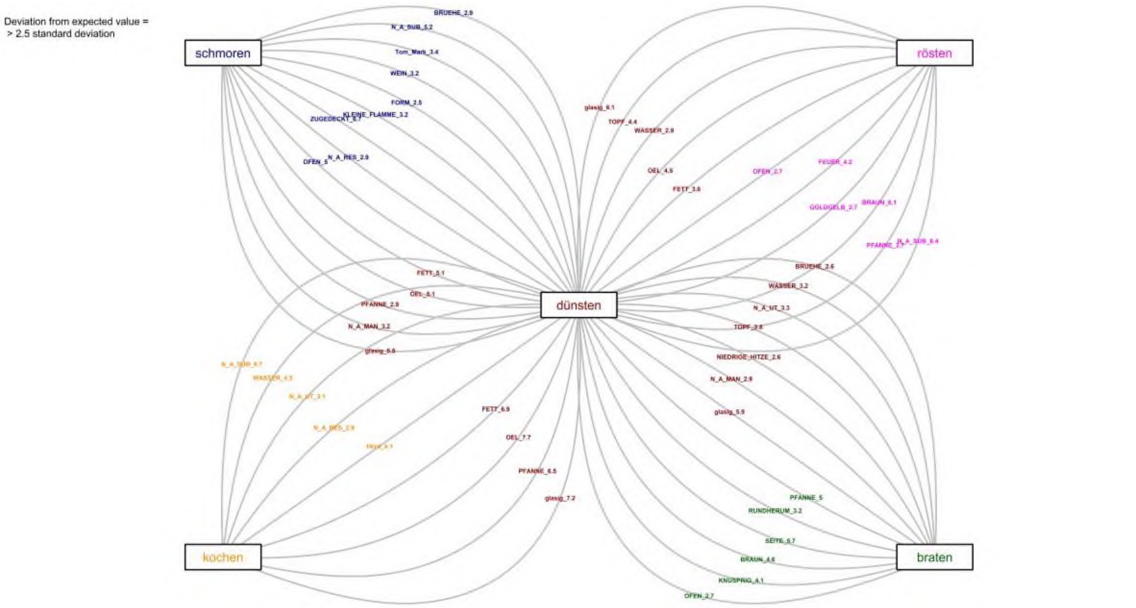


Fig. 3. *CCCG* analysis of the German culinary verb *dünsten*

All of the annotation parameters except for *Heat intensity*[88] impacted the distribution of the occurrences in the verb pair *dünsten-rösten*, illustrating their semantic difference (Fig.3, partial Table 3.3). The parameter *Substance* excludes any semantic closeness between these two verbs. While the verb *dünsten* correlates with *Substance* parameter values denoting both different kinds of *oil* and *fat* {Öl, Fett} as well as some *liquid* substances {Wasser} (water), the verb *rösten* creates a strong correlation with {N_A_SUB} denoting *no verbalization of any substance values*.[89] Nevertheless, the parameter *Resultative (adjectives)* also influenced the distribution of the occurrences for this verb pair, resulting in semantic difference here as well, where two different values of one and the same parameter determine the verb, namely, {glasig} (translucent) correlates with *dünsten* while {braun} (brown) with *rösten* (cf. examples 14 and15).

34	**duensten**	roesten	**glasig_6.1**	0.365155320446515
37	duensten	roesten	ZUGEDECKT_2.3	0.4752032410031

[88] Generally, the parameter *Heat intensity* had a few verbalized values that were verb-specific, in the most cases creating strong correlations with the German verb *schmoren* (e.g. {bei kleiner Hitze} *schmoren*) (**braise* in low heat)) or *dünsten* (e.g. {bei niedriger Hitze dünsten} (*steam/braise/cook in low heat) or *braten* (e.g. {bei hoher Hitze} *braten*) (**fry* in high heat)). See *Correspondence analysis* (CA) in Chapter 2.3 of this work (Fig.2, CA of German culinary verbs in relation to the parameters *Heat intensity* and *Substance*).

[89] Here too, see the *Mosaic-Plo*t visualization method on the distribution of the occurrences of the verb *rösten* for the differentiation between *non-verbalized values* vs. *no substance*, both annotated as {N_A_SUB} in the parameter *Substance* (see Chapter 2.4.2).

40	duensten	roesten	TOPF_4.4	0.390916491933447
43	duensten	roesten	WASSER_2.9	0.437849172848249
44	duensten	roesten	OEL_4.5	0.388252033534574
46	duensten	roesten	FETT_3.6	0.40969156073201
29	roesten	duensten	ROESTEREI_2.4	0.469301909434262
30	roesten	duensten	OFEN_2.7	0.446619754479283
31	roesten	duensten	FEUER_4.2	0.395985160468165
33	roesten	duensten	GOLDGELB_2.7	0.450424193587763
35	**roesten**	duensten	BRAUN_6.1	0.365941840656152
42	roesten	duensten	PFANNE_3.7	0.408083970451304
45	roesten	duensten	N_A_SUB_8.4	0.347763244452165

Table (partial) 3.3. Actual distribution of the annotated occurrences in *dünsten-rösten* verb pair

14. Die Zwiebel klein würfeln und in Öl *glasig dünsten*. *(REZ_DE)*
[Dice the onion and sauté* it in oil until translucent].

15. Rösten Sie die Mandelstifte in einer Pfanne, bis sie leicht hellbraun sind. *(detenten13)*
[Roast the almond slivers in a pan untill light braun].

In the verb pair *dünsten-braten*, the *Resultative (adjective)* parameter had statistically the most influence in determining the semantic difference between the verbs: {glasig} (translucent) and {weich} (soft) correlate strongly with *dünsten* while {knusprig} (crispy) and {braun} (brown) with *braten,* even though the verb *dünsten* differs from *rösten* in all six annotation parameters (cf. examples 16 and 17). Moreover, the parameter *Utensil* and *Manner* semantically differentiate *braten* from *dünsten* with different values correlating with either of the verbs. For instance, while the *Manner* parameter value {Seite} (side)[90] correlates strongly with *braten*, {zugedeckt} (closed/covered) creates a strong correlation with *dünsten* (cf. examples 18 and 19). It is worth noting, however, that *dünsten* differs from *braten* in the parameter *Heat intensity*, which in the annotation of most of the German verbs had no verbalized values. Thus, when comparing the verbs *dünsten* and *braten,* the co-occurrences of the *Heat intensity* value

[90] The *Manner* parameter value {Seite} has been neutralized (generalized) and might include such values as {auf jeder Seite}, {von beiden Seiten}, {beideseitig}, {rundherum} in the German corpus. For a full list of generalization see Appendix 2.

{niedrige_Hitze}[91] (low heat) with *dünsten* exceed the expected value by 2.6 times the standard deviation (example 20) (Fig. 3, partial Tab. 3.4).

81	braten	duensten	OFEN_2.7	0.44777121130882
83	**braten**	duensten	**KNUSPRIG_4.1**	0.397741469014478
85	**braten**	duensten	**BRAUN_4.6**	0.3861961767679
88	**braten**	duensten	**SEITE_5.7**	0.369660017053745
89	braten	duensten	RUNDHERUM_3.2	0.42363426557461
94	braten	duensten	PFANNE_5	0.380551745303778
82	**duensten**	braten	**WEICH_2.3**	0.474522089729501
84	**duensten**	braten	**glasig_5.9**	0.368171506151818
87	**duensten**	braten	**ZUGEDECKT_2.5**	0.460536973416639
90	duensten	braten	N_A_MAN_2.9	0.436792744998868
91	**duensten**	braten	**NIEDRIGE_HITZE_2.6**	0.453781994641004
93	duensten	braten	TOPF_3.8	0.404891997876753
95	duensten	braten	N_A_UT_3.3	0.420260055927432
96	duensten	braten	WASSER_3.2	0.426309815219991
97	duensten	braten	BRUEHE_2.6	0.454697298397103

Table (partial) 3.4. Actual distribution of the annotated occurrences in *dünsten-braten* verb pair

16. Öl in einer großen Pfanne oder Topf erhitzen und die Zwiebel darin 5 Minuten *weich dünsten*. *(detenten13)*

[Heat the oil in a large pan or pot and sauté the onion for 5 minutes until tender/soft*].

17. Öl in einer Pfanne erhitzen und die Speckwürfel darin *knusprig braten*. *(detenten13)*

[Heat the oil in a pan and fry the diced bacon until crispy]. (Translation mine)

18. Das Öl in einer Pfanne erhitzen und die Fischfilets *auf beiden Seiten braten*]. *(detenten13)*

[Heat the oil in a pan and fry the fish filets on both sides. *(detenten13)*

[91] The *Heat intensity* parameter value {niedrige_Hitze} (low heat) has been neutralized (generalized) and might include such values as {auf kleiner Stufe}, {bei niedriger Hitze}, {bei niedriger Temperatur}, {bei schwacher Hitze} in the German corpus. For a full list of neutralization (generalization) of annotation parameters see Appendix 2.

19. 2 EL Wasser in die Pfanne geben und den Spinat *zugedeckt* 1 Minute *dünsten*. *(detenten13)*

[Pour 2 tbsp. water in a pan and sauté the spinach closed/covered for 1 minute].

20. In einer Pfanne etwas Butter erhitzen und die fein gehackten Schalottenwürfel darin *auf niedriger Stufe* glasig *dünsten*. (*(detenten13)*

[Heat some butter in a pan and sauté the finely diced shallots over low heat until translucent].

51	duensten	kochen	glasig_7.2	0.35556481689575
53	duensten	kochen	MITTLERE_HITZE_2	0.496615570728583
54	duensten	kochen	PFANNE_6.5	0.36139294286677
57	duensten	kochen	OEL_7.7	0.35198691903071
59	duensten	kochen	FETT_6.9	0.357756941604588
48	kochen	duensten	Herd_4.1	0.396411107369755
50	kochen	duensten	N_A_RES_2.9	0.43837909316984
55	kochen	duensten	N_A_UT_3.1	0.431062825102113
56	kochen	duensten	WASSER_4.3	0.392655016147625
58	kochen	duensten	N_A_SUB_6.7	0.359563873892813

Table (partial) 3.5. Actual distribution of the annotated occurrences in *dünsten-kochen* verb pair

In comparing the verb *dünsten* to *kochen*, again the parameter *Substance* would suffice to convey the semantic difference between these two verbs. Here, *kochen* either correlates with the verbalized *Substance* parameter value {Wasser} (water) or non-verbalized {N_A_SUB}, which is in the overwhelming majority of cases also {Wasser} (water), as presumed by the verb itself (cf. examples 21 and 22), while *dünsten* correlates with some type of *fat* or *oil*, e.g. {Öl} (oil) and {Fett} (fat) (partial Table 3.5).

21. Kartoffeln schälen, waschen und in Salzwasser ca. 20 Minuten *kochen*. *(REZ_DE)*

[Peel the potatoes, wash and cook* in salted boiling water for around 20 minutes].

22. Die Eier hart kochen, abschrecken, schälen und in Scheibe schneiden. *(REZ_DE)*

[Boil the eggs hard, rinse in cold water and cut them into slices].

The verb pair *dünsten-toasten* was discussed in detail at the beginning of this chapter (Fig. 1) when we compared *toasten* to the rest of the German five culinary verbs.

Thus, the more edges the verb has, the more parameters in which it differs from the other verbs being compared, which makes it more specific. This statement is true for all hyponym verbs since they are semantically denser and have considerably more specific parameter values differentiating them from the other verbs.

2.4.4 CCCG of the German Culinary Verb *braten*

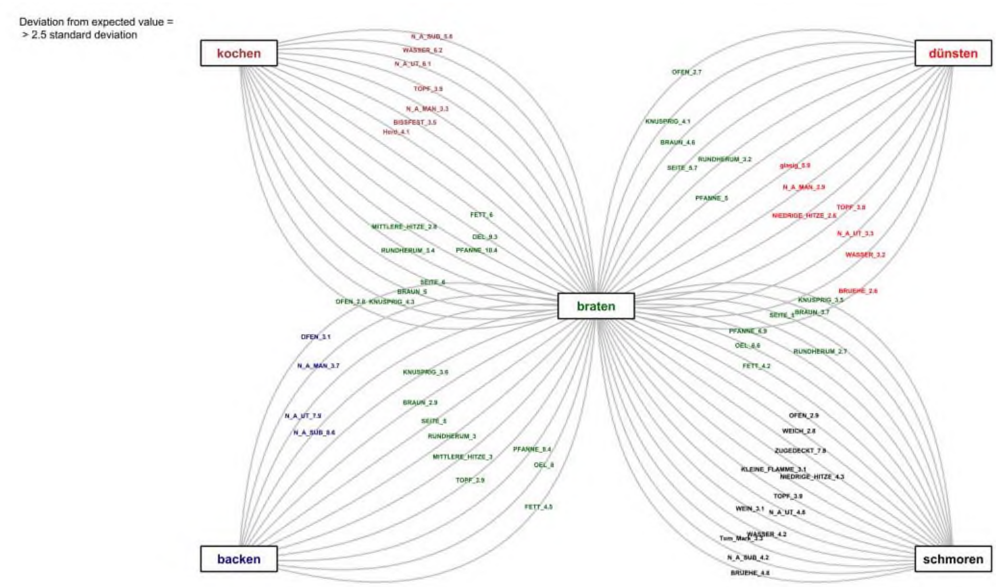


Fig. 4a. *CCCG* analysis of the German culinary *dünsten*

This *CCCG* illustrates in pairs the comparison of the German verbs *braten* with *kochen*, *dünsten*, *backen*, and *schmoren,*[92] displaying the semantic differences in the respective parameters (Fig. 4a). In the verb pairs *braten-backen, braten-dünsten, braten-kochen,* and *braten-schmoren,*[93] the semantic differences are observed in all the six annotation parameters. However, the semantic difference in the parameter *Substance* would suffice to distinguish *braten* from the other verbs being compared, as *braten* correlates with some type of *oil* or *fat* cooking substance as opposed to, for instance, *backen* correlating with no substance at all. *Kochen*, on the contrary, correlates with some type of *liquid* cooking substance, whereas *dünsten* might correlate either with *liquid* or *fat/oil* depending on the context, i.e. the other annotation parameters. For instance, when comparing the verbs *braten* and *kochen*, the number of co-occurrences of the *Substance* parameter value {Öl} (oil) and {Fett} (fat) with *braten* is superior to the expected value by 9.3 time and 6 times, respectively, of the standard deviation under the hypothesis of an independent uniform distribution. The number of co-occurrences of

---

[92] Cf. the classification of German culinary verbs in *Cooking Vocabularies and the Culinary Triangle of Lévi-Strauss* (Lehrer 1972).

[93] Except for the parameter *Heat source*, as the semantic closeness in this parameter is the outcome of non-verbalized annotated values denoting the non-differentiating, non-decisive character of this parameter for the German verb pairs *schmoren-braten.*

{Wasser} (water) and {N_A_SUB}[94] (no verbalized substance), on the contrary, is overrepresented for the verb *kochen* by 6.2 times and 5.8 times, respectively (cf. examples 23 and 26).

23. 3 EL *Öl* in einer großen Pfanne erhitzen und Garnelen darin rundherum ca. 6 Minuten *braten*. *(REZ_DE)*
    [In a large pan heat 3 tbs. oil and fry* the shrimps all around for about 6 minutes].

24. In einer Pfanne *Butterschmalz* erhitzen, die Steaks darin drei Minuten von jeder Seite *braten*. *(REZ_DE)*
    [Heat the ghee in a pan and fry* the steak for 3 minutes on each side].

25. Nudeln zugeben und 3–4 Minute bei mittlere Hitze *kochen* lassen. *(REZ_DE)*
    [Add the noodles and cook*/boil* for 3-4 minutes over medium heat].

26. Für Süßkartoffelpurée *kocht* man die Kartoffeln 5 Minuten länger, püriert sie mit den übrigen Zutaten und stellt sie bei geringer Hitze noch 15 Minuten auf den Herd. *(detenten13)* (no verbalized *Substance* parameter value, however, water is presumed by the valence of the verb *kochen* itself, see in Chapter 2.4.3 Fig. 3, partial Table 3.4).
    [For sweet potato puree one should cook*/boil* the potatoes for 5 minutes, purée them with the other ingredients and leave it on the stove for another 15 minutes over low heat].

41	braten	kochen	OFEN_2.8	0.405571913459486
44	braten	kochen	KNUSPRIG_4.3	0.369829257111758
46	braten	kochen	BRAUN_5	0.359550590555228
48	braten	kochen	SEITE_6	0.349767077273361
49	braten	kochen	RUNDHERUM_3.4	0.388327799916266
51	braten	kochen	kurz_2.5	0.422234690714885
52	braten	kochen	MITTLERE_HITZE_2.8	0.405986351842393
55	braten	kochen	PFANNE_10.4	0.328963826601522
58	**braten**	kochen	**OEL_9.3**	0.332419647437438
60	**braten**	kochen	**FETT_6**	0.35016681925545
42	kochen	braten	Herd_4.1	0.374006278164911
47	kochen	braten	BISSFEST_3.5	0.384715204411756

[94] The peculiarities of the non-verbalized values for the verb *kochen* were discussed in this chapter in comparing *dünsten* with the rest of the German culinary verbs (Fig.3 and accompanying partial Table 3.4).

50	kochen	braten	N_A_MAN_3.3	0.389891758557943
54	kochen	braten	TOPF_3.9	0.377690560359721
56	kochen	braten	N_A_UT_6.1	0.349517610301189
57	**kochen**	braten	**WASSER_6.2**	0.348088264899874
59	**kochen**	braten	**N_A_SUB_5.8**	0.351627701636025

Table (partial) 4.1. Actual distribution of the annotated occurrences in *braten-kochen* verb pair

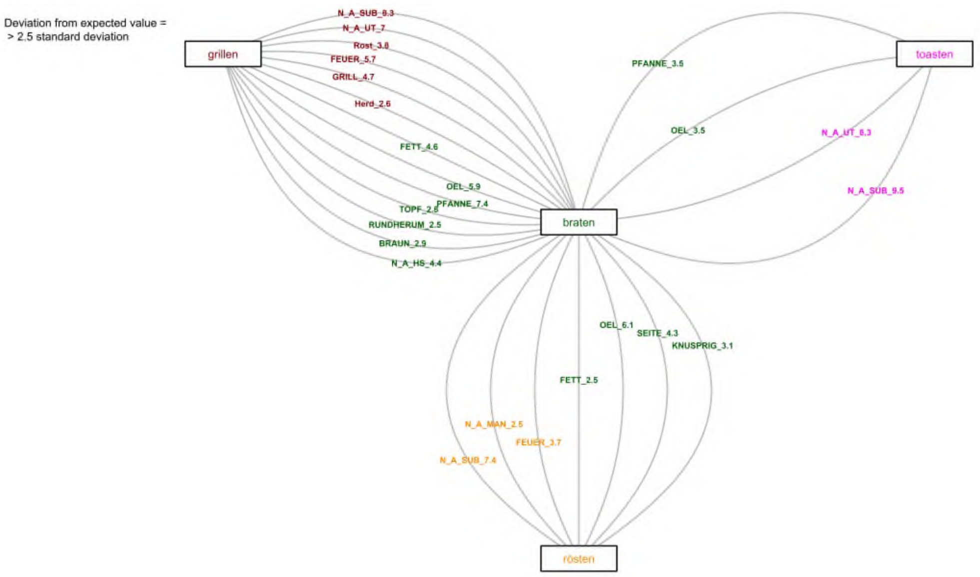


Fig. 4b. *CCCG* analysis of the German culinary verb *dünsten*

Semantic similarities are observed in pairs *braten-grillen*, *braten-rösten*, and *braten-toasten*. In the verb pair *braten-grillen*, there is no statistically significant difference in the distribution of the annotated occurrences for the parameters *Heat intensity*, while in *braten-rösten* there is also no significant difference in the parameter *Utensil* (Fig. 4b, partial Table 4.2).

63	grillen	braten	Herd_2.6	0.417260393995586
64	grillen	braten	GRILL_4.7	0.364274690412824
66	**grillen**	braten	**FEUER_5.7**	0.353023325829981
73	grillen	braten	Rost_3.8	0.379056941504209
75	grillen	braten	N_A_UT_7	0.343095437929944
77	grillen	braten	N_A_SUB_8.3	0.336283127192528
62	braten	grillen	N_A_HS_4.4	0.368014637070635
68	braten	grillen	BRAUN_2.9	0.402180129180536
69	braten	grillen	RUNDHERUM_2.5	0.418816024270125

71	braten	grillen	TOPF_2.8	0.40636863125135
74	braten	grillen	PFANNE_7.4	0.340353403120727
76	braten	grillen	OEL_5.9	0.350679098126549
78	braten	grillen	FETT_4.6	0.365428497218545
32	braten	roesten	KNUSPRIG_3.1	0.395621007328924
34	braten	roesten	SEITE_4.3	0.370215631375408
35	braten	roesten	RUNDHERUM_2.5	0.422254998792619
38	braten	roesten	OEL_6.1	0.349160811023924
40	braten	roesten	FETT_2.5	0.417872928904961
31	**roesten**	braten	**FEUER_3.7**	0.38138645127136
33	roesten	braten	GOLDGELB_2.4	0.425320095004359
36	roesten	braten	N_A_MAN_2.5	0.418079025785495
39	roesten	braten	N_A_SUB_7.4	0.340332657342963

Table (partial) 4.2. Actual distribution of the annotated occurrences in *braten-grillen* and *braten-rösten* verb pairs

In both verb pairs, despite the semantic differences in almost all six annotation parameters,[95] *Substance* is the most decisive and distinguishing one since *braten* correlates strongly with some type of *oil*, while *grillen* and *rösten* mostly show *no verbalized Substance* parameter values.[96] However, the parameter *Heat source* also plays a decisive role in the distribution of occurrences in the verb pairs *braten-grillen* and *braten-rösten*. For instance, when comparing the verbs *braten* and *rösten*, the co-occurrences of {Feuer} (fire) with *grillen* surpass the expected value by 5.7-fold the standard deviation under the hypothesis of an independent uniform distribution (examples 27 and 28).

27. Zum Abschluss machen wir ein *Lagerfeuer* und *grillen* die Bratwürste. *(detenten13)*
    [At the end we make a campfire and grill the bratwursts].
28. Dort hatten die Kinder die Gelegenheit am offenen *Feuer* Würstel zu *grillen*. *(detenten13)*

[95] The semantic difference between the German cooking verb *braten* and *grillen* in the parameter *Utensil* is also quite decisive as the overwhelming majority of the annotated examples of *braten* in our German sub-corpus co-occur with some type of a pan neutralized as {Pfanne} (pan) while *grillen* co-occurs with {Rost} (grid/rack). (For the list of all generalizations/neutralizations see Appendix 2). The following examples from our German sub-corpus *detenten13* is the evidence that in the production of authentic culinary texts, the choice of just the utensil might differentiate one verb from the other that could not be used interchangeably. “Während das Fleisch auf dem *Rost gegrillt* wird, *braten* Sie den Räucherspeck in einer *Pfanne* aus”. [While the meat grills* on the *grid**, fry* the smoked bacon in a *pan*]. For more analysis on the correlations of the verb *grillen* and *braten* and the differences in the annotation parameter see also Chapter 2.2. of this work on *Correspondence analysis* (CA) vs. *Multiple-correspondence analysis* (MCA) *of the German cooking verbs*. Fig.1 and Fig.5 here focus among others on the aforementioned verbs and their correlations with the *Utensil* parameter.

[96] The peculiarities of the parameter *Substance*, especially its *non-verbalized values* initially all annotated as {N_A_SUB} as opposed to {ohne Fett} (without oil or fat) co-occurring the verb *rösten* were addressed and elaborated upon followed by a detailed qualitative analysis in Chapter 2.4.2 (Fig. 2, 2.1, 2.2, and 2.3).

[There the children have the opportunity to grill sausages on an open fire].

In the verb pair *braten-toasten,* the semantic difference is observed in the parameters *Utensil* and *Substance* (Fig. 4). However, as most of the occurrences for both *braten* and *toasten* have no verbalized values in the parameters *Heat intensity* and *Heat source,* the lack of data does not allow for a conclusive assessment of the semantic closeness of these verbs in the aforementioned parameters. Instead, it could be stated that there is no statistically significant difference between *braten* and *toasten* in the parameters *Resultative (adjectives)* and *Manner*.[97]

[97] For a more detailed comparison of the verb *toasten* with the other German culinary verbs see Fig. 1 and its interpretation at the beginning of this chapter.

2.4 *Conditional Inference Trees* (CIT) of the German Culinary Verbs

Conditional inference trees (CITs) constitute "a non-parametric[98] class of regression trees" anchored in tree-based structures available in the programming language *R* for statistical computation, particularly the *partykit* package (Hothorn, T. et al. 2006:1).

Schweinberger points out the significance of the *Conditional Inference Trees* by contrasting them with other tree-based structures, such as decision trees, stating that CITs "represent recursive partitioning of the data to minimize residual deviance". Being a non-parametric type of test, *CITs* "use significance tests, e.g. the p-values, to select variables rather than some information measures like the Gini coefficient. [The] p-values determine the splits in the data"[99] (cf. Strobl et al. 2009, Gries 2021) (https://slcladal.github.io/tree.html#Conditional_Inference_Trees), meaning that the p-value serves as a predictor to determine whether a node shall be split into further inner or leaf nodes. If the predictor finds the splitting parameter significant, the latter is included in the conditional inference tree. The lower the p-value, the lower the probability that the distribution is an accidental result of a random data selection. In our case, lower p-values mean that the split in the tree reflects a real difference between two nodes as far as the verb distribution is concerned rather than an accidental outcome of a random data selection.

When comparing the *Conditional Inference Trees* (CITs) and *Conditional Random Forests* (CRFs) to other recursive partitioning methods, Levshina states that CITs and CRFs "allow [researchers] to model and interpret the relationships between a numeric or categorical response variable and various predictors," especially in subfields such as corpus linguistics, where the collected data may be limited in size or require considerable time and resources (2020:612). CITs and CRFs assist researchers in identifying which "linguistic factors help to predict the use of particular linguistic variant" (ibid. 614). Lohmann (2013:1) unequivocally recommends classification trees and random forests in linguistic studies when regression modeling becomes challenging to interpret due to the nature of data, such as "the strong correlation of predictor variables and/or complex interaction of the predictor variables". He employed the classification trees and random forests as a methodology in his corpus-based analysis of English word

[98] Non-parametric test differs from the parametric one in that the former does not assume that the distribution is normal whereas the latter does. Unlike parametric test, the non-parametric modeling does not make any assumption on the distribution of the data being observed (see Kotz, S. et al. 2006).

[99] In his article *Tree-Based Models in R* published on the webpage of the Language Technology and Data Analysis Laboratory, Martin Schweinberger introduces several tree-structured analysis models in *R*, among them the CART (classification and regression trees) and the CITs (Conditional Inference Trees) as an extension of basic decision trees (Schweinberger 2021).

formation choices, in particular in back- and fore clipping in proper and common nouns (ibid. 5-12).

Our *CIT* model shows how and what factors determine the choice of a verb in a certain context type (understood as clusters with specific combinations of parameter values).

However, before going into detail on the *CIT* analysis, its structure is introduced below based on one of our examples, followed by the interpretation of all *CIT* clusters arranged in three groups.

As mentioned above, in our case, the *CIT* model contributes to the illustration of those context clusters where, provided a particular distribution of verbs, a quantitatively dominant verb stands out. The following CIT modeling is one of the outcomes of our previously annotated German original corpora, described at the beginning of Chapter 2. Such modeling could be useful not only in original German text production, but also in finding possible equivalents in the target language(s). In this model, six of our eight parameters (variables) determine the clusters of possible contexts with essentially different distributions of the German culinary verbs.

The parameters *Ingredient* and *Dish* proved difficult to integrate into this model, mainly for two reasons based on the annotation of our two integrated German corpora, viz. the *German Web 2013 (detenten13)* and the *REZ_DE*. First, the annotation of the parameter *Ingredient* revealed too many distinct values to generalize into more or less large groups, resulting in too small clusters with no significant results. Second, the lack of data for the parameter *Dish* hindered its comparability with the rest of the parameters.

The following CIT modeling was generated with the same 1999 manually annotated occurrences extracted from the aforementioned *detenten13* and *REZ_DE* German corpora. However, due to the reduction of parameter values, the quantity of the annotated examples was reduced to 1809;[100] the same set of data used for the *Mosaic-Plot* and the *M(CA).*

[100] For a more detailed information of the German corpus description and parameter reduction see Chapter 2.1 of this work.

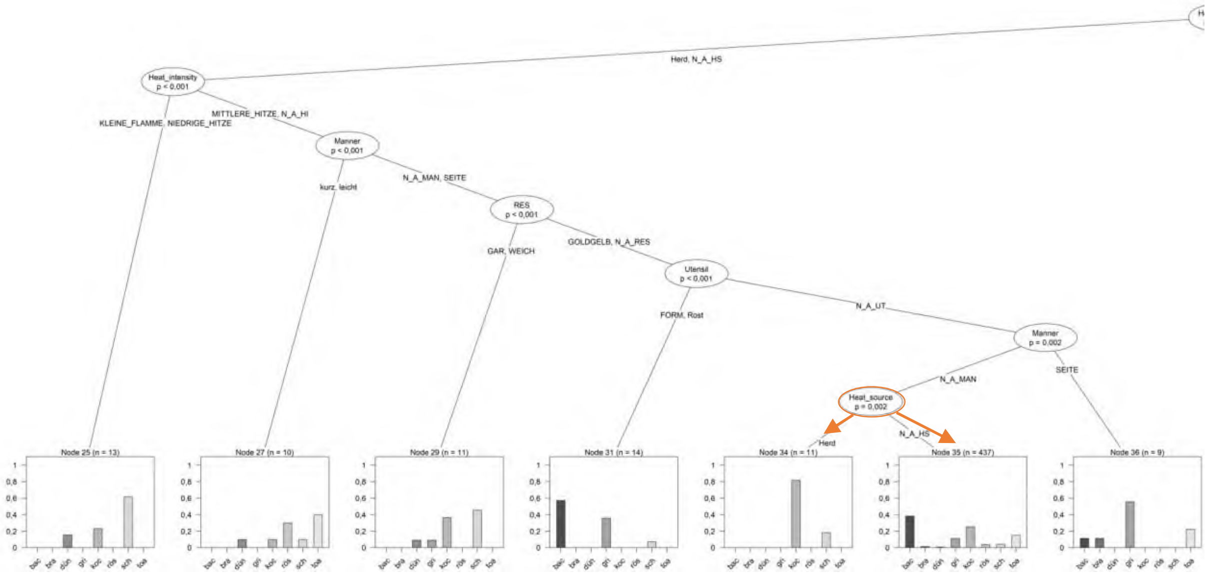


Fig. 1. Partial *CIT*. Distribution model of the German culinary verbs in context[101]

The initials of the six parameters (variables) ***S**ubstance*, ***U**tensil*, ***H**eat Intensity*, ***M**anner*, ***R**esults*, ***H**eat source* (Fig. 1) contributing to the illustration of our *CIT* model are written in the circles placed above the nodes from which the partitioning takes place, and where the inner or terminal nodes (leaves) are generated from. This splitting into further nodes continues until there is any statistically significant partitioning no longer possible. The p-values (probability value) under the parameter initials show the probability of the split being accidental. For instance, p<0.002 under the parameter *Heat source* splitting the nodes 34 and 35 shows that the likelihood of a merely accidental difference in the distribution between the nodes is smaller than 0,2%. Lower p-values imply a high probability that the split in the tree reflects a real difference. The splitting in this *CIT* model resulted in 36 inner nodes and 37 terminal nodes, allowing, among others, to identify the most frequently used verb in context, represented by more prominent columns at the bottom of the graphic. The verbs are placed in alphabetical order, starting with *backen* and ending with *toasten* with *braten*, *dünsten*, *grillen*, *kochen*, *rösten*, and *schmoren* in between. Let us now elaborate on those parameter-defined context clusters which reveal one of the verbs as the most frequently used. The overall frequency of the dominant verbs in the identified clusters is 61,5%, meaning that using a particular parameter combination as a heuristic approach to "guess" an adequate lexical choice in German – or to

[101] For the full CIT model of German culinary verbs see Appendix 4. For the complete CIT model of the German verbs in real size as well as the accompanying table of the split nodes comprising the parameter defined context clusters see Appendix 4.

find an adequate translation in German – by simply selecting the most frequent verb in each cluster would result in an average "error" rate of ca. 38%. Choosing the most frequent verb in a cluster instead of another does not automatically imply that the choice would be inadequate or incorrect. The error rate here denotes the degree of deviation from the most frequently used verb in context.

This model could be compared to both the distribution of the German culinary verbs according to their frequency in the *deTenTen13* corpus and to the random choice of any verb from the eight ones outside the corpus under the hypothesis of an independent discrete uniform distribution. In the first case, one could theoretically choose the most frequent verb, viz. *kochen* in the *deTenTen 13* corpus (Fig. 2), which would result in around 51% correct, and consequently 49% incorrect verb choice.

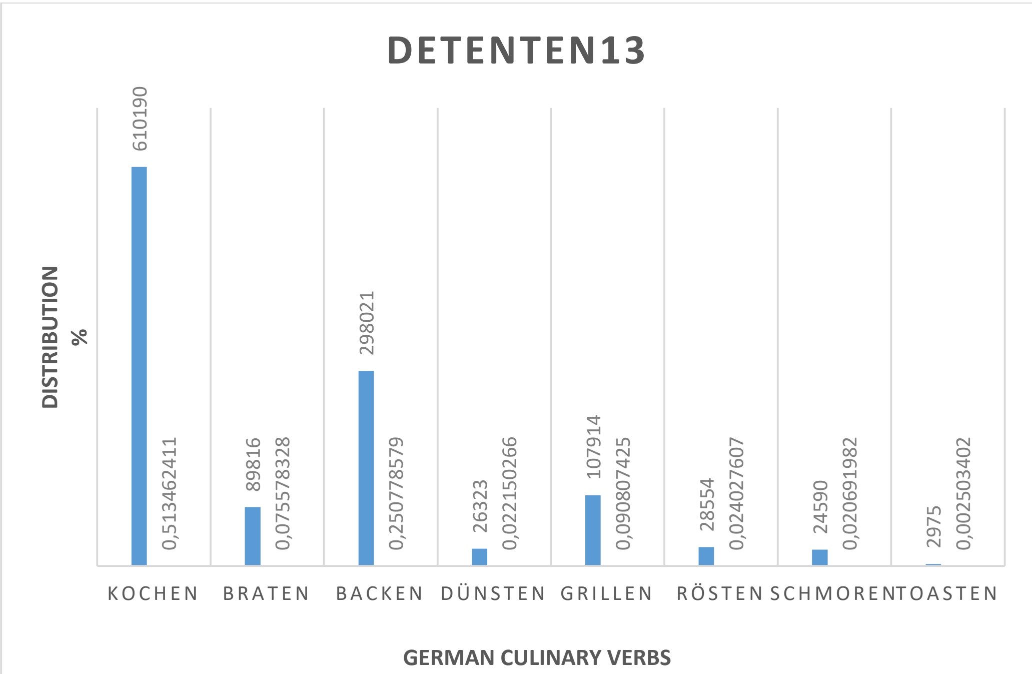


Fig. 2. Distribution of the German culinary verbs in the *deTenTen13* corpus

The *CIT* model also identifies the most frequent verb in a given cluster; however, the partitioning of the *CIT* is based on our manual semantic annotation and not merely on the distribution of the German culinary verbs in the *deTenTen13* corpus according to frequency. This allowed the percentage of the correctly identified verbs to rise to 61,5%, reducing the error rate by approximately 10,5% (Fig. 3).

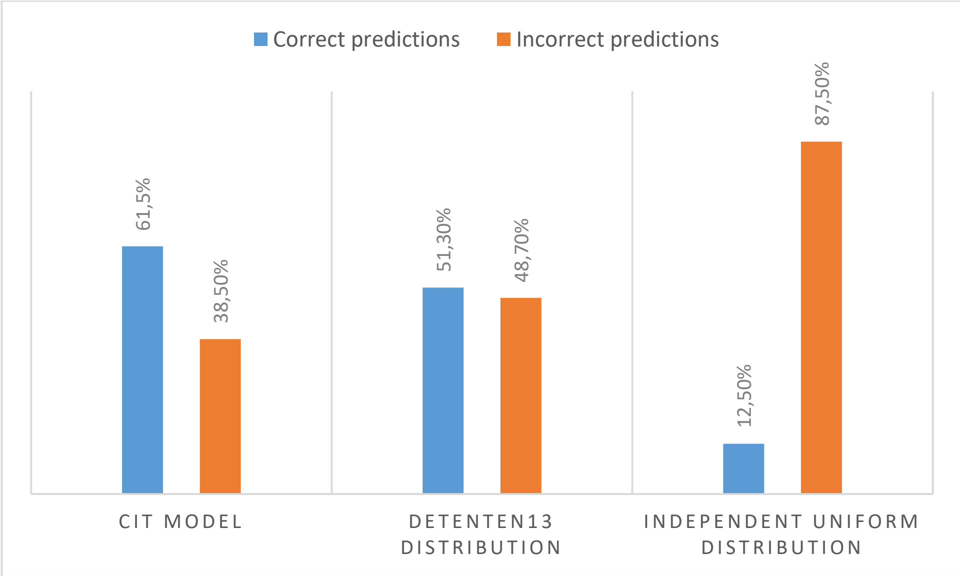


Fig. 3. *CIT* model in comparison: correct and incorrect prediction of the verb

Our *CIT* model is a significant improvement with respect to the hypothesis of an independent distribution of the verbs, where, under the hypothesis of a discrete uniform distribution, the probability of choosing the correct verb would be 1/8, i.e. 12,5%, so that a random guess would produce an error rate of 87,5%. Even if we consider those *CIT* clusters with a considerably higher error rate, for instance, node 35 with 437 occurrences and a 61,8% error rate (Fig. 1), the percentage of incorrectly identified verbs in *CIT* would still be much lower, viz. approx. 38,5% (Fig. 3).

The identified clusters determined by context parameters are elaborated in three groups with respect to their error rate. The first group includes clusters with the lowest error rate, ranging from 0% to approximately 1/3. The second group comprises those clusters with a moderate error rate (from 1/3 up to 1/2), while the third group consists of clusters with a relatively higher error rate (above 1/2).

2.5.1 The First Group of Clusters Determined by Context Parameters

As mentioned above, this group includes clusters of contexts with the lowest error rate, ranging from 0% to approx. 30%. Our *CIT* model identified 14 clusters with 776 occurrences in this group.

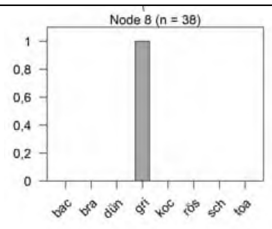
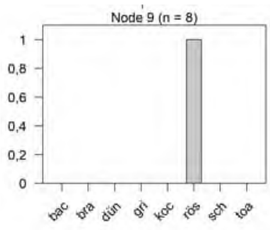
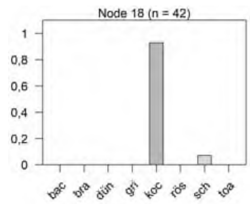
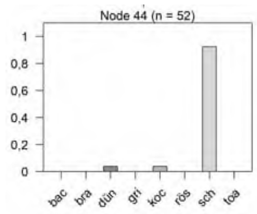
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
8	gri	38	0			Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Utensil (FORM, N_A_UT), Heat source (Glut, GRILL)
9	rös	8	0	bac = 0, bra = 0, dün = 0, gri = 0, koc = 0, rös = 8, sch = 0, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Utensil (FORM, N_A_UT), Heat source (ROESTEREI)
18	koc	42	7,14285714	bac = 0, bra = 0, dün = 0, gri = 0, koc = 39, rös = 0, sch = 3, toa = 0		Heat source (Herd, N_A_HS, OFEN), Manner (kurz, leicht, N_A_MAN, RUEHREN, SEITE, WENDEN), RES (BISSFEST, GAR, glasig, GOLDGELB, N_A_RES, WEICH), Substance (WASSER), Utensil (TOPF)
44	sch	52	7,69230769	bac = 0, bra = 0, dün = 2, gri = 0, koc = 2, rös = 0, sch = 48, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Heat source (Herd, N_A_HS, OFEN), Manner (ZUGEDECKT), Utensil (FORM, TOPF)

Table 1.1. The first group of context-parameter determined clusters with the lowest error rate

The first cluster in this group represents a context type (node 8) where *grillen*, with 38 annotated occurrences, is the only verb used in the corpus. The occurrences in this cluster are determined by the combination of the *Substance*, *Utensil*, and *Heat source* parameters with their {Brühe, N_A_substance, Tomatemark, Wasser, Wein}x{Form, N_A_utensil}x{Grill, Glut} values respectively. The second cluster, with a 0% error rate (node 9) (Table 1.1), illustrates the context type determined by almost the same parameter combinations as the first cluster, but with a slight alteration of the parameter *Heat source*, where the partitioning of the *CIT* model is influenced by the value {Rösterei}. This is, however, a rather small cluster with only 8 examples, all of them featuring the verb *rösten*.

The next cluster, with an overwhelming majority of occurrences (39 out of 42) featuring the verb *kochen* and with merely 7,15% error rate, can be tracked in the complete *CIT* model (Appendix 4) under node 18 (Fig. 1.1). Almost all of our six parameters with their relevant values, viz. possible combinations of {Herd, N_A_HS, Ofen}x{kurz, leicht, N_A_MAN, rühren, Seite, wenden}x{bissfest, gar, glasig, goldgelb, N_A_RES, weich}x{Wasser}x{Topf} – values of *Heat source, Manner, Resultive, Substance,* and *Utensil* parameters respectively – influenced the distribution of the verbs in this cluster by identifying *kochen* as the dominant one.

In the next cluster, under node 44, another context type with only around 7,7% error rate is observed. In this context type, the parameters *Substance*, *Heat source*, *Manner,* and *Utensil* with their values {Brühe, N_A_substance, Tomatenmark, Wasser, Wein}x{Herd, N_A_heat source, Ofen}x{zugedeckt}x{Form, Topf} played a significant role in the splitting of the nodes and proved the distribution to be not an accidental result of a random selection, thus illustrating *schmoren* as the most frequent verb in context. This is a rather big node with overall 52 examples, where 48 are with the verb *schmoren* (Table 1.1).

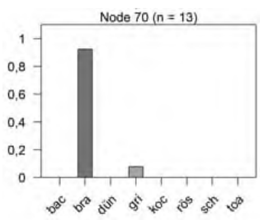
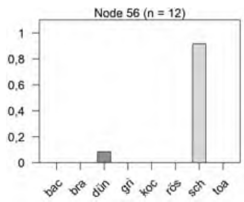
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
70	bra	13	7,69230769	bac = 0, bra = 12, dün = 0, gri = 1, koc = 0, rös = 0, sch = 0, toa = 0		Substance (FETT, OEL), Heat source (FEUER, N_A_HS, OFEN), Utensil (N_A_UT, TOPF), RES (GAR, N_A_RES), Manner (RUNDHERUM, SEITE)
56	sch	12	8,33333333	bac = 0, bra = 0, dün = 1, gri = 0, koc = 0, rös = 0, sch = 11, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Utensil (PFANNE, SPIESS), Manner (ZUGEDECKT)

Table 1.2. The first group of context-parameter determined clusters with the lowest error rate

The next small cluster, with 7,7% error rate, is observed under node 70, where the verb *braten* is identified as the most frequently used verb. This small cluster, featuring almost exclusively the verb *braten* (12 out of 13 examples), is determined by a specific combination of context parameters *Substance, Heat source, Utensil, Resultative*, and *Manner*, with {Fett, Öl}x{Feuer, N_A_HS, Ofen}x{N_A_UT, Topf}x{gar, N_A_RES}x{rundherum, Seite} values, respectively (Table 1.2).

The verb *schmoren* is represented as the most frequent verb in the next context type (node 56, error rate=8,3%) (Table 1.2). This is a rather small cluster with only 12 examples, 11 of which feature the verb *schmoren*. Here, the possible combination of parameters *Substance, Utensil*, and *Manner* with their respective values {Brühe, N_A_substance, Tomatenmark, Wasser, Wein}x{Pfanne, Spiess}x{zugedeckt} influenced the partitioning of the data.

Another cluster in the first group of context types, with the lowest error rate (13,6%), could be traced under node 61 (see Table 1.3. below; cf. Appendix 4). This is a considerably big cluster with as many as 110 occurrences, where the overwhelming majority features the verb *dünsten* and only 8 and 7 examples feature *braten* and *schmoren*, respectively. A combination of four parameters, namely *Substance, Utensil, Heat source*, and *Resultative (adjectives)* with their respective values {Fett, Öl}x{N_A_UT, Pfanne, Topf}x{Feuer, N_A_HS, Ofen}x{bissfest, glasig, weich} impacts the distribution in this cluster.

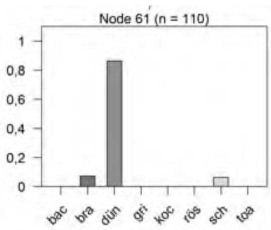
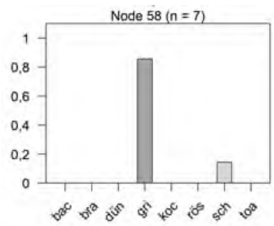
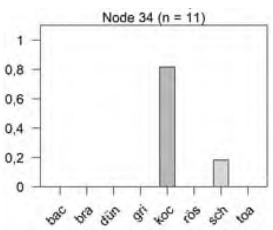
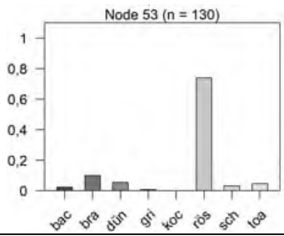
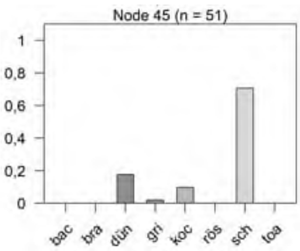
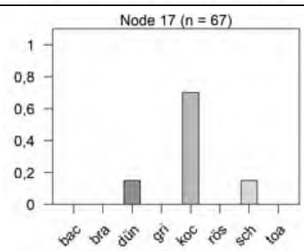
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
61	dün	110	13,6 3636 36	bac = 0, bra = 8, dün = 95, gri = 0, koc = 0, rös = 0, sch = 7, toa = 0		Substance (FETT, OEL), Utensil (N_A_UT,PFANNE, TOPF), Heat source (FEUER, N_A_HS, OFEN), RES (BISSFEST, glasig, WEICH)
58	gri	7	14,2 8571 43	bac = 0, bra = 0, dün = 0, gri = 6, koc = 0, rös = 0, sch = 1, toa = 0		Substance (FETT, OEL), Utensil (FORM, Rost)
34	koc	11	18,1 8181 82	bac = 0, bra = 0, dün = 0, gri = 0, koc = 9, rös = 0, sch = 2, toa = 0		Substance (N_A_SUB), Heat intensity (MITTLERE_HITZE, N_A_HI), RES (GOLDGELB, N_A_RES), Utensil (N_A_UT), Manner (N_A_MAN), Heat source (Herd)

Table 1.3. The first group of context-parameter determined clusters with the lowest error rate

Two rather small clusters are traced under nodes 58 and 34 (with 14,2% and 18,2% error rates, respectively), where *grillen* and *kochen* are identified the most frequent verb (Table 1.3). Parameter combinations determining the context type for the former cluster are *Substance* and *Utensil* with values {Fett, Öl}x{Form, Rost}, while for the latter, all six annotation parameters impacted the distribution. Node 53 exposes the next big cluster, with 26,1% error rate, where the distribution of the German culinary verbs is influenced by the parameter combinations[102] of *Substance*, *Resultative*, *Utensil*, *Manner*, and *Heat source*, and their respective values described in the table below. In this context type, with 96 examples out of the total 130 occurrences, *rösten* is identified as the most frequent verb (Table 1.4).

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
53	rös	130	26,1538462	bac = 3, bra = 13, dün = 7, gri = 1, koc = 0, rös = 96, sch = 4, toa = 6		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), RES (BRAUN, GOLDGELB, N_A_RES), Utensil (PFANNE), Manner (kurz, leicht, N_A_MAN, RUEHREN, RUNDHERUM, WENDEN), Heat source (N_A_HS, ROESTEREI)
45	sch	51	29,4117647	bac = 0, bra = 0, dün = 9, gri = 1, koc = 5, rös = 0, sch = 36, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Heat source (Herd, N_A_HS, OFEN), Manner (ZUGEDECKT), Utensil (N_A_UT, Rost)
17	koc	67	29,8507463	bac = 0, bra = 0, dün = 10, gri = 0, koc = 47, rös = 0, sch = 10, toa = 0		Heat source (Herd, N_A_HS, OFEN), Manner (kurz, leicht, N_A_MAN, RUEHREN, SEITE, WENDEN), RES (BISSFEST,GAR, glasig, GOLDGELB,

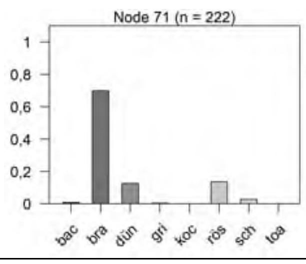
[102] It is important to note that by saying "combination with these parameter values", not all values necessarily create a possible context combination for exactly the dominant verb identified in the cluster. On the contrary: certain parameter value combinations create context(s) for the dominant verb, while others are for the other verbs sharing the distribution of occurrences in that specific node. For instance, *rösten* excludes any substance containing water in its description (cf. *M(CA)* the German verbs in Chapter 2.3 and the *Mosaic-Plot* of the German verbs in relation to the parameter *Substance* in Chapter 2.4.1 of this work) but is included in this cluster, since four occurrences in node 53 also feature the verb *schmoren*, which allows the use of water-based substances in its semantic field. Here a further partitioning of the node based on the parameter *Substance* is not possible due to a small number of occurrences.

						N_A_RES, WEICH), Substance (WASSER), Utensil (N_A_UT)

Table 1.4. The first group of context-parameter determined clusters with the lowest error rate

In the next cluster (node 45), the combination of the parameters *Substance*, *Heat source, Manner,* and *Utensil*, with their values {Brühe, N_A_Substance, Tomatenmark, Wasser, Wein}x{Herd, N_A_HS, Ofen}x{zugedeckt}x{N_A_Utensil, Rost}, affected the distribution of the verbs, identifying *schmoren* with 36 occurrences as the most frequent one. Node 17 with overall 67 occurrences illustrates the next cluster, where the verb *kochen*, with as many as 47 examples, is singled out as the most frequent verb in context. Both clusters have a low error rate of around 29% (Table 1.4).

The last two clusters in our first group of context types have an error rate of around 30%. Node 71 illustrates the biggest cluster of the first group, with overall 222 examples (Fig. 15). The verb *braten* stands out in this cluster as the most frequently used verb, with the overwhelming majority of occurrences (155 examples). The parameters that defining this cluster are *Substance, Heat source*, *Resultative* (adjectives), *Manner*, and *Utensil*, with their values {Fett, Öl}x{Feuer, N_A_HS, Ofen}x{braun, gar, goldgelb, knusprig, N_A_Resultative}x{kurz, leicht, N_A_Manner, rühren, rundherum, Seite, wenden}x{Pfanne}, respectively. Due to so many various parameter value combinations, less frequent verbs of the cluster, such as *dünsten*, *grillen*, *rösten*, and *schmoren*, also share a part in the distribution with a much lower number of occurrences.

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
71	bra	222	30,18018 02	bac = 2, bra = 155, dün = 28, gri = 1, koc = 0, rös = 30, sch = 6, toa = 0		Substance (FETT, OEL), Heat_source (FEUER, N_A_HS, OFEN), RES (BRAUN, GAR, GOLDGELB, KNUSPRIG, N_A_RES), Manner (kurz, leicht, N_A_MAN, RUEHREN, RUNDHERUM, SEITE, WENDEN), Utensil (PFANNE)

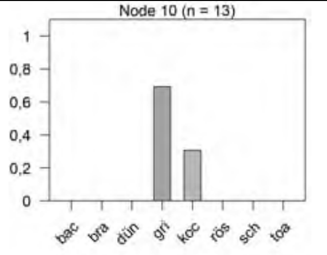
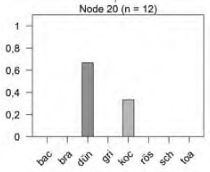
10	gri	13	30,76923 08	bac = 0, bra = 0, dün = 0, gri = 9, koc = 4, rös = 0, sch = 0, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Heat_source (FEUER, Glut, GRILL, ROESTEREI), Utensil (Rost, TOPF)

Table 1.5. The first group of context-parameter determined clusters with the lowest error rate

Node 10 (Table 1.5) exposes another context type where, given certain parameter combinations, the distribution of the German culinary verbs in question is determined. The parameters *Substance*, *Heat source* and *Utensil*, with their respective value combinations {Brühe, N_A_Subatance, Tomatenmark, Wasser, Wein}x{Feuer, Grill, Glut, Rösterei}x{Rost, Topf} were decisive for the distribution of this node, with the overwhelming majority of occurrences featuring the verb *grillen* (9 out of 13) and an additional 4 with *kochen*. Consequently, *grillen* is identified as the most dominant verb for this cluster.

2.5.2 The Second Group of Clusters Determined by Context Parameters

As mentioned earlier in this chapter, the second group of clusters exposes context types with a moderate error rate (from 33,3% to 50%), where certain parameter combinations impacted the distribution of the German culinary verbs. This group overall comprises 15 clusters with 461 occurrences. The clusters in this group are worth considering, since even with a 50% error rate, they led to a noticeable improvement of 37,5% if compared to the random distribution of the German culinary verbs outside the corpus (error rate without the model=87,5%), and allowed for correctly identifying the verb in at least 12 clusters in this group, leaving only 3 clusters with the same error rate as the incorrect predictions if we were to take the most frequent verb in the *deTenTen13* corpus (see Fig. 2 and Fig. 3) .

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
20	dün	12	33,333 3333	bac = 0, bra = 0, dün = 8, gri = 0, koc = 4, rös = 0, sch = 0, toa = 0		Utensil (FORM, N_A_UT, Rost, TOPF), Heat source (Herd, N_A_HS, OFEN), Manner (kurz, leicht, N_A_MAN, RUEHREN, SEITE, WENDEN), Substance (N_A_SUB), RES (BISSFEST, glasig)

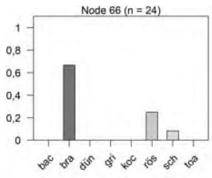
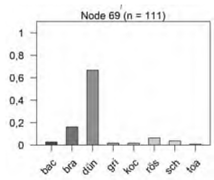
66	bra	24	33,3333333	bac = 0, bra = 16, dün = 0, gri = 0, koc = 0, rös = 6, sch = 2, toa = 0		Substance(FETT, OEL), Heat source (FEUER, N_A_HS, OFEN), Manner (kurz, leicht, N_A_MAN, RUEHREN, RUNDHERUM, SEITE, WENDEN), Utensil (N_A_UT, TOPF), RES (BRAUN, KNUSPRIG)
69	dün	111	33,3333333	bac = 3, bra = 18, dün = 74, gri = 2, koc = 2, rös = 7, sch = 4, toa = 1		Substance (FETT,OEL), Heat source (FEUER, N_A_HS, OFEN), Utensil(N_A_UT, TOPF), RES (GAR, N_A_RES), Manner (kurz, N_A_MAN, RUEHREN)

Table 2.1. The second group of context-parameter determined clusters with moderate error rate

The first three clusters in this group (nodes 20, 66, and 69) all have the same 33,3% error rate (Table 2.1). The first two are relatively smaller clusters, while the third one is rather big, with 111 examples in total. Despite a somewhat higher error rate, the distribution of the verbs in these clusters is also affected by certain parameter combinations where the most frequent verb is identified. In the first cluster, with 12 examples, 8 feature the verb *kochen*. Here five of our six parameters influenced the distribution, where *dünsten* also shares the distribution due to common parameter values with *kochen*, for instance, {Topf} from the parameter *Utensil* and {bissfest} from *Resultative (adjectives)*.

The parameter combinations having impacted the distribution of the German culinary verbs in the cluster under node 66 are *Substance, Heat source, Manner, Utensil,* and *Resultative (adjectives)*, with their respective values {Fett, Öl}x{Feuer, N_A_heat source, Ofen}x{kurz, leicht, N_A_manner, rühren, rundherum, Seite, wenden}x{N_A_utensil, Topf}x{braun, knusprig}, thus identifying *braten* as the most frequent verb in the cluster, with 16 examples (out of a total of 24) in the node.

Node 69 exposes, as mentioned above, a rather big cluster where the overwhelming majority of the examples (76) are with the verb *dünsten*. The distribution in this cluster is influenced by the parameters *Substance, Heat source, Utensil, Resultative,* and *Manner*, with the values {Fett, Öl}x{Feuer, N_A_heat source, Ofen}x{N_A_utensil, Topf}x{gar, N_A_RES}x{kurz, N_A_manner, rühren}, respectively. It is worth noting, however, that all our 8 German culinary verbs are also identified in the distribution of this cluster, but the most frequent verb *dünsten* has four times more examples than the second most frequent, *braten*.

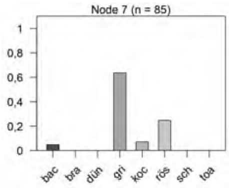
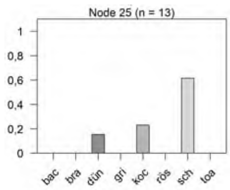
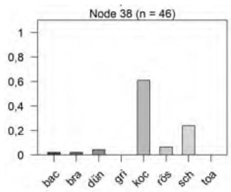
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
7	gri	85	36,4705882	bac = 4, bra = 0, dün = 0, gri = 54, koc = 6, rös = 21, sch = 0, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Utensil (FORM, N_A_UT), Heat source (FEUER)
25	sch	13	38,4615385	bac = 0, bra = 0, dün = 2, gri = 0, koc = 3, rös = 0, sch = 8, toa = 0		Substance (N_A_SUB), RES (GAR, GOLDGELB, N_A_RES, WEICH), Manner (kurz, leicht, N_A_MAN, SEITE), Utensil (FORM, N_A_UT, Rost), Heat source (Herd, N_A_HS), Heat intensity (KLEINE_FLAMME, NIEDRIGE_HITZE)
38	koc	46	39,1304348	bac = 1, bra = 1, dün = 2, gri = 0, koc = 28, rös = 3, sch = 11, toa = 0		Heat source (Herd, N_A_HS, OFEN), Substance (N_A_SUB), RES(GAR, GOLDGELB,N_A_RES, WEICH), Manner (kurz, leicht, N_A_MAN, SEITE), Utensil (TOPF)

Table 2.2. The second group of context-parameter determined clusters with moderate error rate

The distribution in the next cluster, represented under node 7, is influenced by the parameters *Substance, Utensil*, and *Heat source* with their respective values {Brühe, N_A_substance, Tomatenmark, Wasser, Wein} x {Form, N_A_utensil}x{Feuer}, where *grillen* was identified as the most frequent verb, with more than half of the overall examples in the cluster. Decisive for this context type, however, is the *Heat source* parameter value {Feuer}, resulting in the verb *rösten*, with 21 examples, being the second most frequent verb in the cluster (Table 2.2).

The next cluster under node 25 is rather small, with only 13 examples and around 38,5% error rate, where the most frequent verb *schmoren* is singled out. The parameters having impacted the distribution of verbs in this cluster are represented in the table above.

Another cluster in this group of context types can be traced under node 38. With around 39% error rate, the combination of *Heat source, Substance, Resultative (adjectives), Manner* and *Utensil,* with their values {Herd, N_A_HS, Ofen}x{N_A_substance}x{gar, goldgelb, N_A_resultative, weich}x{kurz, leicht, N_A_manner, Seite}x{Topf}, respectively, influenced the distribution of the German culinary verbs in this cluster, identifying *kochen* as the most frequent verb in context. Out of the total 46 annotated occurrences in the node, 28 feature the verb *kochen*, even though due to common parameter values other verbs share the distribution as well; only *grillen* and *toasten* are not represented in the distribution in this cluster (Table 2.2).

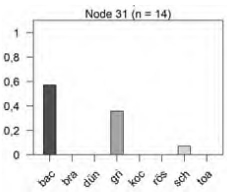
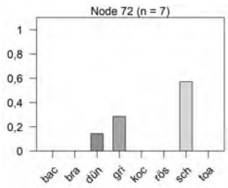
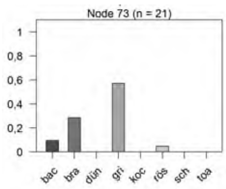
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/path
31	bac	14	42,8 5714 29	bac = 8, bra = 0, dün = 0, gri = 5, koc = 0, rös = 0, sch = 1, toa = 0		Substance (N_A_SUB), Heat source (Herd, N_A_HS), Heat intensity (MITTLERE_HITZE, N_A_HI), Manner (N_A_MAN, SEITE), RES (GOLDGELB, N_A_RES), Utensil (FORM, Rost)
72	sch	7	42,8 5714 29	bac = 0, bra = 0, dün = 1, gri = 2, koc = 0, rös = 0, sch = 4, toa = 0		Substance (FETT, OEL), Utensil (N_A_UT, PFANNE, TOPF), Heat source (FEUER, N_A_HS, OFEN), RES (BRAUN, GAR,GOLDGELB, KNUSPRIG, N_A_RES), Manner (ZUGEDECKT)
73	gri	21	42,8 5714 29	bac = 2, bra = 6, dün = 0, gri = 12, koc = 0, rös = 1, sch = 0, toa = 0		Substance (FETT,OEL), Utensil (N_A_UT, PFANNE, TOPF), Heat source (GRILL, Herd)

Table 2.3. The second group of context-parameter determined clusters with moderate error rate

The next three clusters, under nodes 31, 72, and 73, are rather small, ranging from 7 to 21 examples. All clusters have error rate of around 43%. The parameters affecting the distribution of the smallest of the aforementioned clusters, under node 72, can be traced through its path represented in the table above (Table 2.3). The same table also illustrates the path of the cluster under node 31, where *backen,* with 8 examples out of 14, is the most frequent verb in context. Three parameters, i.e. *Substance, Utensil,* and *Heat source,* with values such as {Fett, Öl}x {N_A_Pfanne, Topf}x{Grill, Herd}, influenced the distribution of verbs in this cluster, where *grillen* is identified as the most frequent verb. However, the verbs *backen*, *braten*, and *rösten* also have a share in the distribution. One common feature observed in the clusters described above is that, due to the numerous combinations of parameter values defining these clusters (especially nodes 31 and 72), almost all of our eight German culinary verbs are represented as contextually relevant.

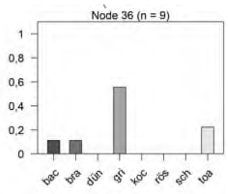
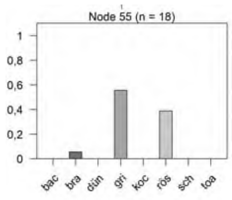
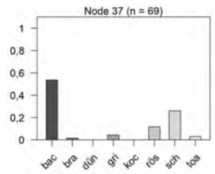
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
36	gri	9	44,4444444	bac = 1, bra = 1, dün = 0, gri = 5, koc = 0, rös = 0, sch = 0, toa = 2		Substance (N_A_SUB), Heat source (Herd, N_A_HS), Heat intensity (MITTLERE_HITZE, N_A_HI), RES (GOLDGELB, N_A_RES), Utensil (N_A_UT), Manner (SEITE)
55	gri	18	44,4444444	bac = 0, bra = 1, dün = 0, gri = 10, koc = 0, rös = 7, sch = 0, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Manner (kurz, leicht, N_A_MAN, RUEHREN, RUNDHERUM, SEITE, WENDEN), RES (BRAUN, GOLDGELB, N_A_RES), Utensil (SPIESS)
37	bac	69	46,3768116	bac = 37, bra = 1, dün = 0, gri = 3, koc = 0, rös = 8, sch = 18, toa = 2		Substance (N_A_SUB), RES (GAR, GOLDGELB N_A_RES, WEICH), Manner (kurz, leicht, N_A_MAN, SEITE), Utensil (FORM, N_A_UT, Rost), Heat source (OFEN)

Table 2.4. The second group of context-parameter determined clusters with moderate error rate

The first two clusters in the sub-group presented in the table above are relatively small, with only 9 and 18 examples respectively and around 44,5% error rate; therefore, they will not be described in detail. However, the parameters determining these clusters can be traced through their path (see nodes 36 and 55) (Table 2.4). The third cluster, on the contrary, is rather big with 69 examples. The distribution in this cluster under node 37 is also influenced by combination(s) of certain context parameters, namely *Substance*, *Resultative*, *Manner*, *Utensil*, and *Heat source*, with their values {N_A_Substance}x{gar, goldgelb, N_A_Resultative, weich}x{kurz, leicht, N_A_Manner, Seite}x{Form, N_A_utensil, Rost}x{Ofen}, identifying the verb *backen*, with an error rate of around 46,3% as the most frequent verb, accounting for more than half of the overall occurrences (Table 2.4). Due to shared parameter values, such as the absence of substance during cooking (annotated as {N_A_SUB}) or {Ofen}, almost all verbs are included in the distribution of this cluster, except for *dünsten* and *kochen*.

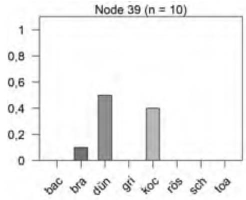
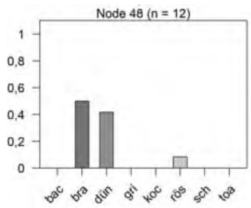
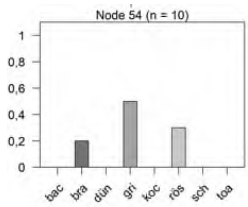
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
39	dün	10	50	bac = 0, bra = 1, dün = 5, gri = 0, koc = 4, rös = 0, sch = 0, toa = 0		Utensil (FORM, N_A_UT, Rost, TOPF), Heat source in (Herd, N_A_HS, OFEN), Substance (N_A_SUB), RES (GAR, GOLDGELB, N_A_RES, WEICH), Manner (RUEHREN)
48	bra	12	50	bac = 0, bra = 6, dün = 5, gri = 0, koc = 0, rös = 1, sch = 0, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), Utensil (PFANNE, SPIESS), Manner (kurz, leicht, N_A_MAN, RUEHREN, RUNDHERUM, SEITE, WENDEN), RES (BISSFEST, GAR, glasig, KNUSPRIG, WEICH)
54	gri	10	50	bac = 0, bra = 2, dün = 0, gri = 5, koc = 0, rös = 3, sch = 0, toa = 0		Substance (BRUEHE, N_A_SUB, Tom_Mark, WASSER, WEIN), RES (BRAUN, GOLDGEL, N_A_RES), Utensil (PFANNE), Manner (SEITE)

Table 2.5. The second group of context-parameter determined clusters with moderate error rate

In the last three clusters of this group of context types (nodes 39, 48, and 54), the distribution of verbs with respect to the occurrences generates an error rate of 50%. Consequently, in these clusters the most frequent verbs, namely *dünsten*, braten, and *grillen*, account for only half of the occurrences in these nodes. Since the clusters are relatively small, with only 10-12 examples, a more detailed description of possible parameter combinations will not be elaborated upon here. However, it is possible to trace their paths in the table above.

2.5.3 The Third Group of Clusters Determined by Context Parameters

This group comprises 8 clusters with overall 572 occurrences, of which 437 are under node 35, mentioned at the beginning of this chapter. These are apparently the least accurately described clusters in our CIT model, with the highest error rate ranging from 51% up to 63,2%. As mentioned above, it was decided not to include the largest cluster in this group under node 35, with an error rate of around 62%, due to insufficient data resulting from almost exclusively non-verbalized parameter values annotated as N_As for all parameters. Exceptions are the values {mittlere_Hitze} and {goldgelb} in the parameters *Heat intensity* and *Resultative (adjective)*, respectively (Table 3.1).

The next cluster, nevertheless, with a 51% error rate, is worth our attention since it is relatively large, with a total of 49 examples, where *schmoren* is identified as the most frequent verb in context due to the defining parameter combinations provided in the path. Moreover, besides *schmoren*, only *dünsten* and *kochen*, which share parameter values with the verb *schmoren*, are represented in the distribution in this cluster. Therefore, even with a considerably higher error rate, this cluster illustrates semantically rather overlapping (synonymous) verbs, where *kochen* is the hyperonym of both *schmoren* and *dünsten*.

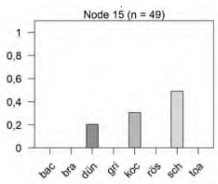
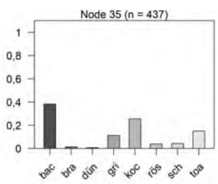
Node	Prediction	Weight	Error	Distribution	P_value	Node_cluster/Path
15	sch	49	51,02 04082	bac = 0, bra = 0, dün = 10, gri = 0, koc=15, rös = 0, sch = 24, toa = 0		Utensil (FORM, N_A_UT,Rost,TOPF), Heat source (Herd, N_A_HS, OFEN), Manner (kurz, leicht, N_A_MAN,RUEHREN, SEITE, WENDEN), RES (BISSFEST, GAR, glasig, GOLDGEL, N_A_RES, WEICH), Substance (BRUEHE, Tom_Mark, WEIN)
35	bac	437	61,78 4897	bac = 167, bra = 6, dün = 3, gri = 49, koc = 111, rös = 17, sch = 19, toa = 65		Substance (N_A_SUB), Heat intensity (MITTLERE_HITZE, N_A_HI), RES (GOLDGELB, N_A_RES), Utensil (N_A_UT), Manner (N_A_MAN) Heat source (N_A_HS)

Table 3.1. The third group of clusters with the highest error rate

It is less reasonable to consider the remaining 6 clusters of this group due to a deviance of more than 50% in the most frequent verb, resulting from combinations of certain parameter values that influence the distribution of the occurrences. However, it is worth noting some common features among the clusters. All of them (except for node 35 and 15 mentioned above) have a low number of occurrences, totaling only 86 examples. Many verbs share a part in the distribution of separate clusters due to insufficient data, as was the case with node 35 (Table 3.1).

At this point, it is also necessary to provide some clarifying information about the cluster under node 27, which results in a high error rate. There are just 10 overall occurrences in this node, with as few as four examples featuring the verb *toasten*, and this is due to several factors. First, we initially had only 100 annotated occurrences of the verb *toasten* extracted solely from *detenten13* because it was not featured in our *REZ_DE* corpus. Second, some annotated entries have been omitted as a result of quantity reduction of occurrences, based on varying frequencies according to their degree of relevance.[103] Last but not least, the absence of any values in all six parameters makes it difficult for the *CIT* model to suggest *toasten* as the most frequently used verb in context. As a result, verbs with no common semantic features with the verb *toasten* (e.g. *kochen*) also have a share in the distribution of this cluster.

In order to appropriately evaluate the *CIT* model, especially the generated error rate, the largest cluster under node 35, with 437 examples, was omitted with the intention to test the model by leaving out the examples that lack significant information in the annotation parameters. Even though the overall number of predictions was reduced from 1114 occurrences to 947, the proportion of correct predictions increased to 69% (Table 4), resulting in an error rate of 31%.

The description of the *CIT* model for the German culinary verbs in this chapter, which identifies the most frequent verb in clusters with certain parameter combinations, also revealed that in clusters with relatively higher error rates, the verbs sharing the distribution have either a formal lexical relation of synonymy or taxonomy (hypernyms and hyponyms), which could be also regarded as semantically synonymous verbs. Therefore, another *CIT* model (see Appendix 4.1) was developed, neutralizing the eight German verbs into synonymous groups, based on the hypothesis that the error rate was likely to be reduced because the incorrect predictions would remain within the synonymous group, thus avoiding severe mistakes in verb choice.

	CIT model	CIT model without the largest N_A cluster (node 35)	CIT model of synonymous verb groups	CIT model of synonymous verb groups without N_A cluster (node 30)
Annotated examples	1809	1372	1809	1292
Correct predictions	1114	947	1328	1048
Incorrect predictions	695	425	481	244
Correct predictions in %	61,50%	69%	73,40%	82%
Incorrect predictions in %	**38,50%**	**31%**	**26,60%**	**18%**

Table 4. *CIT* model of German culinary verbs vs. *CIT* model of synonymous verb groups

For instance, if we were to always take the most frequent verb in *deTenTen13* corpus, viz. *kochen*, which has an error rate of 48,70% (see Fig. 2) – a rate that is even lower than the incorrect predictions of some clusters of the *CIT* model – we would generate overtly incorrect

[103] For a detailed list of generalization and reduction of occurrences see Appendix 2 as well as Chapter 2.2 of this work.

verb choices since this approach of always giving preference to the most frequent verb *kochen* excludes any semantical annotation with no chance of other synonymous verbs to be used instead. For example, we might end up with *Kaffeebohnen kochen*** [104] instead of *Kaffeebohnen rösten*, which would be the correct choice in context. The *CIT* model with German verbs neutralized in synonymous verb groups was generally reduced the overall error rate from our initial *CIT* model down to 26,60%, even with the largest cluster of 517 occurrences with no information, annotated as N_As (Table 4). Excluding this cluster, the error rate dropped further to 18,8%. Besides, our approach excludes any serious semantic errors in the verb choice, even in clusters with considerably higher error rate. This is because the most frequent verb group identified by the *CIT* model generates an error rate that remains within the scope of semantically synonymous verbs, thus avoiding severely wrong verb choices. For instance, if we consider the following example from the *deTenTen13* corpus: *"Ein Kochtopf oder eine Pfanne ohne Öl auf der Herdplatte erhitzen und darin die Sesamsamen hellbraun rösten"*, with our semantic annotation of *Utensil*, *Substance*, *Resultative (adjectives)*, and *Heat source* parameters, almost identical to the cluster path under node 34 of the *CIT* model of synonymous verbs (see the table below), *grillen_rösten* is identified as the most frequent synonymous verb group with an error rate of 54,5%; however, even the other verb groups sharing the distribution under this node would not generate a severe semantical mistake, especially when compared to simply choosing the most frequent verb in the *deTenTen13*, viz. *kochen*.

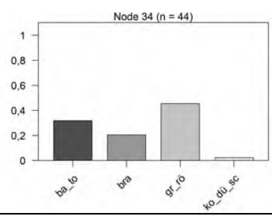
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
34	grillen_rö sen	44	54,545 4546	ba_to =14, bra = 9, gr_rö= 20, ko_dü_sc = 1		Substanc (N_A_SUB) Utensil (FORM, N_A_UT, TOPF), Heat_source (Herd, N_A_HS, OFEN), RES (BRAUN, KNUSPRIG)

Table 5. *CIT* model of German culinary synonymous verb groups

Thus, the CIT model of synonymous verb groups has allowed to reduce the number of actual incorrect predictions from 425 to 244 occurrences, excluding the annotated examples with no verbalized values in the relevant parameters. Consequently, the error rate of actual incorrect prediction of the model has been reduced to 13%. In conclusion, it could be stated that

[104] Two asterisks (**) here denote completely wrong verb choice

approximately 10% of the overall incorrect predictions (181 occurrences) in the initial *CIT* model are synonymous or semantically closer verbs (Fig. 4).

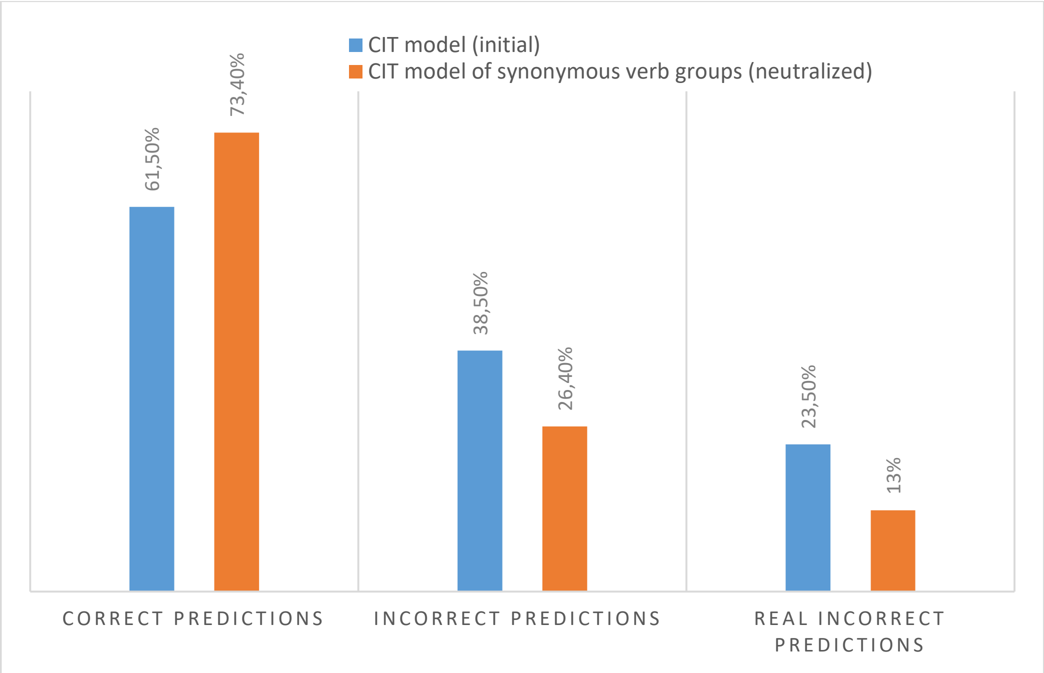


Fig. 4. *CIT* models in comparison with respect to *real*/*actual* incorrect predictions

Chapter 3. Description of the English Web 2015 (enTenTen15) and EN_REZ Corpora

Since no English parallel corpora to the German Web 2013 (*deTenTen13*) are available, it was required to find comparable ones in English which can serve as the basis for the intra- and interlinguistic analysis of the following corpus-based analysis. For this reason, among the available English corpora accessible via the *Sketch Engine* corpus manager and text analyzer software, the English Web 2015 (*enTenTen15*) with 15,411,682,875 tokens was selected as it was assessed as comparable to the *deTenTen13* corpus. Additionally, a manually compiled English corpus, i.e. the *EN_REZ* with 1,588,769 tokens, was also used for collocation extraction and concordancing as well as for annotation and further analysis. This approach was adopted with the dual objective of maintaining a comparable proportion with the German *REZ_DE* corpus and of elucidating potential differences between web culinary texts and cookbooks, whether in hard copy or digital format. In order to avoid redundancy, the criteria of corpus selection as well as the methodology of corpus annotation underlying the following study will not be elaborated here in detail since they were comprehensively and thoroughly described in Chapters 2.1 and 2.2.

The same functions were employed for the analysis of the English culinary texts as for the German corpora, namely *Word sketch* and *Concordance* functions in the *Sketch Engine* corpus query system. Chapter 2.1 illustrates all necessary information about the aforementioned functions and formulae in detail together with an exposition of the underlying calculations. It is nevertheless important to restate that *Word sketch* (Fig. 1) made the quantitative part of the analysis for the two English corpora mentioned above possible by showing in one page the summary of the grammatical and collocational behavior of the given word (see Killgariff et al. 2010). The function *Concordance* allowed to extract occurrences (examples) within contexts which were subsequently manually annotated, contributing to the qualitative analysis of the gathered data. Thus, the two functions contributed to the identification of the collocations for six English culinary verbs, viz. *bake*, *braise*, *cook*, *fry*, *roast*, and *sauté*.

Fig. 1. Partial *Word sketch* of the English culinary verb *sauté* in Sketch Engine

As it was explained in Chapter 2.1, the direct objects' column of collocations was taken into consideration in the following analysis, given that the nature of the culinary texts in question implies that an ingredient or a dish is being prepared throughout the recipe or text. The partial *Word sketch* above (Fig. 1) also illustrates, in addition to the collocations of the verb *sauté*, the number of occurrences (the frequency) of the given verb with a specific collocation in the given corpus as well as the degree of association between these two words (measured by the *logDice* score).[106] As illustrated in the partial *Word sketch* (Fig. 1), the collocations of the English culinary verb *sauté* are ordered according to the association score l*ogDice*, descending from highest to lowest. The strongest collocation is the word *shallot*, which occurs 43 times in the overall 2,931 entries for the verb *sauté* in the *enTenTen15* corpus. Nevertheless, despite the automatic pre-selection through the *Word sketch* function, a manual annotation of all concordances would have been unfeasible due to the high number of occurrences, it would have been practically unfeasible to manually annotate all concordances for all six English culinary verbs. Therefore, the same criterion of collocation reduction above a *logDice* score of 5.0 was

[106] See Chapter 2.1 of this study as well as (Rychlý 2008) and (Evert 2005) for a more detailed description on different association scores between two neighboring words, particularly on l*ogDice*.

applied as for the German corpora. Chapter 2.2 also elaborates on why the *logDice* score 5.0 in particular is the most relevant for reducing the number of collocations. The lower the collocation association score (*logDice*) gets, the less strong the collocations are. This strategy proved successful to reduce the English collocation lists, too. Either hyponyms or meronyms of a much stronger collocation are included in the list of collocations, or another spelling of one of the same collocation candidates, or merely misspellings.

However, even with the aforementioned reduction, it would have been impossible to annotate all occurrences due to their volume. Therefore, the model used for the German corpora, namely of determining the distribution of the annotated examples proportionally in relation to the collocations, so that the ones with higher association scores get more annotated occurrences, was applied to the English data as well. This model allowed to get a representative image of the corpora being annotated. Because of the considerable differences in the overall tokens of *enTenTen15* and the *EN_REZ* corpora with around 15 billion and 1.5 million tokens respectively, the annotation of the extracted concordances was carried out with differences in their quantity of examples. More precisely, 200 occurrences for each English culinary verb were annotated in relation to their collocations generated with *Word sketch* to represent the *enTenTen15* corpus, while for the *EN_REZ* one, 100 occurrences (for each verb) were considered sufficient; primarily because of the insufficiency of the number of available examples in the *EN_REZ* corpus due to its size. It is worth noting, however, that while the collocations for the *enTenTen15* were reduced by including only those with a *logDice* association score above 5.0, as mentioned above, for the *EN_REZ* corpus all collocations and therefore their concordances were considered for annotation irrespective of their *logDice* score. Nonetheless, for some verbs, for instance *braise*, even with the aforementioned strategies, the necessary 100 occurrences could not be annotated as the number of examples available in the *EN_REZ* corpus was lower than that. Instead, only the available 37 occurrences with the verb *braise* in the *EN_REZ* corpus were annotated.[107]

However, like in the case of the German culinary verbs and their collocations, it was still difficult to determine the distribution of the 200 or 100 occurrences in relation to the collocates and the node (the specific culinary verb) to ensure that the ones with higher *logDice* score are represented by more examples for further annotation while those with lower scores are proportionally represented by fewer occurrences. Thus, the same annotation quantity

[107] For the manual semantic annotation of the English culinary verbs for both the *enTenTen15* and *EN_REZ* corpora in *.xlsl* format see Appendix 5.

determination model, developed for the German culinary verbs annotation, was successfully applied also for the English verbs. The model was introduced and elaborated in detail in Chapter 2.2. However, the following example of the annotation quantity model briefly illustrates the distribution of the collocations and their concordance lines in relation to the context of the verb *sauté* in the *enTenTen15* corpus (Fig.2). The initial *logDice* association score, which indicates the collocational behavior of the node (the culinary verb) and its collocates, was converted into *Dice*, another measurement of association strength between two neighboring words (Evert 2005:200) to best illustrate the real differences between the collocates of the same node. Then a 2D-line diagram, representing quantitatively the distribution of collocates in relation to the verb context in the whole *enTenTen15* corpus, was generated followed by a trend line to best match the 2D-line diagram. Based on the formula of the given trend line, a model was developed to determine the distribution of the collocates in the given English sub-corpus. For instance, in the case of all the collocates in relation to the context of the verb *sauté*, the model resulted in the selection of 32 occurrences of the collocate *onion* in the 200 examples of the sub-corpus to represent the 394 occurrences of the whole *enTenTen15* one (see Fig.2).

Initially, two types of corpora were annotated to also reveal the differences between the collocational behavior in the culinary texts in the web corpus, where everyone may create content, and the one based on cookbooks available both as hard copies and in digital format. However, upon the completion of the annotations carried out separately for the two corpora, it became evident that the results do not differ in a statistically significant way. The comparison of the English *enTenTen15* and the *EN_REZ* was carried out with the same analysis method as the German *deTenTen13* and the *REZ_DE*, i.e. the *Mosaic plot* visualization and showed no statistically significant difference between the two aforementioned corpora in the parameters *Resultative* (adjectives) as well as *Utensil* and slight overrepresentations of just one parameter value in the *Manner*, *Substance*, *Heat intensity* and *Heat source* parameters (see Appendix 6 for the respective *Mosaic plots*).[108]

Therefore, the annotated examples of both the *enTenTen15* and the *EN_REZ* were then compiled into a single document, in which the corpus origin of each occurrence is provided.

---

[108] The Mosaic-Plots comparing the *enTenTen15* and *the EN_REZ* corpora in terms of their distribution regarding the annotation parameters were generated with 2:1 ration, viz. for each English culinary verb twice as many examples have been annotated in the *enTenTen15* corpus as in *EN_REZ* one. For instance, for the verb *braise* 34 annotated examples of the *enTenTen15* and proportionally 68 of the *EN_REZ* corpus served as basis for generating the *Mosaic-Plots*.

Word	Freq	logDice*1000	Number	Dice	Model	Designation in 200	Rounded	Zero→1
shallot	43	9.060	1	0,032577055	0,0419	71,58043511	72	72
onion	394	8.610	2	0,0238478	0,018685902	31,92231576	32	32
leek	29	8.350	3	0,01991501	0,011651122	19,90435332	20	20
garlic	69	7.010	4	0,00786684	0,008333245	14,23621191	14	14
veggies	27	6.200	5	0,004487103	0,006425604	10,9772683	11	11
clove	10	6.180	6	0,004425328	0,005195984	8,876630194	9	9
mushroom	53	6.080	7	0,004128977	0,00434185	7,417458367	7	7
celery	6	5.880	8	0,003594483	0,003716329	6,348841707	6	6
kale	8	5.760	9	0,003307599	0,003239825	5,534798448	6	6
zucchini	5	5.540	10	0,00283979	0,00286559	4,895469329	5	5
spinach	11	5.470	11	0,002705292	0,002564434	4,380985904	4	4
eggplant	6	5.260	12	0,002338826	0,002317223	3,958659813	4	4
shrimp	14	5.070	13	0,002050228	0,002110911	3,606204314	4	4
vegetable	90	5.010	14	0,00196671	0,00193631	3,30792133	3	3
bacon	11	5.010	15	0,00196671	0,001786766	3,052446193	3	3
								200

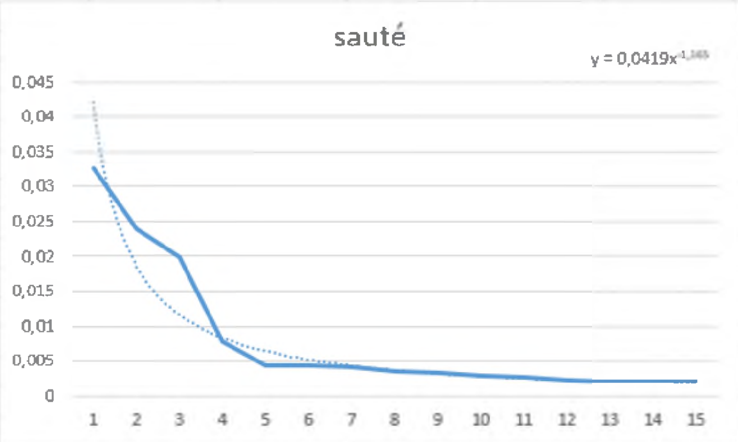


Fig. 2. Annotation quantity determination model for the verb *sauté* in the *enTenTen15* corpus

3.1 Neutralization (Generalization) of Annotation Parameters and Reduction of Annotated Occurrences

Sentences	Verb	Ingredient	Substance	Dish	Utensil	Heat intensity	Manner	RES	Heat source	Corpus
Click here to order "So, you want a wood-fired oven to bake bread and have pizza parties?	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	wood-fire ove	preloaded/ententen15_
Free range meat, game, poultry and eggs; locally grown fruit and vegetables, artisan baked bread , cakes and pastries; cheese, wine, pickles and preserves and even local [...]	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
Bake the bread for 60 minutes.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
All ovens are different, so be careful the bottom doesn't burn, but be sure that bread is baked properly.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
Patisserie, San Fransciso Baking Institute and Della Fattoria where they bake their breads in wood-fired ovens in a middle of a farm in Petaluma, California.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	wood-fire ove	preloaded/ententen15_
There will be harvesting, cooking, baking bread from locally grown heritage wheat, sharing recipes, hearing food stories and connecting with what we eat and where it comes from.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
We shall not bake all bread in one oven.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
That's why at Panera Bread bakery-cafes, bakers spend their nights baking breads in stone deck and convection ovens before we open our doors each morning.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
And I am the one to bake bread , or toffee, or cookies, or all of them.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
We bake fresh artisan breads and exquisite European-style pastries every day.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
Now he must build an oven for warming and baking bread .	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
The first loaves of bread baked in the newly completed oven Tidings from the 18th Century (p 205, 1993)	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
Then dough is made and bread is baked .	bake	dough	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
Students learn to bake traditional Irish brown bread and scones and partake in the traditional Irish skill of butter making.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
We purchased flour and meal, baked our own bread , and cooked our own food, generally, which was good, though sometimes scanty; and sometimes we had johnny-cake, or corndodger, instead of flour bread.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
Steam and the odour of fresh baked bread broke the spell and I hopped off the stool and sped back up to Francesca, at once invigorated and ashamed.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_
We shall not bake all bread in one oven.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
That's why at Panera Bread bakery-cafes, bakers spend their nights baking breads in stone deck and convection ovens before we open our doors each morning.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	oven	preloaded/ententen15_
And I am the one to bake bread , or toffee, or cookies, or all of them.	bake	N_A	N_A	bread	N_A	N_A	N_A	N_A	N_A	preloaded/ententen15_

Table 1. Annotation of the *EN_REZ* corpus concordances (occurrences) of the verb *bake*

After the number of occurrences to be annotated based on the previously introduced model (Fig.2) was determined, the examples extracted from both the *enTenTen15* and *EN_REZ* corpora were annotated according to the eight parameters presented in Table 1. Besides the columns of the verb itself and the annotation parameters, Table 1 also includes the extracted concordance lines in the first column and the name of the specific corpus in the last column.

Since the annotation parameters were introduced in detail in Chapter 2.1., they are not discussed further here. For the English *intra*- and *inter*linguistic analysis six of the eight initial annotation parameters were also taken into consideration, namely *Substance*, *Utensil*, *Manner*, *Heat intensity*, *Resultative* (adjectives), and *Heat source*. The parameters *Ingredient* and *Dish* were left out due to a very high number of different values resulting in no significant correlations. Besides, with the attempt to generalize the values of the parameter *Dish*, most of them at some point of the generalization process would become an *Ingredient*. For instance, the German dish 'Bratkartoffel' (roast potato) would be have generalized as 'Kartoffel' (potato) coinciding with the *Ingredient* parameter value 'Kartoffel', thus proving the irrelevance of the parameter *Dish*. Still, each of the remaining six annotation parameters included vastly different values to the point that statistically no significant results could have been obtained without neutralization of some of these. For instance, the parameter *Substance* has values such as *olive oil*, *oil*, *sunflower oil*, *coconut oil*, *palm oil*, and *canola oil*, which are connected with each other in a taxonomic relation of co-hyponymy. Therefore, these values were generalized (neutralized) with their hypernym *oil.*[109] The generalization allowed to reduce the types of values and resulted in fewer types, which led to more informative results, while at the same time maintaining generally the semantics of the original content. However, the generalization could not entirely solve the issue of having too many different variables in small numbers that not only do not contribute to the analysis of the English cooking collocations but rather obstruct the analysis. Therefore, the English annotated examples were also reduced by excluding the values with very low occurrences from the statistical analysis.[110]

As a result, the initially annotated 1683 occurrences were reduced to 1273, which serve as the basis for all statistical analyses in this chapter.

3.2. *Correspondence Analysis* (CA) vs. *Multiple-Correspondence Analysis* (MCA) of the English Culinary Verbs

Correspondence analysis (CA) and *Multiple Correspondence analysis* (MCA) contributed to the visualization of the distribution of the English culinary verbs with regard to relevant parameters. As mentioned above, the *enTenTen15* and *EN_REZ* corpora were separately annotated according to eight parameters (not counting the verb itself), i.e. *Ingredient, Substance, Dish, Utensil, Heat intensity, Manner, Results,* and *Heat source*.

Chapter 2.3. of this research reports extensively on the details of both CA and MCA based on the German culinary verbs and the same eight annotation parameters. Therefore, details on (M)CA analysis are omitted here. It is worth highlighting, however, that while CA showcases the correlations between the dependent (the verbs) and the independent (the parameter values) variables, MCA first illustrates the correlations among the independent ones, and in the following a verb or verbs (the dependent variables) are distributed around the identified cluster as qualitative supplementary variables, which were not taken into consideration for the analysis as such, but are rather placed on the plot a posteriori.

---

[109] The list of all generalizations is provided in Appendix 7.

[110] The same criteria served as basis for the reduction of the annotated parameter values for the English annotated examples as for the German and Armenian occurrences. On one side, the considerable difference in the numbers of the preceding and succeeding collocates was taken into account, on the other side, the maximal coverage for all the frequencies was considered, both ensuring minimal loss for the further analysis and their visualization. For instance, the *Manner* parameter was reduced to the value {gently} as the difference in frequencies between {gently} and the succeeding collocate {deep} is 11 annotated occurrences. The parameter *Resultative* (adjectives), for example, was reduced to the value {done} covering still 97% of the annotated occurrences. Both criteria of parameter reduction ensure at least around 97%-98% coverage of the initial annotated occurrences, for the parameters *Heat intensity*, *Substance*, *Utensil*, *Heat source* up to 99%.

In both CA and MCA, the strength of correlations is identified by both the distance of the dependent and independent variables from the intersection as well as the angle between them. In detail, for statistically relevant correlations, the dependent and independent variables should be far from the intersection. The sharper the angle is between the independent and dependent variables, the stronger correlated they are.

1,273 contextualized examples resulted from the generalization and reduction of the parameters of the initially annotated 1,683 occurrences (see Chapter 3.1.). They served as the basis for both the CA and the MCA analysis.

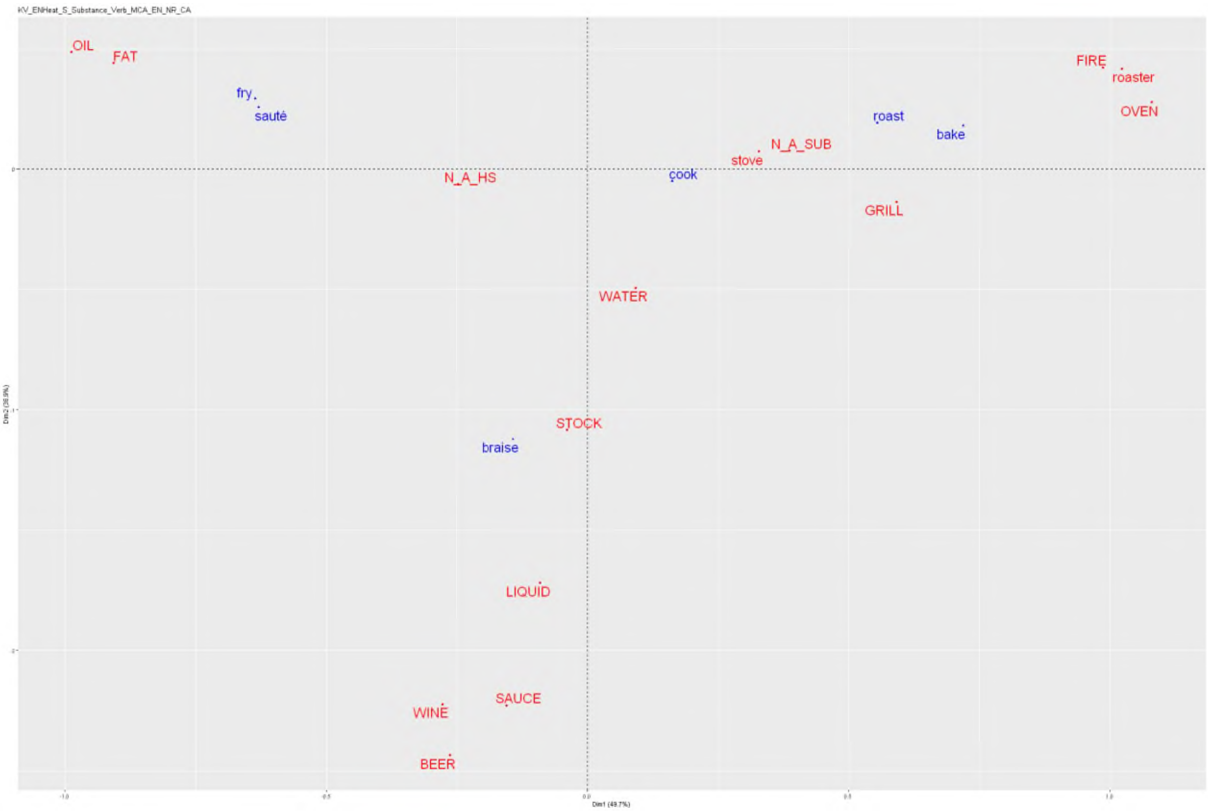


Fig.1. CA: English culinary verbs (parameters *Substance* and *Heat source*)

The *CA* graphical visualization (Fig.1) illustrates strong correlations between the dependent variables (verbs marked in blue) and the independent ones (parameter values marked in red). {Oil} and {fat}[111] as examples of the *Substance* parameter are strongly correlated with the English culinary verbs *sauté* and *fry*. However, judging from the distance of the aforementioned verbs and parameter values from the intersection as well as from the sharpness of angle they

[111] *Substance* parameter value {fat} was generalized and includes values such as *ghee, lard* and *butter*. For all English parameter generalizations see Appendix 7.

form when drawing an imaginary line between them, both {oil} and {fat} are correlated with the English culinary verbs *fry* and *sauté* as well as {N_A_HS} of the *Heat source* parameter. {N_A_HS} is the annotation category for the non-verbalized values in this parameter, which "pulls" the verbs *fry* and *sauté* down to the X axis of the CA graph (Fig. 1; examples 1-7). The first dimension of this CA graph (Dim1) illustrates 49.7%, while the second one (Dim2) 35.9% of variance in the data, resulting in overall 85,6% representation of the "explained inertia" (terminology adopted from Jenset and McGillivray (2017)). The distribution of the *Substance* parameter values {oil} and {fat} are almost identical both for the English verbs *fry* and *sauté*, resulting in nearly the same correlation as they are both simultaneously correlated with the {N_A_HS} of the *Heat source* parameter mentioned above. However, since for the *CA* visualization technique the number of the annotated examples also plays a decisive role in the positioning of the variables on this two-dimensional scatterplot, the *Substance* parameter value {oil} and the verb *fry* are placed a bit further in the upper-left corner compared to the verb *sauté* and {fat}. Thus, in the English sub-corpus, overall 95 co-occurrences of *fry* and {oil} as well as 30 ones of *fry* and {fat} were annotated with a slight difference in distribution for the verb *sauté*, i.e. 113 co-occurrences of *sauté* and {oil}, and 44 ones with *sauté* and {fat}. Summing up, it can be stated that this group of correlation is gathered around some type of *oil*-like, *fat*-like cooking substance.

1. Gently *fry* the onion in *butter* in a large saucepan. *(EN_REZ)*
2. *Fry* the meat in the *ghee*, stirring occasionally until brown all over. *(EN_REZ)*
3. Quickly *fry* the prawns in the *butter* and *garlic oil* until just cooked. *(enTenTen15)*
4. Heat the dry wok, add the *oil* in a stream and *sauté* the shallots for 1 minute. *(enTenTen15)*
5. *Sauté* the vegetables in the *olive oil*. *(EN_REZ)*
6. Add a good lug of *oil* to the frying pan and *fry* the spring onion and garlic over a medium-low heat until softened and golden. *(EN_REZ)*
7. Heat the *oil* in a large pan and *fry* the onions until soft. *(enTenTen15)*

The *CA* graph (Fig. 1) identifies another group of strong correlations, i.e. between the English culinary verb *braise* and several values of the parameter *Substance* denoting some type of *liquid*-like cooking substance. For instance, the verb *braise* correlates with the *Substance* parameter values {wine, sauce, liquid[112], beer, stock}, while being simultaneously "pulled" up from the {N_A_HS} value of the *Heat source* parameter (examples 8-13).

8. Add mirepoix to the hotel pan (including apple and fennel) and add pork *stock* as needed to *braise* cheeks. *(enTenTen15)*
9. *Braise* the beef brisket in *beer* for two hours. *(enTenTen15)*
10. Pour in 1 cup of *broth*, cover the pan, and *braise* the squash for 10 minutes until partially cooked. *(EN_REZ)*
11. Cover with water, and *braise* on the lowest possible temperature for 3 1/2- 4 hours (results are best when the pork is chilled in the *braising liquid*, then reheated). *(enTenTen15)*
12. The beef *braised* in red *wine* melts in your mouth and the owner hand-makes the selection of charcuteries. *(enTenTen15)*
13. A common combination might be soy *sauce braised* pork, Chinese greens and a tea. *(enTenTen15)*

What is interesting is the effect of the *Substance* parameter value {water}: it simultaneously "pulls" the verbs *braise* as well as *cook*. As *braise* also correlates with other *liquid*-like cooking substances, e.g. {stock} and *cook* correlates with {N_A_SUB} in addition to its strong correlation with the *Heat source* parameter {stove}, {water} remains just between the two verbs (examples 14-15).

14. *Cook* the rice in *boiling salted water* for 5 minutes. *(enTenTen15)*

15. After *cooking* the soup on the *stove*, the children served the soup to their friends at our dining table. *(enTenTen15)*

The last group of strong correlations identified in this CA graph (Fig. 1) is determined by the absence or non-verbalization of any cooking substance and specific values of the *Heat source* parameter. In this group, the English culinary verb *bake* correlates with {oven}x{N_A_SUB}

[112] In order to avoid too much generalization of the *Substance* parameter values with the objective to keep the verb-specific values, such as {beer, stock, wine, water}, the latter were not generalized as {liquid}. {Liquid} here comprises a small group of *Substance* parameter values such as *syrup, milk, gravy,* and *braising liquid.*

(examples 16-17), while *roast* correlates with {fire, roaster}x{N_A_SUB} (examples 18-20) of the *Heat source* and *Substance* parameters respectively.

16. *Bake* in a pre-heated 350degree *oven* 40 minutes, then reduce heat to 300 degrees and bake 35 to 40 minutes more, or until cake tests done. *(EN_REZ)*

17. *Bake* the tart for 30 minutes in the 325F *oven* or until golden on the edges. *(enTenTen15)*

18. The families also bonded while rafting, hiking, biking, and *roasting* marshmallows over the *campfire*. *(enTenTen15)*

19. We also had a *campfire* and *roasted* marshmallows and put them on chocolate, peanut butter cups or York peppermint patties and graham crackers. *(enTenTen15)*

20. Something you could try as a substitute would normally be home *roasted* chocolate malt in a wok or nut *roaster*. *(enTenTen15)*

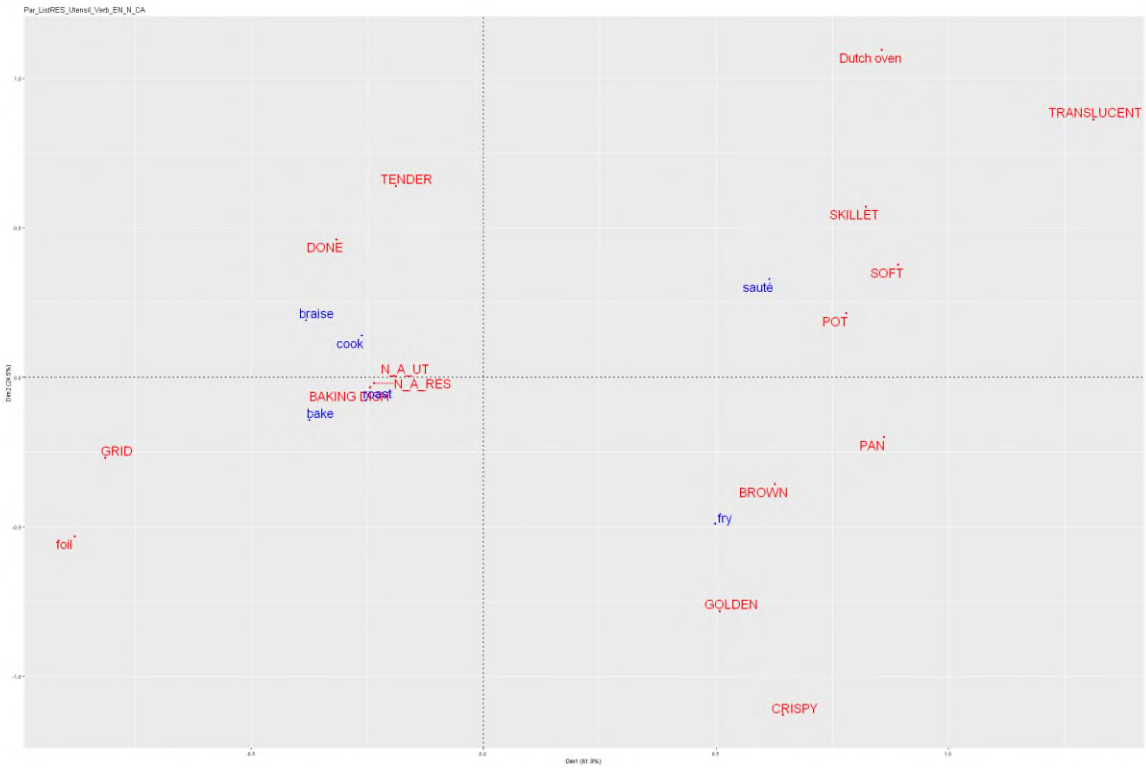


Fig. 2a. CA: English culinary verbs (parameters *Resultative* (adjectives) and *Utensil*))

A number of clear-cut strong correlations between the English culinary verbs and certain values of the *Resultative* (adjectives) and *Utensil* parameters are observed in Fig.2a. This CA two-dimensional visualization graph illustrates strong correlations between the verbs *sauté* and *fry* on one side while grouping shared correlations of the verbs *braise* and *cook* on the upper-left as well as those of the verbs *bake* and *roast* on the lower-left sides of this scatterplot. The {golden}, {brown}, and {crispy} values of the *Resultative* (adjectives) parameter as well as {pan} are displayed in the CA as correlating strongly with the English culinary verb *fry* (examples 21-23).

21. *Fry* the bacon in a dry pan until *crispy*. *(enTenTen15)*

22. Using a medium-size heavy-based non-stick *frying pan*, *fry* the bacon in a tablespoon oil until *crisp*. *(enTenTen15)*

23. Fry the meat in the ghee, stirring occasionally until brown all over. *(EN_REZ)*

The verb *sauté* is indicated to strongly correlate with the *Resultative* (adjectives) parameter values {translucent, soft} as well as the *Utensil* parameter ones {skillet, pot, Dutch oven} (examples 24-29).

24. *Sauté* the onion, salt, pepper, red pepper flakes, and zest, stirring occasionally, until the onion is *translucent*. *(EN_REZ)*

25. Heat the olive oil in a *skillet* over medium high heat and *sauté* the shallots until they are *soft* and *translucent*, about five minutes. *(enTenTen15)*

26. *Sauté* the onion and garlic in oil in large soup pot until *soft*. *(enTenTen15)*

27. In a large *skillet*, heat olive oil over medium heat and *sauté* the leeks or scallions until just *softened*, then add the garlic and *sauté* another 30 seconds or so. *(enTenTen15)*

28. In a *Dutch oven*, *sauté* the onions and garlic in the oil until *soft*. *(EN_REZ)*

29. Heat remaining oil in the same *pan* and sauté the meat until *brown*. *(EN_REZ)*

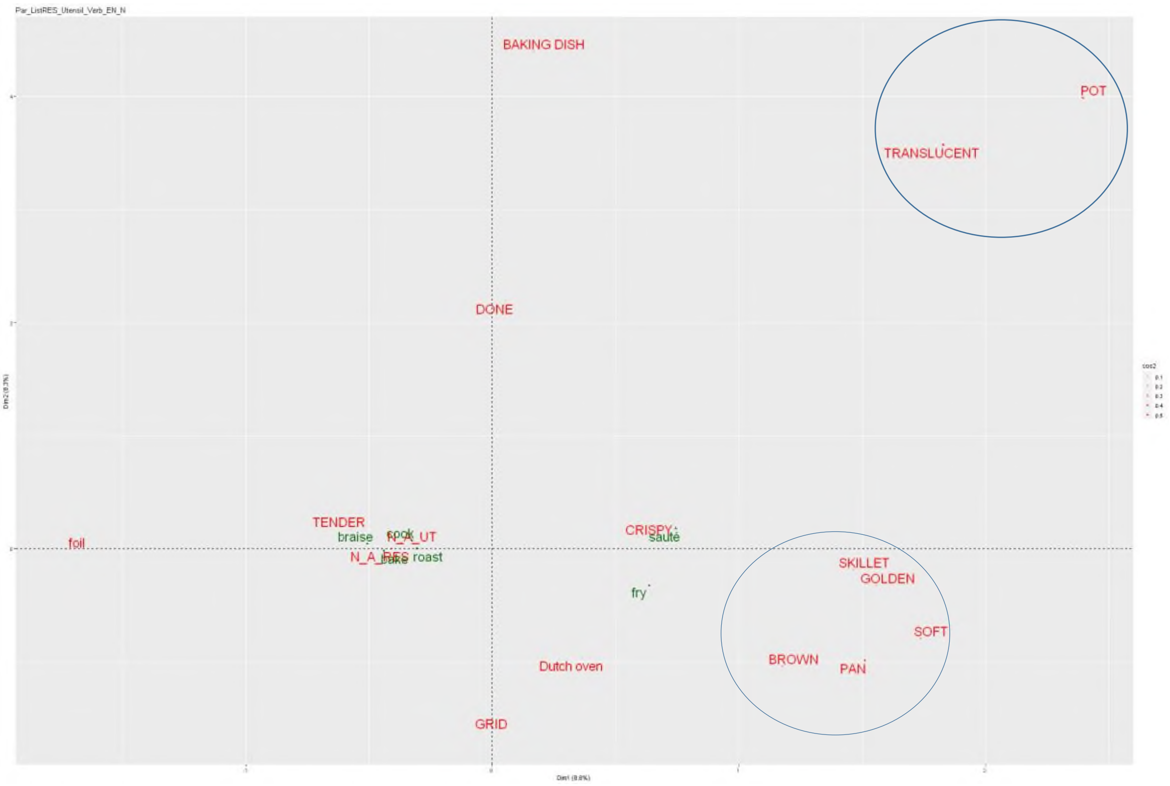


Fig. 2b. MCA: English culinary verbs (parameters *Resultative* (adjectives) and *Utensil*)

The *MCA* (Fig. 2b) of the same data, on the contrary, demonstrates a cluster of parameter values (independent variables) identifying either their co-occurrence in the same contexts, viz. strong correlation, or their profile similarities, viz. similar syntagmatic distribution. For instance, {brown, golden, crispy, skillet, pan} form a cluster of strongly correlated independent variables (examples 30-31).

30. In a medium-sized *skillet*, melt the fat and *sauté* the onions until limp and *golden brown*. *(enTenTen15)*
31. Heat a *pan* with olive oil and *fry* the prawns seen until *golden brown*. *(enTenTen15)*

When analyzing the distance of the *Utensil* parameter value {pot} and the *Resultative* (adjectives) one {translucent}, it becomes clear that they form a cluster of strongly correlated independent variables, which expresses their co-occurrences in the same context even without the distribution of the dependent variables, viz. the verbs.
Another relatively large cluster is gathered around the non-verbalized values in the *Utensil* and *Resultative* (adjectives) parameter values of {N_A_UT} und {N_A_Res, tender} as well as the English culinary verbs *bake*, *braise*, *roast*, and *cook* (Fig. 2b). However, this cluster is

generated primarily due to the absence of data in form of {N_A} in both aforementioned parameters and is located relatively close to the intersection, demonstrating a rather neutral correlation between them. {tender} is included in this cluster due to its co-occurrence with the *Utensil* parameter value {N_A_UT} (example 32-33).

32. Chicken should be *tender*, and *cooked* through. *(enTenTen15)*

33. Meanwhile toss 1 tablespoon salt into the pasta water and *cook* the linguine until just *tender*. *(EN_REZ)*

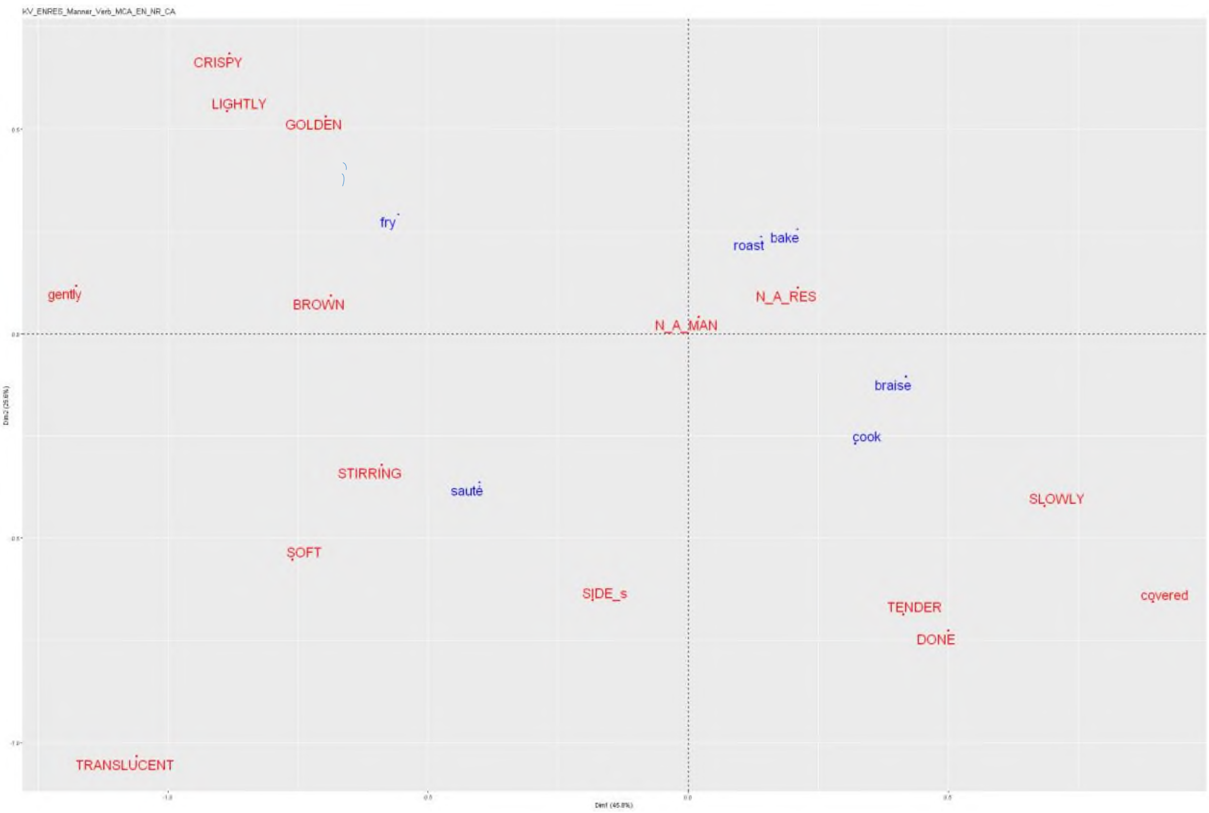


Fig. 3. CA: English culinary verbs (parameters *Resultative* (adjectives) and *Manner*)

The distribution of the *Resultative* (adjectives) and *Manner* parameter values in relation to the six English culinary verbs in the framework of CA and MCA visualization techniques is almost identical, therefore only the CA is presented in detail.
Fig. 3 (CA) illustrates several strong correlations between the English culinary verbs and values of the parameters *Resultative* and *Manner*. As mentioned at the beginning of this chapter, two factors are taken into account when judging whether the correlation is strong or not. First, the distance of the dependent variable (the verbs) and the independent ones (the parameter values)

from the intersection; distributions close to the intersection are statistically not relevant. Second, the angle between the aforementioned variables when drawing imaginary lines to the intersection point. The sharper the angle is, the stronger correlated they are. For instance, the English culinary verb *saut*é is demonstrated to strongly correlate with the *Manner* parameter value {stirring, side_s} (examples 34-36) as well as the *Resultative* (adjectives) parameter {soft}.

34. Continue to *sauté* the onions, *stirring* occasionally, until the wheat berries start to become tender and the onions begin to caramelize. *(enTenTen15)*

35. *Sauté* the fish, 2-3 minutes *on each side*. *(EN_REZ)*

36. With the pan still on medium heat, *sauté* the shallots for 30 seconds or so, *stirring* constantly. *(enTenTen15)*

The English verb *saut*é is also strongly correlated with another value of the *Resultative* parameter value, i.e. {translucent}, as the latter is situated the furthest from the intersection and "pulls" *saut*é down the most, which expresses the high number of annotated co-occurrences of *saut*é and {translucent}. The correlation is so strong that they form a line (example 37).

37. *Sauté* the onion 3 minutes or until it starts to turn *translucent*. *(enTenTen15)*

The English culinary verb *fry* creates a strong correlation with the *Manner* parameter values {lightly, gently} as well as {crispy, golden} of the *Resultative* (adjectives) parameter. However, these two values are located rather near the horizontal axis of the graph since they are "pulled" down by another culinary verb *saut*é. The verb *saut*é co-occurs 32 times with the *Resultative* parameter value {brown}, while the verb *fry* does so 42 times. In the case of the *Manner* parameter value {gently} (examples 38-40), seven contextual examples with *saut*é and 13 with *fry* were identified in our English sub corpus.

38. *Gently fry* the leek and celery for 5 minutes in the butter. *(enTenTen15)*

39. In the same oil, *gently sauté* the onion and garlic for the sauce. *(EN_REZ)*

40. Meanwhile, heat the olive oil in a small pan and *gently sauté* the garlic cloves, paprika and the whole chili. *(enTenTen15)*

The English verb *cook* correlates strongly with the values {tender} and {done} as well as {covered} and {slowly} of the *Resultative* and *Manner* parameters respectively (examples 41-43). The verb *braise* is strongly correlated with the *Manner* parameter values {slowly} and {covered} (examples 44-46).

41. Bring to a simmer, *cover* and *cook* for 1,5 hours, or until the meat is *tender*. *(EN_REZ)*

42. Make a syrup of one cupful each of sugar and water and *cook* the apples in it very *slowly* until tender. *(enTenTen15)*

43. Stir in the potatoes and garam masala, *cover* and *cook* for a further 10 minutes or until the potatoes are *tender* and the meat fully cooked. *(enTenTen15)*

44. *Slowly braised* beef brisket in red wine and rosemary is the ultimate special occasion dinner. *(enTenTen15)*

45. *Cover* the Dutch oven and *braise* the beef brisket for 45 more minutes. *(enTenTen15)*

46. *Cover* pot, turn on LOW and *cook* 4-6 hours, or until chicken is tender. *(EN_REZ)*

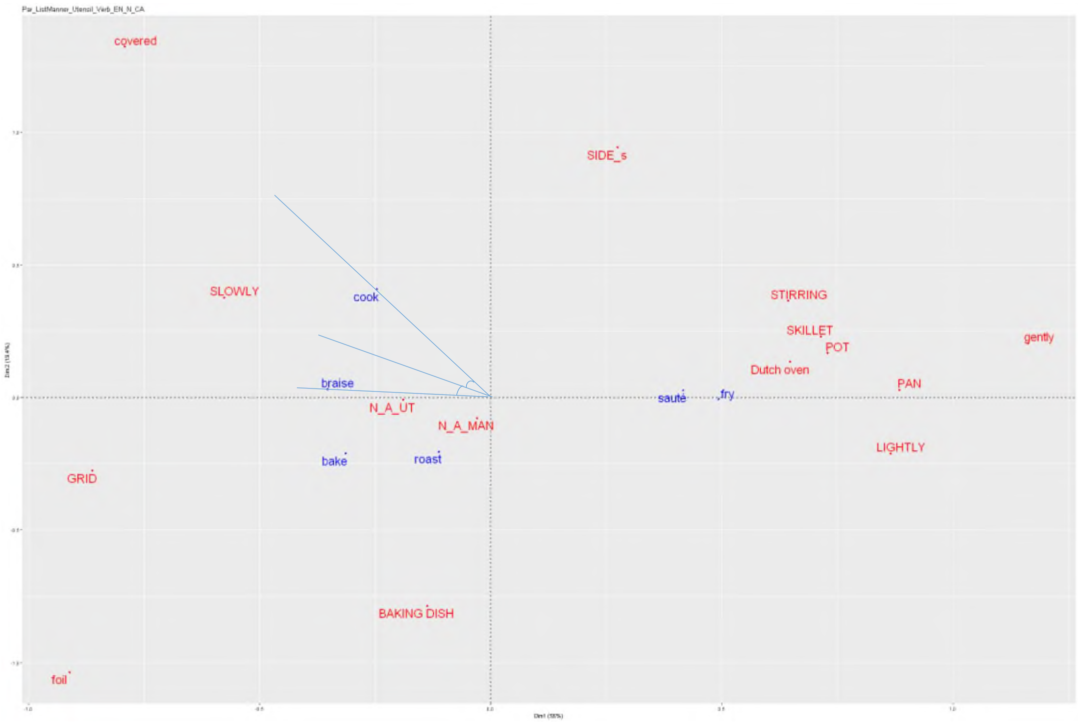


Fig. 4a. CA: English culinary verbs (parameters *Manner* and *Utensil*)

Among the strong correlations demonstrated on the CA visualization graph (Fig. 4a), the verbs *sauté* and *fry* create almost identically strong correlations with the *Utensil* parameter values {pan, skillet, pot, Dutch oven} as well as the {lightly, stirring, side_s,[113] gently} values of the *Manner* parameter. The CA graph illustrates strong correlations between the English culinary verb *cook* and the *Manner* parameter values {slowly, covered} with the latter being the strongest one judging both from its distance from the intersection as well as drawing an imaginary a line with *cook*. The aforementioned parameter values correlate also with the verb *braise*, however, if {covered} "pulls" *cook* to the upper-left hand corner, the {N_A_UT} value of the *Utensil* parameter draws *braise* down, close to the horizontal axis of this two-dimensional diagram and creates a less sharp angle, therefore indicating a less strong correlation of *braise* and {covered} compared to *cook* and {covered}. The English culinary verb *bake* correlates with the *Utensil* parameter value {foil} and {grid}, while {baking dish} simultaneously "pulls" both the verb *bake* and *roast* closer to its location.

Fig. 4b. CA: English culinary verbs with concatenated *Manner* and *Utensil* parameters

The CA with concatenated parameter values (Fig. 4b) introduces another form of data visualization where strong correlations between specific parameter combinations (independent variables) and the English culinary verbs (dependent variables) are illustrated. The following

[113] Neutralized parameter values are written in capitals to avoid unnecessary generalizations as the programming language *R*, which has been used to generate among others the (M)CA graphs, is case sensitive.

comparison of the first CA and the CA with concatenated parameter values illustrates the differences of correlations based on the analysis of the same data from another perspective. On the CA graph with concatenated parameter values (Fig. 4b), *cook* is displayed to correlate strongly with combinations of different *Utensil* and *Manner* parameter values, i.e. with {pot}x{covered}, {skillet}x{side_s} as well as {N_A_UT}x{covered} and {N_A_UT}x{slowly}. Here, for instance, the *Manner* parameter value {side_s} as well as the *Utensil* parameter values {pot, skillet, N_A_UT} identified as strong correlations with the verbs *saut*é, *braise* and *bake* on the CA graph above, are also included as parameter combinations to correlate strongly with the English verb *cook*. Both the first CA graph (Fig. 4a) and the second CA with concatenated parameter values (Fig. 4b) above demonstrate almost the same positioning of two dependent variables, i.e. the verbs *saut*é and *fry* (examples 47-49).

47. Next, *lightly sauté* the leeks and garlic in a *sauté pan* over medium-low heat for just 1-2 minutes. *(enTenTen15)*

48. Melt the butter in a *frying pan*, lay in the cutlets, which should be smartly trimmed, and *fry lightly* till cooked. *(EN_REZ)*

49. Heat about 1 tsp of oil in a *pan* and *lightly fry* the eggs for 1-2 minutes and remove. *(enTenTen15)*

The CA graph with concatenated parameter values (Fig. 4b) illustrates, on one hand, the aforementioned idea of overlapping correlations between *saut*é and *fry* and, on the other hand, sets more or less specific parameter combinations for each verb. For instance, it exposes distinctive strong correlations between the two verbs with their respective *Utensil* and *Manner* parameter values, e.g. the verb *fry* correlating with {pan}+{lightly} while *sauté* with {skillet}+{stirring} (cf. examples 50 and 51).

50. Melt the butter in a frying *pan*, lay in the cutlets, which should be smartly trimmed, and *fry* lightly till cooked. (EN_REZ)

51. In a 7- or 8-inch skillet, *sauté* the onions in the oil over moderate heat, *stirring*, until very soft, 7 to 8 minutes. (EN_REZ)

The verb *roast* is identified on the CA with concatenated parameter values graph to correlate with such values of the aforementioned two parameters not illustrated on the CA graph (cf. Fig. 4a and 4b), for instance, {skillet}+{lightly} (examples 47).

52. *Slightly roast* the pistachios in a *skillet*. *(enTenTen15)*

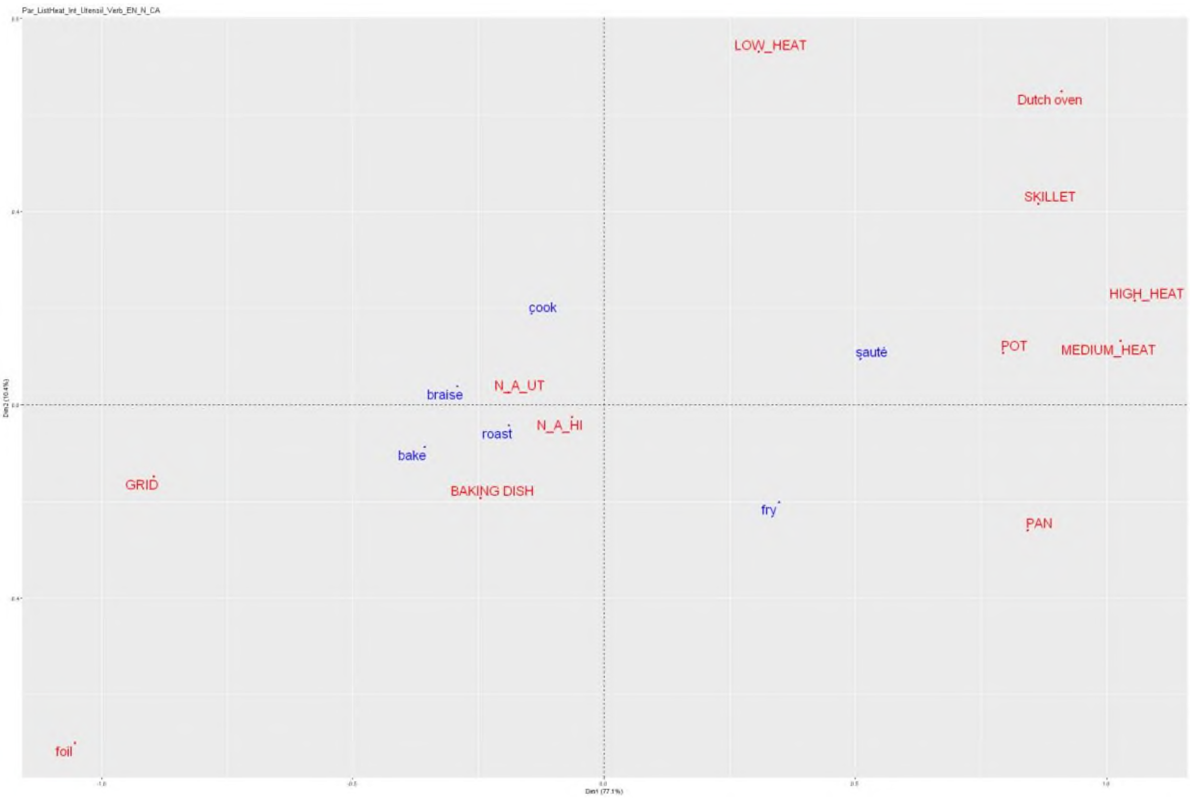


Fig. 5a. CA: English culinary verbs (parameters *Utensil* and *Heat intensity*)

Strong correlations between the dependent and independent variables illustrated on this CA graph (Fig. 5a) are those with the verb *bake* and the *Utensil* parameter values {grid,[114] foil, baking dish} (examples 53-54). {Baking dish} is also strongly correlated with the verb *roast*.

53. *Bake* the bread on the middle *rack* of oven 40-45 minutes or until golden brown and cooked through. (enTenTen15)

54. The other is to *bake* the potatoes in *foil* in Convection 175 °C for 75 minutes. (EN_REZ)

The verb *sauté* correlates strongly with the *Utensil* parameter values {skillet, pot, Dutch oven} as well as the {high heat, medium heat} ones of the *Heat intensity* parameters (examples 55-

[114] Among other parameter generalization, *Utensil* parameter value *grid* has also been generalized comprising also the utensil *rack*. For the whole list of English parameter generalizations see Appendix 7.

56). *Sauté* correlates also with {low heat}, which is also “pulled” by the verb *cook*. The verb *fry* is displayed as strongly correlated solely with the *Utensil* parameter value {pan} (example 57-58), however, *fry* is also “pulled” by {pot} as well as the *Heat intensity* parameter values {medium heat, high heat}.

55. *Sauté* the garlic and cayenne pepper in the oil over *medium heat* for about 3-4 minutes. *(enTenTen15)*

56. In a sauté pan, over *high heat*, heat the olive oil and *sauté* the oranges, onion, tomato and garlic. *(EN_REZ)*

57. Add some oil to a *pan* and *fry* the onions until softened and they have a bit of colour. *(enTenTen15)*

58. Heat the oil in a *pan* and fry the ginger until pale brown. *(EN_REZ)*

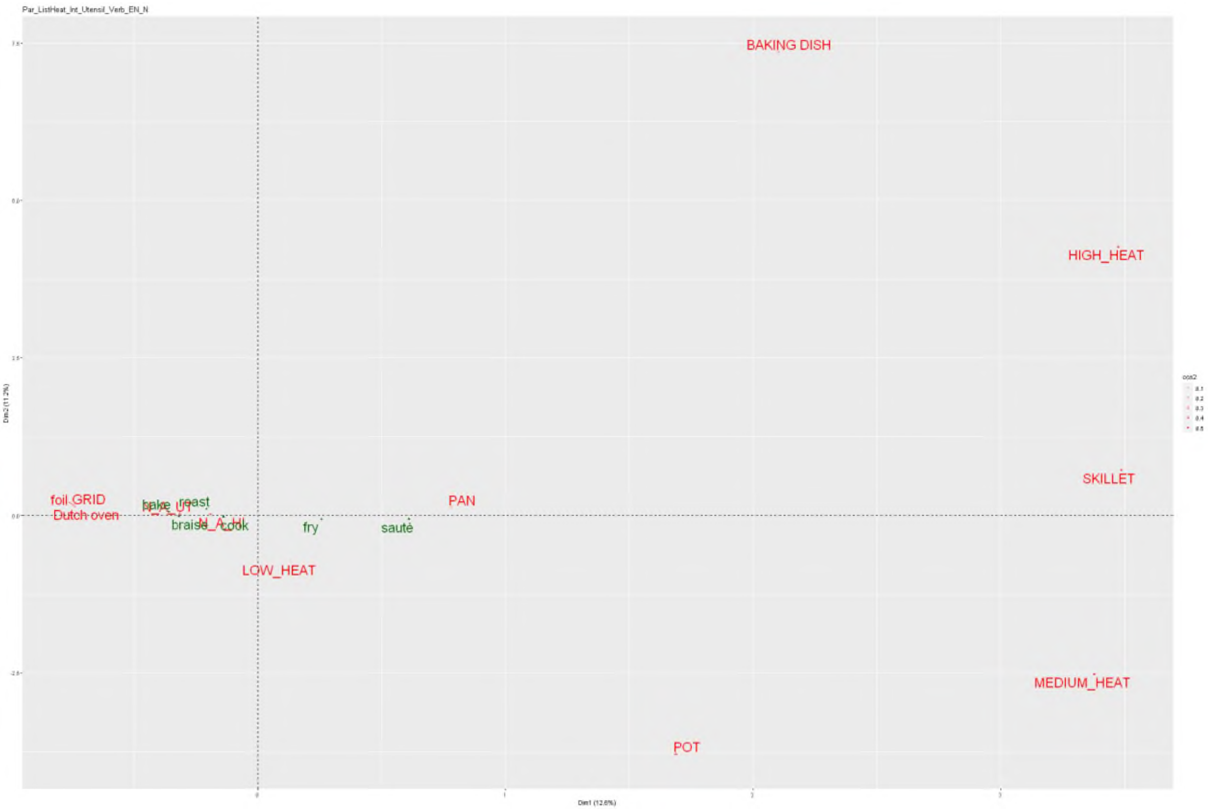


Fig. 5b. MCA: English culinary verbs (parameters *Utensil* and *Heat intensity*)

The MCA graph, on the contrary, first shows strong correlations between the independent variables, i.e. the values of *Heat intensity* and *Utensil* in clusters, followed by the respective verbs that are "pulled" to them as qualitative supplementary variables. There is not much difference between the CA and MCA of the English six verbs in relation to the two aforementioned parameters, except that because of the loss of percentage of explained inertia, the image is shrunk. Among the strong correlations between the independent variables, a small cluster of the *Utensil* parameter values {grid, foil, Dutch oven} and the {N_A_HI, low heat} values of the *Heat intensity* parameter demonstrates their co-occurrence in the same contexts.

3.3 *Context-Conditional Correlation Graph* (CCCG) Visualization of the English Data Analysis

CCCG analysis contributed to the visualization of the comparison of the English culinary verbs *bake*, *braise*, *cook*, *fry*, *roast*, and *sauté*, particularly illustrating the differences in the verb pairs with respect to differentiating parameters. This type of data visualization allowed to illustrate all strong correlations between the verb and parameter values and compares it with the respective second verb at the same time and within one graph. Each verb is represented in a specific color together with its strongly correlated parameter values, for instance the verb *bake* and {oven} as well as {N_A_Sub} marked in dark green (Fig. 1). The comparison is always carried out in verb pairs where the main verb used as reference for the comparison is displayed at the center of the graph while the other verbs are surrounded around it linked with edges. The difference between two verbs being compared is determined by the number of edges connecting them as well as the value of deviation near each parameter value: More edges and higher-than-standard deviation values correspond to significant differences between the verbs in more parameters. Consequently, fewer edges mean that the two verbs being compared are semantically closer. If the deviation in distribution of the occurrences in relation to the annotation parameters is more than twice the standard deviation, it is considered statistically significant, however, all *CCCGs* describing the differences of English culinary verbs compared in pairs have been generated above 2.5 times the standard deviation to allow better legibility, that is resulting in the reduction of the number of edges connecting verbs being compared. The following *CCCGs* were also generated with the *qgraph* package of the programming language for statistical computation *R* and were exported together with accompanying tables for each graph, where the differences are illustrated more clearly in case the graph is too complicated to do so.

3.3.1 CCCG of the English Culinary Verb *braise*

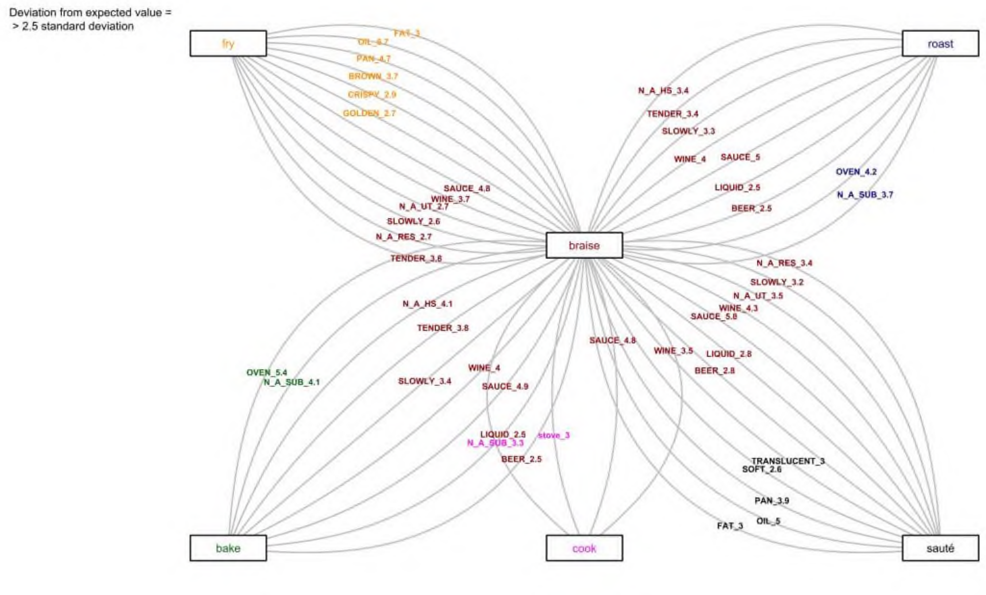


Fig. 1. *CCCG* analysis of the English culinary verb *braise*

In the verb pairs *braise-sauté* and *braise-fry,* the semantic difference between the verbs being compared is illustrated (Fig.1) in all annotation parameters, except for *Heat Intensity* and *Heat Source* as the overwhelming majority of the annotated examples for these three verbs have no verbalized value in the aforementioned two parameters. In both verb pairs, *braise* is observed to correlate with *liquid*-like *Substances,* e.g. {wine}, {sauce}, {liquid},[115] and {beer} parameter values, while *sauté* and *fry* with *oil*-like substances. For instance, in comparing the verb *braise* to *sauté* and *fry,* the co-occurrences of the *Substance* parameter value {wine} with *braise* are overrepresented by respectively 4.3 times and 3.7 times the standard deviation under the hypothesis of independent uniform distribution while the verbs *sauté* and *fry* correlate strongly with {oil} and {fat} (examples 1-5, Fig.1, partially Table 1.1.)).

1. In this video I will show you how I make my *braised* beef shank in red *wine sauce*. *(ententen15)*
2. *Sauté* the onion in the *butter* until it begins to wilt. *(EN_REZ)*

[115] The *Substance* parameter {liquid} was generalized (neutralized) from considerably smaller clusters of *liquid* substances, e.g. {milk, sirup}, which could not be integrated into other *liquid*-like categories. For the list of all generalizations (neutralizations) of parameter values see Appendix 7.

3. Heat the *oil* in large *pan* and *sauté* the onion until golden. *(ententen15)*
4. *Fry* the onion in a hot *pan* with *oil* until nearly brown then add chopped garlic. *(ententen15)*
5. Gently *fry* the onion in *butter* in a large saucepan. *(ententen15)*

3	braise	sauté	1	N_A_RES_3.4	0.422800780954605
4	braise	sauté	1	SLOWLY_3.2	0.427044709817759
6	braise	sauté	1	N_A_UT_3.5	0.42214528843673
7	**braise**	sauté	1	**WINE_4.3**	0.408375609933371
8	**braise**	sauté	1	**SAUCE_5.8**	0.393153133355304
10	**braise**	sauté	1	**LIQUID_2.8**	0.438671259842475
12	**braise**	sauté	1	**BEER_2.8**	0.438671259842475
1	sauté	braise	1	TRANSLUCENT_3	0.432140482245537
2	sauté	braise	1	SOFT_2.6	0.444900726463625
5	sauté	braise	1	PAN_3.9	0.414054402838384
9	**sauté**	braise	1	**OIL_5**	0.400180620596448
11	**sauté**	braise	1	**FAT_3**	0.433002152063737
22	braise	fry	1	TENDER_3.8	0.4153608854239
23	braise	fry	1	N_A_RES_2.7	0.443223829712201
27	braise	fry	1	SLOWLY_2.6	0.446163157018132
29	braise	fry	1	N_A_UT_2.7	0.444247319416156
30	**braise**	fry	1	**WINE_3.7**	0.417264255652944
31	braise	fry	1	SAUCE_4.8	0.402330689907782
24	fry	braise	1	GOLDEN_2.7	0.441506663078462
25	fry	braise	1	CRISPY_2.9	0.436810842922872
26	fry	braise	1	BROWN_3.7	0.417459580424079
28	fry	braise	1	PAN_4.7	0.402914298395299
32	**fry**	braise	1	**OIL_5.7**	0.394230800808648
33	**fry**	braise	1	**FAT_3**	0.43382450895468

Table (partial) 1.1. Actual distribution of the annotated occurrences in *braise-sauté* and *braise-fry* verb pairs

The semantic difference between the verbs *braise* and *sauté* in the parameter *Resultative* (adjectives) is observed in strong correlations between the verb *sauté* and {translucent, soft} while *braise* correlates with *non-verbalized* values in this parameter. In the verb pair *braise-fry* too, *Resultative* (adjectives) semantically differentiates these two verbs and has a significant impact on the distribution of the annotated examples in the English sub-corpus, with the verb *fry* correlating with {golden}, {crispy}, and {brown}. The verb *braise,* on the contrary, correlates with *non-verbalized* values of the aforementioned parameter as well as {tender} with it co-occurrences with the latter exceeding the expected value by the 3.8 times the standard deviation (cf. examples 6-9).

6. While soup is simmering, *fry* bacon in a skillet until *crispy*. *(ententen15)*

7. In a kadhai pour oil and *fry* the baby potatoes till it turns *golden brown*. *(ententen15)*

8. Flour the pieces and *fry* them a good *brown*, in butter seasoned with salt and pepper. *(EN_REZ)*

9. At one visit, we are blown away by beef stew with shredded meat in red wine and beef stock, curried vegetable, and *tender braised* lamb shank. *(ententen15)*

10. *Slowly braised* beef brisket in red wine and rosemary is the ultimate special occasion dinner. *(ententen15)*

In the verb pairs *braise-sauté* and *braise-fry*, the semantic difference is also noted in the parameter *Utensil*, e.g. with the verbs *fry* and *sauté* correlating with {pan} while *braise* correlates with non-verbalized values in this parameter (examples 1; 3-4, Fig.1, partial Table 1.1.). The parameter *Manner* with its value {slowly} as well semantically differentiates the verb *braise* in the verb pairs *braise-fry* and *braise-sauté* as the number of co-occurrences with *braise* are overrepresented by respectively 3.2 times and 2.6 times the standard deviation.

In the verb pairs *braise-bake* and *braise-roast*, the semantic difference is observed in all annotation parameters except for *Heat intensity*. In order to avoid redundancy, only a few parameter differences are elaborated here accompanied by the partial table below showing explicitly the deviations in the distribution of the annotated occurrences with respect to the differentiating parameters in these two verb pairs (partial Table 1.2).

38	**bake**	braise	1	**OVEN_5.4**	0.39669484625398
44	bake	braise	1	N_A_SUB_4.1	0.411704020355447
39	**braise**	bake	1	**N_A_HS_4.1**	0.411530295023635
40	braise	bake	1	TENDER_3.8	0.415994385660089
41	braise	bake	1	SLOWLY_3.4	0.424206392753926
42	braise	bake	1	WINE_4	0.411892410017496
43	braise	bake	1	SAUCE_4.9	0.401540727584989
45	braise	bake	1	LIQUID_2.5	0.448941857005234
46	braise	bake	1	BEER_2.5	0.448941857005234
14	braise	roast	1	N_A_HS_3.4	0.422605226129725
15	braise	roast	1	TENDER_3.4	0.423409644393201

16	braise	roast	1	SLOWLY_3.3	0.424692141051903
17	braise	roast	1	WINE_4	0.412297552104424
18	braise	roast	1	SAUCE_5	0.400150030397239
20	braise	roast	1	LIQUID_2.5	0.449589521402537
21	braise	roast	1	BEER_2.5	0.449589521402537
13	roast	braise	1	OVEN_4.2	0.409795536228679
19	roast	braise	1	N_A_SUB_3.7	0.417268558948385

Table (partial) 1.2. Actual distribution of the annotated occurrences in *braise-bake* and *braise-roast* verb pairs

Fig. 1 identifies the verb pair *braise-cook* as being semantically closest, differing only in the parameters *Substance* and *Heat source*. In comparing the verbs *braise* and *bake*, the co-occurrences of the *Heat source* parameter value {oven} with *bake* are overrepresented by 5.4 times the standard deviation under the hypothesis of independent uniform distribution (example 11).

11. To *bake* the cookies, preheat the *oven* to 350°F (180°C). *(ententen15)*

The semantic closeness in the parameters *Resultative* (adjectives), *Manner*, *Heat intensity,* and *Utensil* in this verb pair is because *cook* and *braise* have both no verbalized values in the aforementioned parameters leading to statistically no significant difference between them. The difference in the parameter *Substance*, however, is common for all three pairs: *braise* correlates with *liquid*-like substances while *bake*, *roast*, and *cook* with *no substance* or *no verbalized* substance (partial Table 1.3). Interesting is the correlation of the *Heat source* parameter value {stove} with the verb *cook* (examples 12-13) also restating that *cook* in English is a rather general verb, super-hypernym, often equaling "to prepare". Chapter 5.2 of this work also discusses the definitions of the English verb *cook* in comparing it with the Armenian culinary verb s, i.e. *ephel* (boil*, cook*).

12. After *cooking* the soup on the *stove*, the children served the soup to their friends at our dining table. *(ententen15)*

13. If you are *cooking* the rice on a conventional *stove*, simmer for about 35 minutes in the microwave, bring the water to a boil, and cook the rice on medium-low for about 30 minutes. *(EN_REZ)*

35	**braise**	cook	1	**WINE_3.5**	0.421178759808693
36	**braise**	cook	1	**SAUCE_4.8**	0.402137470772796
34	cook	braise	1	stove_3	0.433090803538376
37	**cook**	braise	1	**N_A_SUB_3.3**	0.425489114642949

Table (partial) 1.3. Actual distribution of the annotated occurrences in the verb pair *braise-cook*

3.3.2 CCCG of the English Culinary Verb *bake*

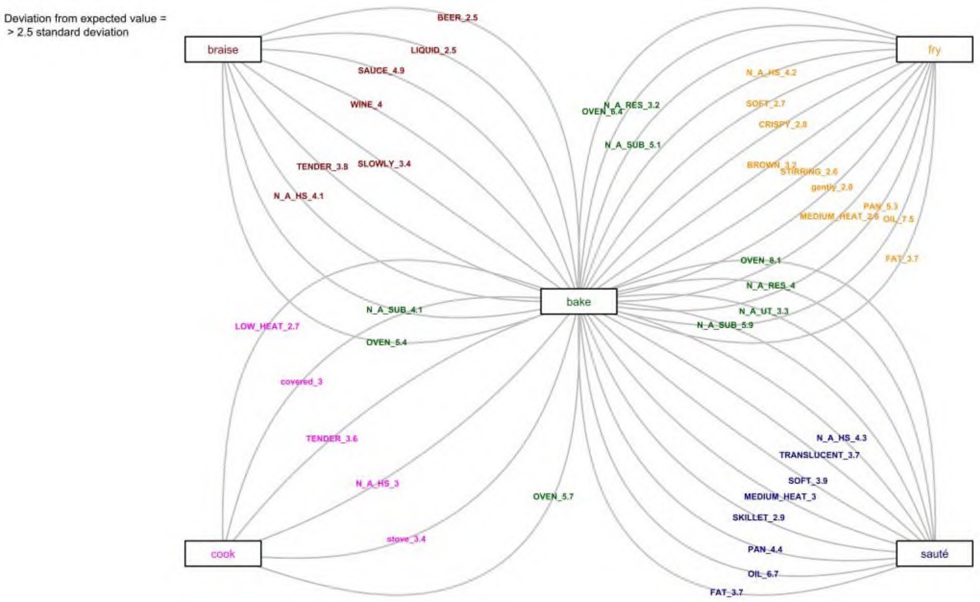


Fig. 2. *CCCG* analysis of the English culinary verb *bake*

The edges connecting the verb *bake* to the rest of the English culinary verbs, corresponding to the statistically significant semantic differences between them, denote that *bake* significantly differs in all parameters from the verbs *sauté* and *fry* (Fig. 2). For instance, in comparing the verbs *bake* and *fry*, the co-occurrences of the *Substance* parameter value{oil} with *fry* surpasses the expected value by 7.5 times the standard deviation while *bake* correlates strongly with *no verbalized* value in this parameter denoting actual absence of a cooking substance (Fig. 2, partial Table 2.1). The differentiation in the parameter *Substance* is also observed in the correlation of *bake* with the value {N_A_SUB}. *Braise* correlates with *liquid*-like while *fry* and *sauté* with *oil*-like substances. All the 113 examples of {oil} co-occur with the verb *sauté*.

13	bake	fry	1	OVEN_6.4	0.388810924656685
16	bake	fry	1	N_A_RES_3.2	0.429124426930784
24	**bake**	fry	1	**N_A_SUB_5.1**	0.398579610783999
14	fry	bake	1	N_A_HS_4.2	0.409571218490948
15	fry	bake	1	SOFT_2.7	0.442073676940214
17	fry	bake	1	CRISPY_2.8	0.439898384438383
18	fry	bake	1	BROWN_3.2	0.429209052980541
19	fry	bake	1	STIRRING_2.6	0.447914707744523
20	fry	bake	1	gently_2.8	0.440068431214891

21	fry	bake	1	MEDIUM_HEAT_2.6	0.447914707744523
22	fry	bake	1	PAN_5.3	0.396792870587611
23	**fry**	bake	1	**OIL_7.5**	0.383318272198782
25	fry	bake	1	FAT_3.7	0.418273444301742

Table (partial) 2.1. Actual distribution of the annotated occurrences in the verb pair *bake-fry*

Despite a relatively low number of edges, there are statistically significant differences between the verbs *bake* and *cook* in all of the six annotation parameters except for *Utensil*. In the verb pair *bake-braise,* the semantic closeness between the verbs being compared is observed only in the parameter *Heat intensity* (Fig. 2). The semantic differences have been discussed in detail above (Fig.1, partial Table 1.2), therefore they will not be considered here to avoid redundancy. The CCCG shows statistically no significant semantic differences in the six annotation parameters between *bake* and *roast* as the latter does not even appear on the graph. The verbs *bake* and *roast* were compared and showed that the deviation in the distribution is double of the standard deviation, exhibiting this semantic difference of the verb pair in the parameter *Substance*, with its value {oil} correlating strongly with the verb *roast*. All of the 12 examples with {oil} in comparing the verbs *roast* and *bake* co-occur with the verb *roast*.

In comparing the verbs *bake* and *fry,* the parameters *Resultative* (adjectives) with its value {N_A_RES}, *Heat source* with {oven}, and *Substance* with {N_A_Sub} had significant impact on the distribution of these two verbs by correlating strongly with the verb *bake* as opposed to *fry*. However, the graph also illustrates that the verb *bake* has *no verbalized* values in the parameters *Resultative* (adjectives) and *Substance*, denoting either simply no verbalization or absence, i.e. no cooking substance, whereas *Heat source* ({oven}) is a distinguishing parameter for *bake* (examples 13-14)

13. I personally like to make *oven baked* sweet potato fries. *(ententen15)*

14. *Bake* in a pre-heated 350degree *oven* 40 minutes, then reduce heat to 300 degrees and bake 35 to 40 minutes more, or until cake tests done. *(EN_REZ)*

3.3.3 CCCG of the English Culinary Verb *fry*

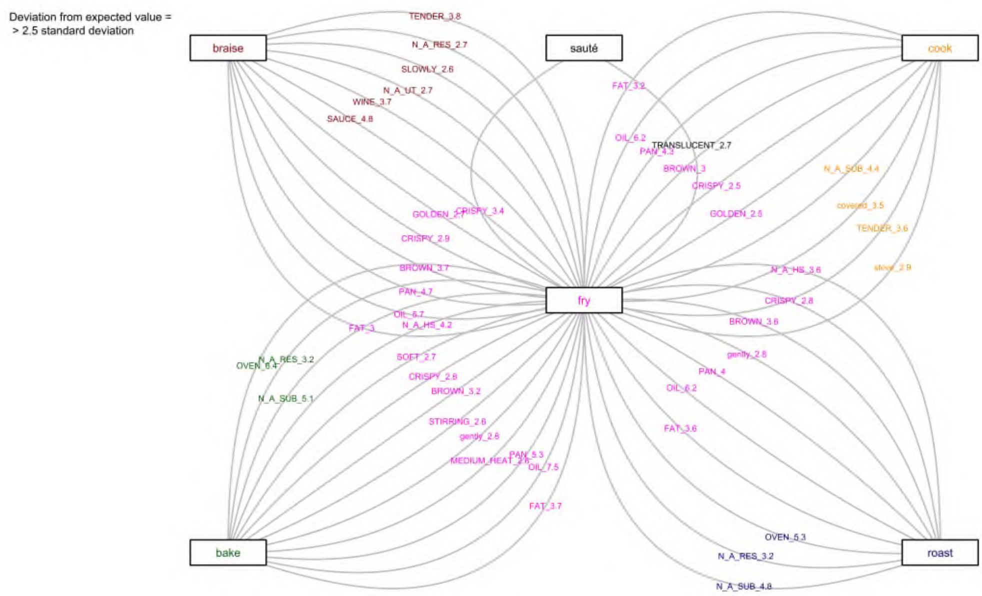


Fig. 3. *CCCG* analysis of the English culinary verb *fry*

The following *CCCG* compares the verb *fry* in pairs with the other English culinary verbs *braise*, *bake*, *sauté*, *roast*, and *cook* (Fig.3). The pairs *braise-fry* and *bake-fry* have already been compared and elaborated in detail above (see Fig.1 and 2, partial Table 1.2 and 2.1), therefore they will not be touched upon here. Given the conventions underlying the interpretation of *CCCG* analysis such as the number of edges connecting two verbs in comparison, Fig. 3 illustrates semantic similarity of the verb *fry* and *sauté* in all of our six annotation parameters, except for *Resultative* (adjectives) as well as differences in many parameters in comparing *fry* to the verbs *braise* and *roast*. The verbs *fry* and *sauté* differ in just one parameter, viz. *Resultative* (adjectives). As displayed on the following graph (Fig. 3, cf. examples 15-16) the *Resultative* (adjectives) parameter value {crispy} (marked in the same pink color) correlates strongly with the verb *fry*, while {translucent} is illustrated as a strong correlate of *sauté* (marked black). Thus, the parameter *Resultative* (adjectives) has a statistically significant impact on the distribution of the occurrences for the verbs *fry* and *sauté* differentiating them from one another with two distinct values.

15. *Fry* the tortillas for 30 to 40 seconds until *crisp*. *(ententen15)*

16. *Sauté* the shallot in canola oil over medium high heat until translucent. *(ententen15)*

In comparing the verb pair *fry* and *roast*, five of the annotation parameters, except for *Heat intensity*, had a statistically significant impact on the distribution of the occurrences for these two verbs. For instance, the co-occurrences of the *Substance* parameter value{oil} with *fry* are overrepresented by 6.2 times of the standard deviation under the hypothesis of independent uniform distribution while *roast* correlates with *non-verbalized* values in this parameter (Fig. 3, Table (partial) 3.1, examples 17-18).

17. Dice and *fry* the turkey fillets in coconut *oil* – we added a little chili powder to give it an added boost. *(ententen15)*

18. Heat the *oil* in a frying pan and *fry* the ginger for 3 to 4 minutes. *(EN_REZ)*

4	**fry**	roast	1	**N_A_HS_3.6**	0.419738084509645
6	fry	roast	1	CRISPY_2.8	0.440626640001591
7	fry	roast	1	BROWN_3.6	0.418868948744572
8	fry	roast	1	gently_2.8	0.440670578724048
9	fry	roast	1	PAN_4	0.412868734762606
10	**fry**	roast	1	**OIL_6.2**	0.390162233910673
12	**fry**	roast	1	**FAT_3.6**	0.4187911210195
3	**roast**	fry	1	**OVEN_5.3**	0.397565009759087
5	roast	fry	1	N_A_RES_3.2	0.429272269548557
11	**roast**	fry	1	**N_A_SUB_4.8**	0.402187287279735

Table (partial) 3.1. Actual distribution of the annotated occurrences in the verb pair *fry-roast*

In general, Fig. 3 illustrates that the verb *roast* correlates with *non-verbalized* values in the parameters *Substance* and *Resultative* (adjectives) while *fry* correlates with a number of values in these parameters. However, *Heat source* with its value {oven} is decisive in the differentiation of *roast* from *fry*, with the deviation from the expected value being 4.3 times of the standard deviation (Fig.1, Table (partial) 3.1, examples 19-20).

19. For a different, more intense flavour, *roast* the peppers in a hot oven for about half an hour before use. *(ententen15)*

20. *Roast* the beef in the middle of the oven for 20 minutes. *(EN_REZ)*

In the verb pair *fry-cook*, the statistically significant semantical differences are observed in all six annotation parameters, except for *Heat intensity*. The parameter *Substance* with its values {fat} and {oil} as well as the parameters *Utensil* and *Resultative* (adjectives) with their respective values {pan} and {brown}, {crispy}, and {golden} correlate strongly with the verb *fry* and have a great impact in differentiating it from the verb *cook* (examples 21-23).

21. *Fry* the onion until the color is *golden*. *(ententen15)*

22. In a large *skillet* over medium heat, *fry* the chicken pieces in 4 tablespoons of oil until they are golden brown on all sides, 5 to 7 minutes to a side. *(EN_REZ)*

23. In a large *skillet* over medium heat, *fry* the chicken pieces in 4 tablespoons of oil until they are golden brown on all sides, 5 to 7 minutes to a side. *(ententen15)*

The verb *cook* differs from *fry* in the parameters *Resultative* (adjectives), where it correlates strongly with its value {tender}, and non-verbalized values in the *Substance* parameter. However, *cook* also differs from *fry* in the parameters *Heat source* and *Manner*, correlating with {stove} and {covered} respectively (Fig. 3, Table (partial) 3.2, examples 24-25).

15	**fry**	cook	1	**GOLDEN_2.5**	0.448386856192829
16	**fry**	cook	1	**CRISPY_2.5**	0.449276125964304
17	**fry**	cook	1	**BROWN_3**	0.432569055524171
19	**fry**	cook	1	**PAN_4.3**	0.408205749242683
20	**fry**	cook	1	**OIL_6.2**	0.390407188203097
22	**fry**	cook	1	**FAT_3.2**	0.428930575546193
13	cook	fry	1	stove_2.9	0.436974438016466
14	**cook**	fry	1	**TENDER_3.6**	0.419434857577019
18	cook	fry	1	covered_3.5	0.42244209294385
21	**cook**	fry	1	**N_A_SUB_4.4**	0.40725084031703

Table (partial) 3.2. Actual distribution of the annotated occurrences in the verb pair *fry-cook*

25. *Cook* the carrots, mushrooms and cabbage until *tender*, add the red pepper and prawns and cook until the prawns are red. *(ententen15)*

26. *Cover* and *cook* until the potatoes are *tender*, about 10 minutes. *(EN_REZ)*

3.4 *Conditional Inference Trees* (CIT) of the English Culinary Verbs

Conditional Inference Trees (CIT) – introduced in Chapter 2.5 of this work – is a form of a tree-based structure of visualization anchored in recursive partitioning of data. Its significance for this research is also illustrated in detail in the same sub-chapter. The *CIT* model of the English culinary verbs was generated with the programming language *R* – like its German and Armenian counterparts– based on previously annotated context examples. *CIT* model is a tree-structured data analysis based on recurrent splitting of the data based on statistically significant tests, particularly p-values (see Gries 2009). The splitting is carried out until it is statistically irrelevant. The lower the p-value, the lesser is the probability that the splitting of the node into further inner nodes is accidental. In the case of the *CIT* model of the English culinary verb, lower p-values indicate that the annotated parameters had a strong (decisive) effect on the distribution of the verbs under the particular node. P-values reflect the real difference between two neighboring nodes proving that the distribution was not merely a result of random data selection. For instance, p<0.06 under the parameter *Heat source*, splitting the nodes 17 and 18 (Fig. 1), show the likelihood of a merely accidental difference in the distribution between the nodes is smaller than 0,6%.

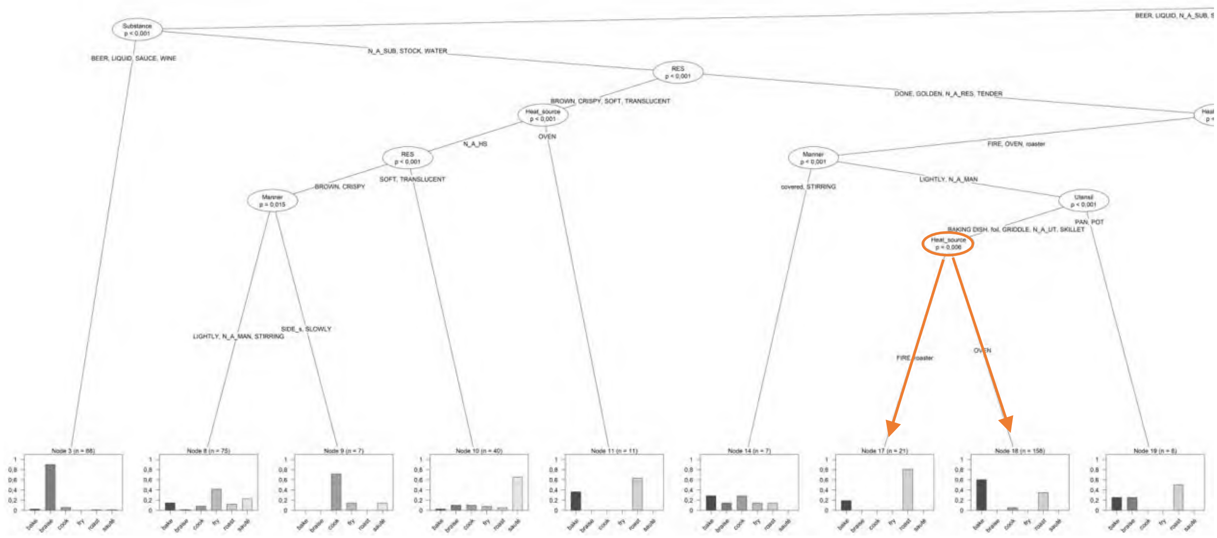


Fig. 1. Partial *CIT*: Distribution model of the English culinary verbs in context

CIT model of the English culinary verbs illustrates those context clusters where the most frequent verb is identified contributing to the intralinguistic analysis of the English culinary fields. As mentioned in Chapter 2.1 of this work, six parameters, i.e. *Substance, Utensil,* *Manner, Resultative* (adjectives), *Heat intensity* and *Heat source* proved to be significant in identifying those clusters of possible context types where one of the six English culinary verbs stood out as a quantitatively dominant verb. The parameters *Ingredient* and *Dish* were evaluated and found to be unsuitable for inclusion in the CIT model, as their high variability and large number of distinct values hindered effective generalization and compromised the model's predictive performance.

CIT model was generated based on previously annotated 1683 occurrences extracted from both the *English Web 2015 (enTenTen15)* and *EN_REZ* corpora described in detail in Chapter 3.1 of this work. However, due to the reduction of parameters – guided by frequency of occurrence and the statistical significance of each parameter – the final annotated examples serving as the basis for this *CIT* model were reduced to 1273[116]. The partitioning of the *CIT* model of the English culinary verbs resulted in 19 terminal and 18 inner nodes, identifying 19 context clusters where the most frequent verb is identified visually illustrated by the highest column at the bottom of the nodes. The verbs are arranged in alphabetical order beginning with *bake* followed with *braise, cook, fry, roast, and sauté.*

In contrast to the *CIT* modeling of the German and Armenian culinary verbs, the English *CIT* analysis yielded comperatively less conclusive results. The overall frequency of the dominant verbs in the identified English culinary context clusters is 47% percent (cf. German 61,5% in Chapter 2.5, and Armenian 69% in Chapter 4.4). Referring to this particular parameter combination as heuristic to "guess" an appropriate lexical choice in English by selecting the most frequent verb in every cluster would result in an average "error rate" of ca. 53% (cf. German 38,5%, and Armenian 31%). By excluding the largest cluster (Node 31, occurrences = 345, error rate = 74.8%), which exhibited the highest error due to the absence of parameter values (annotated as N_As in the respective parameters) rather than a flaw in the model itself, the overall error rate reduced to 44%. Nevertheless, choosing the most frequent verb in a certain context cluster instead of the other, does not automatically imply the inadequate or false choice of the verb. The error rate denotes that semantically closer verb(s) are also identified in the cluster based on frequency of occurrences. The somewhat lower percentage of accuracy of the English *CIT* is mostly due to the peculiarities of the English culinary verb classification and hierarchy. For instance, *cook* in English is considered a super-hypernym for all culinary verbs while *kochen* in German is a hypernym only for the verbs *sieden, dünsten, dämpfen,* and *schmoren* (see Lehrer 1972:161). However, even with a relatively less success than the German

[116] All parameter reductions of the English culinary verbs are available in Appendix 7 of this work.

and the Armenian *CIT* results, the English *CIT* model is still a significant improvement. We compared our *CIT* model to both the distribution of the culinary verbs in the *enTenTen15* corpus where *cook* is identified as the most frequent verb and to any randomly chosen verb out of our six English culinary verbs outside the corpus. Under the hypothesis of an independent discrete uniform distribution, the probability of randomly choosing the correct verb out of the six English culinary verb is 1/6, viz. ≈16,66%, so that a random guess produced an error rate of ≈83,33%.

Comparing our CIT model to the distribution of the English culinary verbs in the *enTenTen15* corpus (Fig. 2), always choosing *cook*, the most frequent verb based on *ententen15* corpus, seems a better verb choice. Even without the worst cluster, mostly with *non-verbalized values* in almost all annotation parameters (node 31, 345 occurrences, error rate=74,8%), the correct predictions of our *CIT* model rose up to max. 55%.

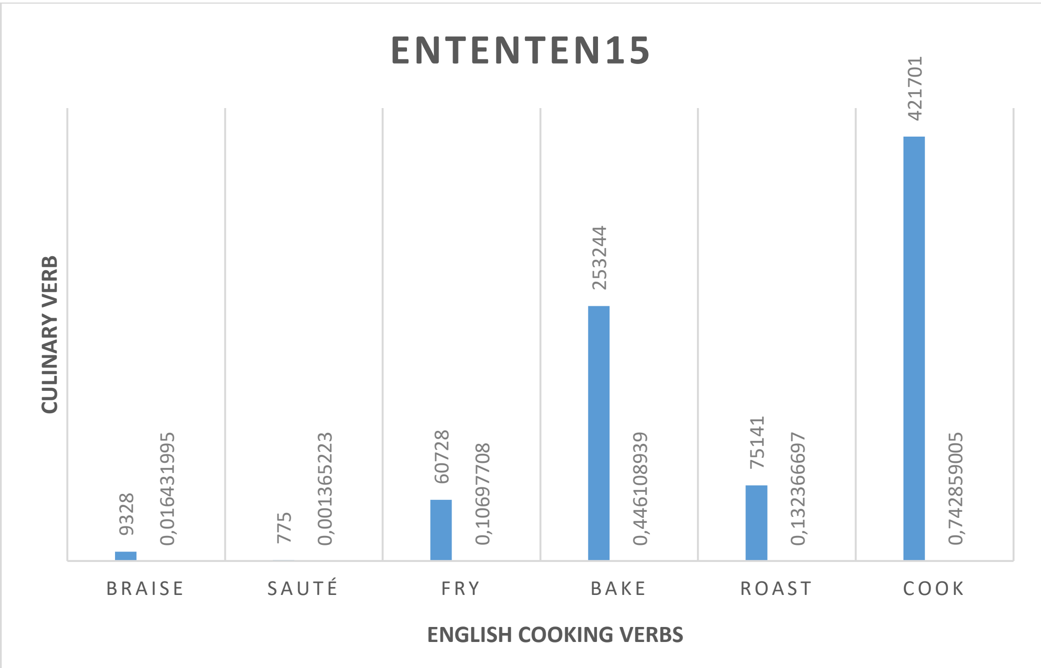


Fig. 2. Distribution of the English culinary verbs in the *enTenTen15* corpus

In this case, the developed *CIT* for the English culinary verbs appears a less successful solution, however, throughout the work it was stated a number of times that *cook* for English is very general verb which practically means “the food is ready for consumption”, “the food is prepared” in some way. It is not a mere a hypernym for *boil* or *simmer* like *kochen* is for *dünsten* but rather a super-hypernym for all the culinary verbs, from *braise* to *bake*. Yet, the most frequent verb *cook* generates both overt and evident mistakes, viz. wrong verb choices and also hidden ones in case a larger context is not provided. For instance, if instead of *'to bake bread'* we use *'to cook bread'* by choosing the most frequent verb in the *enTenTen15* corpus, the mistake is evident. If a larger context is provided, such as *'in the oven'*, *cook* could also be used for *bake*. In *'to cook**[117] coffee beans'* instead of *'to roast coffee beans'*, *cook* could be taken for *boil* [+Water], generating covert/hidden mistakes. A somewhat larger context, such as *'to cook** coffee beans in a pan*' might bewilder about the wrong choice of the verb and make to rethink, however, it is not sufficient to conclude that the correct verb is *roast*. Specific verbs, such as *roast*, need a larger context for the correct verb choice. Therefore, our *CIT* model for identifying the most frequent verb in context was still a success for specific verbs like *roast* avoiding severely wrong verb choices. Fig. 3 compares our *CIT* model of the English culinary verbs to the most frequent verb in the *enTenTen15* corpus as well as to the independent discrete uniform distribution with regard to correct and incorrect predictions.

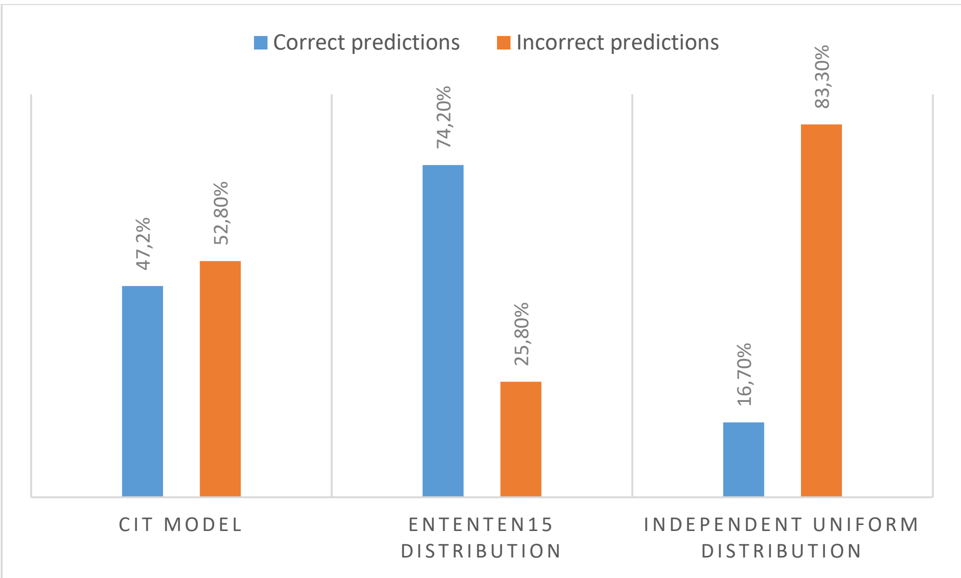


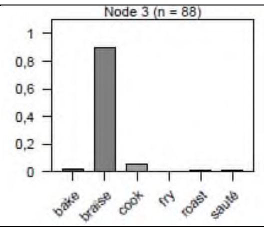
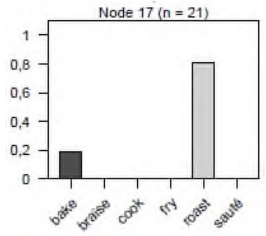
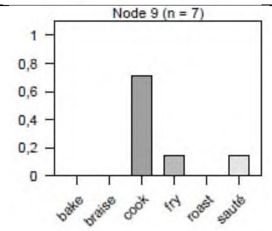
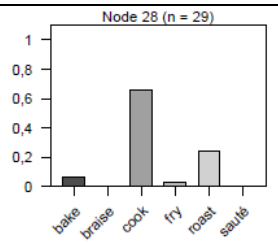
Fig. 3. *CIT* model in comparison: correct and incorrect prediction of the verb

The identified clusters determined by combinations of context parameters were categorized into three groups based on ascending error rates. The first group includes clusters with the lowest error rate under approximately 1/3. The second group constitutes those clusters with moderate (up to 50%) while the third group with considerably higher error rate (above 50%).

[117] Two asterisks (**) here denote completely wrong verb choice.

3.4.1 The First Group of Context Clusters Determined by Parameters

The first group of context clusters comprises those ones with the lowest error rate ranging from 10,2% to 35%. The first group includes 5 clusters with overall 185 occurrences (Table 1). It enables to track the nodes from top to bottom by identifying the path from the root to each terminal node. It also presentes the total number of occurrences under each node, the distribution of verbs within each cluster, the associated error rate, and the p-value influencing the partitioning at that node.

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
3	braise	88	10,2272727	bake = 2 braise = 79 cook = 5 fry = 0 roast = 1 sauté = 1		Substance (BEER, LIQUID, SAUCE, WINE)
17	roast	21	19,0476191	bake = 4 braise = 0 cook = 0 fry = 0 roast = 17 sauté = 0		Substance (N_A_SUB, STOCK, WATER), RES (DONE, GOLDEN, N_A_RES, TENDER) , Manner (LIGHTLY, N_A_MAN) , Utensil (BAKING DISH, foil, GRID, N_A_UT, SKILLET), Heat_source (FIRE, roaster)
9	cook	7	28,5714286	bake = 0 braise = 0 cook = 5 fry = 1 roast = 0 sauté = 1		Substance (N_A_SUB, STOCK, WATER) , Heat_source (N_A_HS), RES (BROWN, CRISPY), Manner (SIDE_s, SLOWLY)
28	cook	29	34,4827586	bake = 2 braise = 0 cook = 19 fry = 1 roast = 7 sauté = 0		Substance (N_A_SUB, STOCK, WATER), Manner (covered, N_A_MAN, SLOWLY), Heat_intensity (N_A_HI), RES (N_A_RES), Heat_source (GRILL, stove)

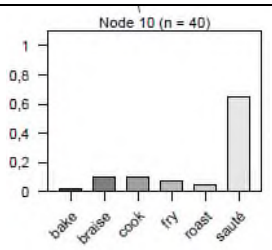
10	sauté	40	35	bake = 1 braise = 4 cook = 4 fry = 3 roast = 2 sauté = 26		Substance (N_A_SUB, STOCK, WATER) , Heat_source (N_A_HS), RES (SOFT, TRANSLUCENT)

Table 1. The first group of context-parameter determined clusters with the lowest error rate

The first cluster in this group illustrates a context type (node 3) where with around 10,2% error rate the English culinary verb *braise* is identified as the most frequent verb. The only determining parameter for this cluster is *Substance* with its values {beer, liquid, sauce, wine}. The overwhelming majority of the overall examples under this node (79 out of 88) are with the verb *braise* representing also the largest cluster in this group.

The second cluster with considerably low error rate (around 19%) is observed under the node 17 where *roast* is singled out as statistically the dominant verb in context. Even though this cluster is relatively smaller than the first one, almost all occurrences are with the verb *roast* determined by five of our six annotation parameters, namely *Substance* x *Heat source* x *Resultative* (adjectives) x *Manner* (Table 1).

The next cluster (node 9) is rather small with only 7 examples and around 28,5% error rate. However, with 5 occurrences the English *cook* is identified as the most frequently used verb in this context type determined by the parameter combinations *Substance* x *Heat source* x *Resultative* (adjectives) x *Manner* with their values {N_A_SUB, stock, water} x {N_A_HS}x {brown, crispy} x {Side_s, slowly} respectively.

There are two more clusters in this group with 34,5-35% error rate, viz. under the nodes 28 and 10 respectively. In the first cluster with 19 examples out of the overall 29 occurrences, the verb *cook* is identified as the most frequently used verb determined by the parameter combinations *Substance* x *Manner* x *Heat intensity* x *Resultative* (adjectives) x *Heat source* with their respective values {N_A_SUB, stock, water} x {covered, N_A_MAN, slowly} x {N_A_HI} x {N_A_RES} x {grill, stove}. The absence of verbalized values in parameters *Substance* and *Resultative* affected the distribution of the verbs in this cluster where along with the most frequent verb *cook*, the English culinary verbs *bake*, *fry*, and *roast* shares the distribution. In the aforementioned second cluster the parameters *Substance* x *Heat source* x *Resultative* (adjectives) with their combinations of values {N_A_SUB, stock, water}x{N_A_HS} x {soft, translucent} determined a context type where *sauté* is identified as the most frequent verb.

Nevertheless, the rest of the occurrences are proportionally distributed to all of the remaining English culinary verbs.

3.4.2 The Second Group of Context Clusters Determined by Parameters

The second group of context types generated by our *CIT* model of the English six culinary verbs included clusters with moderate error rate ranging from 40-50% where certain parameter combinations determined the distribution of the English culinary verbs by identifying the most frequent one. Despite a slightly higher error rate, this cluster group remains worth considering, as it represents a significant 33% improvement compared to the independent discrete uniform distribution of the verbs in question. 7 clusters with overall 565 occurrences constitute this group of context types.

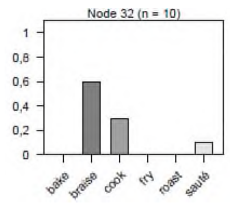
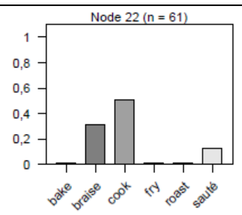
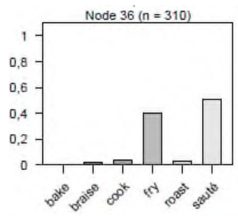
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
32	braise	10	40	bake = 0 braise = 6 cook = 3 fry = 0 roast = 0 sauté = 1		Manner (covered, N_A_MAN, SLOWLY), Heat_intensity (N_A_HI), RES (N_A_RES) , Heat_source (N_A_HS), Utensil (BAKING DISH, Dutch oven, GRID, N_A_UT), Substance (STOCK)
22	cook	61	49,1803279	bake = 1 braise = 19 cook = 31 fry = 1 roast = 1 sauté = 8		Substance (N_A_SUB, STOCK, WATER) , Heat_source (GRILL, N_A_HS, stove), Manner (covered, N_A_MAN, SLOWLY), RES (DONE, TENDER)
36	sauté	310	49,3548387	bake = 1, braise = 7, cook = 11, fry = 124, roast = 10, sauté = 157		Substance (FAT, OIL), Heat_source (FIRE, N_A_HS, stove)

Table 2.1. The second group of context-parameter determined clusters with the lowest error rate

The first cluster in this group is a rather small one with as few as 10 examples where, however, 6 occurrences were with the verb *braise*. All of our six annotation parameter combinations played a role in identifying the verb *braise* as the most frequent one in this cluster under the node 32 (Table 2.1). However, judging from the annotation parameters overlapping *braise* and *cook* (see also

Chapter 3.3), these two verbs might substitute each other in this context types by giving rise to 40% error rate in this cluster.

In the next cluster with 61 overall examples, the verb *cook* is identified as the most frequent verb determined by the parameters *Substance* x *Heat source* x *Manner* x *Resultative* (adjectives) with their combinations {N_A_SUB, stock, water} x {grill, N_A_HS, stove} x {covered, N_A_MAN} x {done, tender} respectively. These parameter values yet also affected the distribution of other verbs in this cluster by identifying the verb *braise* as the next frequent verb due to shared values, for instance, {water, stock}, {covered}, {done, tender} (see Chapter 3.3 on M(CA)). For this reason, the verb *cook* is identified but with around 49,2% error rate as the most frequent verb.

Node 36 displays a considerably larger cluster with overall 310 examples, where the verb *sauté* is identified as the most frequently used verb in context. Although all of the English verbs shared the distribution under this cluster, the verb *fry* is singled out in that context type had but a slightly fewer examples than the most frequent verb *sauté* – *fry* has 124 while *sauté* has 157 occurrences. This form of distribution in the cluster is explained by the parameter combinations determining this type of context. In fact, only two parameters *Substance* x *Heat source* with their values {fat, oil} x {fire, N_A_HS, stove} were decisive for the verb *fry* to be the second most frequent verb in the cluster with almost the same number of occurrences.

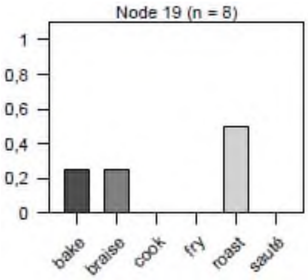
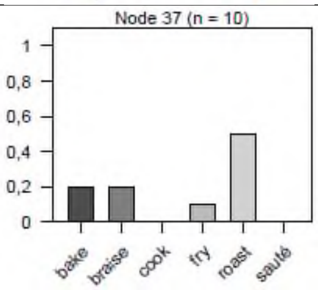
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
19	roast	8	50	bake = 2 braise = 2 cook = 0 fry = 0 roast = 4 sauté = 0		Substance (N_A_SUB, STOCK, WATER), RES (DONE, GOLDEN, N_A_RES, TENDER) , Heat_source (FIRE, OVEN, roaster), Manner (LIGHTLY, N_A_MAN), Utensil (PAN, POT),
37	roast	10	50	bake = 2 braise = 2 cook = 0 fry = 1 roast = 5 sauté = 0		Substance (FAT, OIL), Heat_source (OVEN)

Table 2.2. The second group of context-parameter determined clusters with moderate error rate

The last two clusters in this group are rather small with 8 and 10 examples respectively while both have an error rate of 50% stating that only half of the occurrences are with *roast* identifying it as the most frequent verb in context. Given the small size, it is not worth elaborating in detail on the parameters determining them. However, Table 2.2 illustrates both the path from the verb to the parameter values and the distribution in these clusters.

3.4.3 The Third Group of Context Clusters Determined by Parameters

The third group includes 6 clusters with considerably higher error rate ranging from 58-74% and with overall 520 occurrences. The largest cluster under the node 31 with 345 examples is also included in this group. As mentioned at the beginning of this chapter, this cluster was not included in the analysis of the *CIT* model of the English culinary verbs since its high error rate is related to the absence of verbalized values in almost all of the six parameters hindering the model to identify the dominant verb (Table 3.1). It is also less reasonable to consider all clusters in this group, especially noticeably smaller clusters such as the ones under the nodes 24 and 26. Therefore, this cluster will not considered for further analysis. The clusters are provided in the table below so that relevant information could easily be retrieved when necessary.

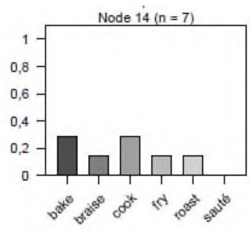
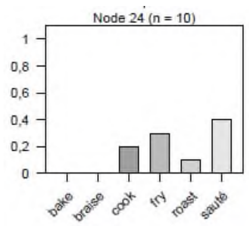
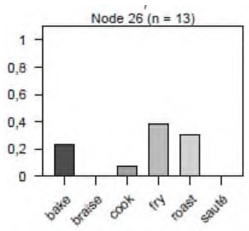
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster_path
14	bake	7	71,4285714	bake = 2 braise = 1 cook = 2 fry = 1 roast = 1 sauté = 0		Substance (N_A_SUB, STOCK, WATER) , RES (DONE, GOLDEN, N_A_RES, TENDER) , Heat_source (FIRE, OVEN, roaster), Manner (covered, STIRRING)
24	sauté	10	60	bake = 0, braise = 0, cook = 2, fry = 3, roast = 1, sauté = 4		Substance (N_A_SUB, STOCK, WATER) , Heat_source (GRILL, N_A_HS, stove), Manner (covered, N_A_MAN, SLOWLY) , RES (GOLDEN, N_A_RES), Heat_intensity (HIGH_HEAT, LOW_HEAT, MEDIUM_HEAT)
26	fry	13	61,5384615	bake = 3, braise = 0, cook = 1, fry = 5, roast = 4, sauté = 0		Substance (N_A_SUB, STOCK, WATER) , Heat_source (GRILL, N_A_HS, stove) , Manner in c(covered, N_A_MAN, SLOWLY) , Heat_intensity in c(N_A_HI) , RES (GOLDEN)

Table 3.1. The smallest clusters with the highest error rate

The next cluster under the node 34 with 34 examples and 70% error rate identifies the verbs *cook* and *sauté*. Nevertheless, all clusters with rather higher error rate are due to the absence of verbalized information in most of the parameters. It impacted the distribution in the clusters where the model managed badly to single out the most frequent verb by allocating occurrences to all of the six verbs. However, the largest cluster in this group under the node 8 with 75 occurrences is worth consideration since even with around 58% error rate it is still an amelioration when compared to the most frequent verb under the hypothesis of independent discrete uniform distribution. Thus, with 31 occurrences the verb *fry* is identified as the most frequent verb in this context type determined by the parameter combinations *Substance* x *Heat source* x *Manner* with their respective values {N_A_SUB, stock, water} x {N_A_HS}x {lightly, N_A_MAN, stirring} (Table 3.2).

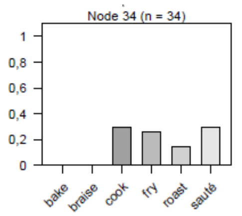
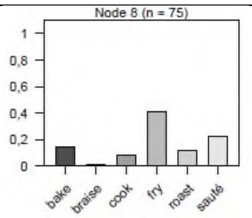
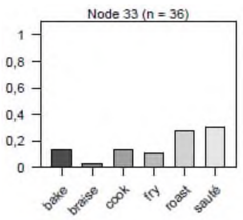
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster_path
34	cook	34	70,588 2353	bake = 0 braise = 0 cook = 10 fry = 9 roast = 5 sauté = 10		Substance (N_A_SUB, STOCK, WATER) , RES (DONE, GOLDEN, N_A_RES, TENDER), Heat_source (GRILL, N_A_HS, stove) , Manner (gently, LIGHTLY, SIDE_s, STIRRING)
8	fry	75	58,666 6667	bake = 11 braise = 1 cook = 6 fry = 31 roast = 9 sauté = 17		Substance (N_A_SUB, STOCK, WATER), Heat_source (N_A_HS), RES (BROWN, CRISPY), Manner (LIGHTLY, N_A_MAN, STIRRING)

Table 3.2. The third group of context-parameter determined clusters with considerably high error rate

The cluster under the node 31 and 33 are be excluded from further consideration as mentioned at the beginning of Chapter 3.5. Most of the parameter combinations here are with *non-verbalized values* resulting in shared distribution for all the English culinary verbs (Table 3.3).

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster_path
33	sauté	36	69,444 4444	bake = 5, braise = 1, cook = 5, fry = 4, roast = 10, sauté = 11		Substance (N_A_SUB, STOCK, WATER), Manner (covered, N_A_MAN, SLOWLY), Heat_intensity (N_A_HI), RES (N_A_RES), Heat_source (N_A_HS), Utensil (PAN, POT, SKILLET)

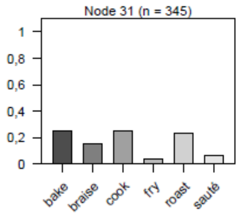
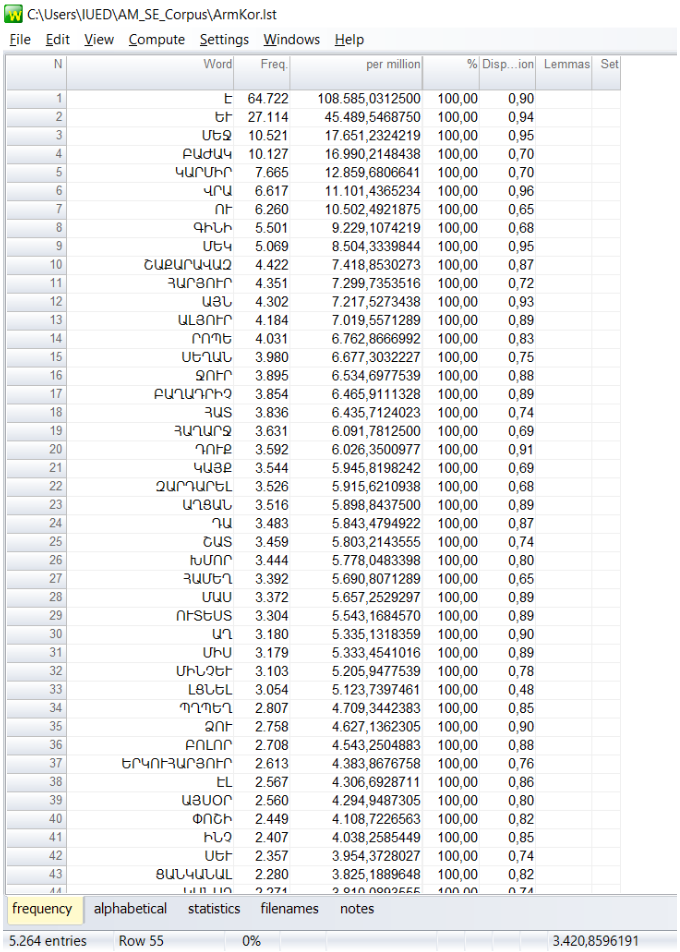
31	bake	345	74,782 6087	bake = 87 braise = 53 cook = 87 fry = 14 roast = 81 sauté = 23		Manner (covered, N_A_MAN, SLOWLY), Heat_intensity (N_A_HI) , RES (N_A_RES), Heat_source (N_A_HS), Utensil (BAKING DISH, Dutch oven, GRID, N_A_UT) , Substance (N_A_SUB)

Table 3.3. The clusters with the highest error rate due to non-verbalized parameter values

Chapter 4. Description of the Armenian *ARM* corpus

It was already mentioned in Chapter 2.1. that due to the absence of parallel corpora, comparable English and German corpora available in *Sketch Engine* together with manually compiled ones served as the basis for intra- and interlinguistic analysis of this work. However, the Armenian language presented more difficulties for a number of reasons. First, the *Sketch Engine* corpus query system and language analyzer did not offer any Armenian corpora giving rise to several problems. It not only excluded the chance of finding a comparable Armenian corpus, preferably approximately of the same size as the German and English counterparts, but also made it impossible to analyze even the one manually compiled by me. The absence of an Armenian corpus of any size in *Sketch Engine* also implied that no Armenian part-of-speech (POS) tagging is available in this software. In general, no Armenian web corpus comparable to the German Web 2013 *(deTenTen13)* as well as English Web 2015 *(enTenTen15*) of around 19,8 and 15,4 billion tokens respectively was available in any platform. The East Armenian National Corpus, 2007-2009 (EANC) with overall 110 million tokens was available at http://www.eanc.net/, however, since the corpus was static and the texts were predominantly either journal articles or literary works and scarcely culinary texts, the extracted culinary sub-corpus was not considered suitable due to its incompatible size. However, the token markup (including lexical morphological labels) underlying the EANC allowed us to upgrade the Armenian POS tagging (Fig. 1) making it possible to analyze our own Armenian corpus.[118] The tagging is available under https://bitbucket.org/timarkh/uniparser-grammar-eastern-armenian/src/master/ link. Consequently, having successfully carried out the pipeline of all processes necessary for a corpus-based analysis, an Armenian culinary corpus of 1,190,779 tokens *(ARM)* was compiled in conformity with the German *(REZ_DE)* and English *(EN_REZ)* manually assembled corpora with around 1,5 million tokens each. The following Armenian corpus includes both cookbooks available in print or digital format and web recipes extracted from culinary websites. Thus, the intra- and interlinguistic analysis of the Armenian culinary fields was carried out based on the aforementioned *ARM* corpus.

[118] Until recently, the Armenia lemmatization and POS-tagging used in this work was the only possibility for tagging Armenian texts which was rather time consuming. However, by now it is possible to run the lemmatization and part-of-speech tagging of large Armenian texts via the *Stanza* natural language analysis package in *Python* (see https://github.com/iued-uni-heidelberg/corpustools/blob/main/S01LemmatizationEnHyV01.ipynb).


C:\Users\IUED\AM_SE_Corpus\ArmKor.lst

File Edit View Compute Settings Windows Help

N	Word	Freq.	per million	%	Disp...ion	Lemmas	Set
1	Է	64.722	108.585,0312500	100,00	0,90		
2	ԵՒ	27.114	45.489,5468750	100,00	0,94		
3	ՄԵՋ	10.521	17.651,2324219	100,00	0,95		
4	ԲԱԺԱԿ	10.127	16.990,2148438	100,00	0,70		
5	ԿԱՐՄԻՐ	7.665	12.859,6806641	100,00	0,70		
6	ՎՐԱ	6.617	11.101,4365234	100,00	0,96		
7	ՈՒ	6.260	10.502,4921875	100,00	0,65		
8	ԳԻՆԻ	5.501	9.229,1074219	100,00	0,68		
9	ՄԵԿ	5.069	8.504,3339844	100,00	0,95		
10	ՇԱՔԱՐԱՎԱԶ	4.422	7.418,8530273	100,00	0,87		
11	ՀԱՐՅՈՒՐ	4.351	7.299,7353516	100,00	0,72		
12	ԱՅՆ	4.302	7.217,5273438	100,00	0,93		
13	ԱԼՅՈՒՐ	4.184	7.019,5571289	100,00	0,89		
14	ՐՈՊԵ	4.031	6.762,8666992	100,00	0,83		
15	ՍԵՂԱՆ	3.980	6.677,3032227	100,00	0,75		
16	ՋՈՒՐ	3.895	6.534,6977539	100,00	0,88		
17	ԲԱՂԱԴՐԻՉ	3.854	6.465,9111328	100,00	0,89		
18	ՀԱՏ	3.836	6.435,7124023	100,00	0,74		
19	ՀԱՂԱՐՋ	3.631	6.091,7812500	100,00	0,69		
20	ԴՈՒՔ	3.592	6.026,3500977	100,00	0,91		
21	ԿԱՅՔ	3.544	5.945,8198242	100,00	0,69		
22	ՁԱՐԴԱՐԵԼ	3.526	5.915,6210938	100,00	0,68		
23	ԱՂՑԱՆ	3.516	5.898,8437500	100,00	0,89		
24	ԴԱ	3.483	5.843,4794922	100,00	0,87		
25	ՇԱՏ	3.459	5.803,2143555	100,00	0,74		
26	ԽՄՈՐ	3.444	5.778,0483398	100,00	0,80		
27	ՀԱՄԵՂ	3.392	5.690,8071289	100,00	0,65		
28	ՄԱՍ	3.372	5.657,2529297	100,00	0,89		
29	ՈՒՏԵՍՏ	3.304	5.543,1684570	100,00	0,89		
30	ԱՂ	3.180	5.335,1318359	100,00	0,90		
31	ՄԻՍ	3.179	5.333,4541016	100,00	0,89		
32	ՄԻՆՉԵՒ	3.103	5.205,9477539	100,00	0,78		
33	ԼՑՆԵԼ	3.054	5.123,7397461	100,00	0,48		
34	ՊՂՊԵՂ	2.807	4.709,3442383	100,00	0,85		
35	ՁՈՒ	2.758	4.627,1362305	100,00	0,90		
36	ԲՈԼՈՐ	2.708	4.543,2504883	100,00	0,88		
37	ԵՐԿՈՒՀԱՐՅՈՒՐ	2.613	4.383,8676758	100,00	0,76		
38	ԵԼ	2.567	4.306,6928711	100,00	0,86		
39	ԱՅՍՕՐ	2.560	4.294,9487305	100,00	0,80		
40	ՓՈՇԻ	2.449	4.108,7226563	100,00	0,82		
41	ԻՆՉ	2.407	4.038,2585449	100,00	0,85		
42	ՍԵՒ	2.357	3.954,3728027	100,00	0,74		
43	ՑԱՆԿԱՆԱԼ	2.280	3.825,1889648	100,00	0,82		

frequency alphabetical statistics filenames notes

5.264 entries Row 55 0% 3.420,8596191

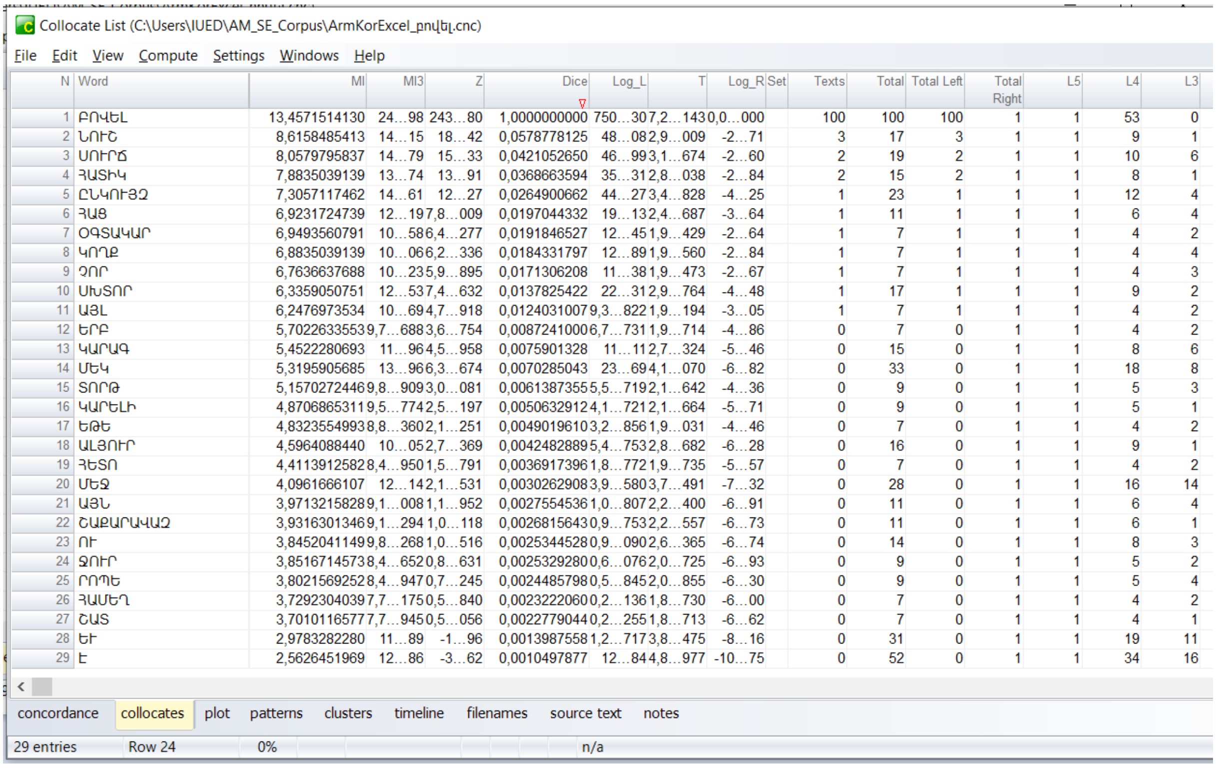
Fig. 1. Partial word list of the Armenian *ARM* sub corpus

As mentioned above, the *Sketch Engine* did not have (still does not have) any Armenian corpus, therefore rather limited language processing functions such as *Concordance*,[119] were available for analyzing Armenian texts. The function *Concordance*, nevertheless, allowed to extract occurrences as context examples which were annotated according to the parameters described in detail at the beginning of Chapter 2.1. However, with the function *Word Sketch*[120] missing, the Armenian culinary corpus data was analyzed with the *Word Smith* (Scott 2020) tool. Even though the *Word Smith* tool also offers among others collocational analysis, it is, however, based on the *Dice* score (Fig. 2). Since all collocations of the Armenian corpus irrespective of

[119] Besides *Concordance*, *Sketch Engine* also offers *N-grams*, *Wordlist*, *Keywords*, *Text type analysis*, *OneClick Dictionary* functions for the Armenian language.

[120] *Sketch Engine* provides a one-page summary of the collocational and grammatical behavior of a word (see Chapter 2.1.) based on one of the association scores between two neighboring words, viz. *LogDice*. For a detailed information on all the association scores including *LogDice*, see Chapter 2 of this work as well as Evert 2005, Killgariff et al. 2010.

their association score have been taken into consideration, there was no necessity to convert the *Dice* score into *LogDice*. This combined method of *Sketch Engine Concordance* function and the *Word Smith* collocation analysis contributed to both the identification of collocations of six Armenian cooking verbs, i.e. *բովել [bovel]*, *եփել [ephel]*, *թխել [thxel]*, *խաշել [xashel]*, *շոգեխաշել [shogexashel]*, and *տապակել [tapakel]* and to the annotation of their contextual examples.



Collocate List (C:\Users\IUED\AM_SE_Corpus\ArmKorExcel_բովել.cnc)

File Edit View Compute Settings Windows Help

N	Word	MI	MI3	Z	Dice	Log_L	T	Log_R	Set	Texts	Total	Total Left	Total Right	L5	L4	L3
1	ԲՈՎԵԼ	13,4571514130	24…98	243…80	1,0000000000	750…30	7,2…143	0,0…000		100	100	100	1	1	53	0
2	ՆՈՒՇ	8,6158485413	14…15	18…42	0,0578778125	48…08	2,9…009	-2…71		3	17	3	1	1	9	1
3	ՍՈՒՐՃ	8,0579795837	14…79	15…33	0,0421052650	46…99	3,1…674	-2…60		2	19	2	1	1	10	6
4	ՀԱՏԻԿ	7,8835039139	13…74	13…91	0,0368663594	35…31	2,8…038	-2…84		2	15	2	1	1	8	1
5	ԸՆԿՈՒՅԶ	7,3057117462	14…61	12…27	0,0264900662	44…27	3,4…828	-4…25		1	23	1	1	1	12	4
6	ՀԱՑ	6,9231724739	12…19	7,8…009	0,0197044332	19…13	2,4…687	-3…64		1	11	1	1	1	6	4
7	ՕԳՏԱԿԱՐ	6,9493560791	10…58	6,4…277	0,0191846527	12…45	1,9…429	-2…64		1	7	1	1	1	4	2
8	ԿՈՂՔ	6,8835039139	10…06	6,2…336	0,0184331797	12…89	1,9…560	-2…84		1	7	1	1	1	4	4
9	ՉՈՐ	6,7636637688	10…23	5,9…895	0,0171306208	11…38	1,9…473	-2…67		1	7	1	1	1	4	3
10	ՍԽՏՈՐ	6,3359050751	12…53	7,4…632	0,0137825422	22…31	2,9…764	-4…48		1	17	1	1	1	9	2
11	ԱՅԼ	6,2476973534	10…69	4,7…918	0,0124031007	9,3…822	1,9…194	-3…05		1	7	1	1	1	4	2
12	ԵՐԲ	5,7022633553	9,7…688	3,6…754	0,0087241000	6,7…731	1,9…714	-4…86		0	7	0	1	1	4	2
13	ԿԱՐԱԳ	5,4522280693	11…96	4,5…958	0,0075901328	11…11	2,7…324	-5…46		0	15	0	1	1	8	6
14	ՄԵԿ	5,3195905685	13…96	6,3…674	0,0070285043	23…69	4,1…070	-6…82		0	33	0	1	1	18	8
15	ՏՈՐԹ	5,1570272446	9,8…909	3,0…081	0,0061387355	5,5…719	2,1…642	-4…36		0	9	0	1	1	5	3
16	ԿԱՐԵԼԻ	4,8706865311	9,5…774	2,5…197	0,0050632912	4,1…721	2,1…664	-5…71		0	9	0	1	1	5	1
17	ԵԹԵ	4,8323554993	8,8…360	2,1…251	0,0049019610	3,2…856	1,9…031	-4…46		0	7	0	1	1	4	2
18	ԱԼՅՈՒՐ	4,5964088440	10…05	2,7…369	0,0042482889	5,4…753	2,8…682	-6…28		0	16	0	1	1	9	1
19	ՀԵՏՈ	4,4113912582	8,4…950	1,5…791	0,0036917396	1,8…772	1,9…735	-5…57		0	7	0	1	1	4	2
20	ՄԵՋ	4,0961666107	12…14	2,1…531	0,0030262908	3,9…580	3,7…491	-7…32		0	28	0	1	1	16	14
21	ԱՅՆ	3,9713215828	9,1…008	1,1…952	0,0027554536	1,0…807	2,2…400	-6…91		0	11	0	1	1	6	4
22	ՇԱՔԱՐԱՎԱԶ	3,9316301346	9,1…294	1,0…118	0,0026815643	0,9…753	2,2…557	-6…73		0	11	0	1	1	6	1
23	ՈՒ	3,8452041149	9,8…268	1,0…516	0,0025344528	0,9…090	2,6…365	-6…74		0	14	0	1	1	8	3
24	ՋՈՒՐ	3,8516714573	8,4…652	0,8…631	0,0025329280	0,6…076	2,0…725	-6…93		0	9	0	1	1	5	2
25	ՐՈՊԵ	3,8021569252	8,4…947	0,7…245	0,0024485798	0,5…845	2,0…855	-6…30		0	9	0	1	1	5	4
26	ՀԱՄԵՂ	3,7292304039	7,7…175	0,5…840	0,0023222060	0,2…136	1,8…730	-6…00		0	7	0	1	1	4	2
27	ՇԱՏ	3,7010116577	7,7…945	0,5…056	0,0022779044	0,2…255	1,8…713	-6…62		0	7	0	1	1	4	1
28	ԵՒ	2,9783282280	11…89	-1…96	0,0013987558	1,2…717	3,8…475	-8…16		0	31	0	1	1	19	11
29	Է	2,5626451969	12…86	-3…62	0,0010497877	12…84	4,8…977	-10…75		0	52	0	1	1	34	16

concordance collocates plot patterns clusters timeline filenames source text notes

29 entries Row 24 0% n/a

Fig. 2. Identification of collocations of the Armenian culinary verb *bovel* with *Word Smith* based on *Dice*

The collocates of the respective nodes (the specific culinary verbs) were filtered out by leaving only those belonging to the semantic field of the verb, so that, e.g. *ev*, *ajl*, *e*, *mek* (conjunctions, numerals, auxiliary verbs) were excluded from the collocational list (cf. Table 1 and Fig. 3).

Nr.	word	dice	log_l	texts	total	total left	total right
1	bovel	1	750,237488	100	100	100	1
2	nush	0,05787781	48,9712601	3	17	3	1
3	surtch	0,04210527	46,8190536	2	19	2	1
4	hatik	0,03686636	35,5460167	2	15	2	1
5	ynkujz	0,02649007	44,1969795	1	23	1	1

6	hac	0,01970443	19,0641537	1	11	1	1
7	ogtakar	0,01918465	12,8308411	1	7	1	1
8	koghq	0,01843318	12,4937134	1	7	1	1
9	chor	0,01713062	11,8840723	1	7	1	1
10	sxtor	0,01378254	22,0133495	1	17	1	1
11	ajl	0,0124031	9,32557678	1	7	1	1
12	erb	0,0087241	6,77298927	0	7	0	1
13	karag	0,00759013	11,3813334	0	15	0	1
14	mek	0,0070285	23,2703896	0	33	0	1
15	torth	0,00613874	5,56124687	0	9	0	1
16	kareli	0,00506329	4,18470907	0	9	0	1
17	ethe	0,00490196	3,20486999	0	7	0	1
18	aljur	0,00424829	5,41035748	0	16	0	1
19	heto	0,00369174	1,81482518	0	7	0	1
20	medj	0,00302629	3,98375416	0	28	0	1
21	ajn	0,00275545	1,06570208	0	11	0	1
22	shaqaravaz	0,00268156	0,94779778	0	11	0	1
23	u	0,00253445	0,95216781	0	14	0	1
24	djur	0,00253293	0,60625911	0	9	0	1
25	rope	0,00244858	0,50392818	0	9	0	1
26	hamegh	0,00232221	0,29453161	0	7	0	1
27	shat	0,0022779	0,25685433	0	7	0	1
28	ev	0,00139876	1,21252477	0	31	0	1
29	e	0,00104979	12,4343071	0	52	0	1

Table 1. Transliteration of the collocations of *bovel* identified by *Word Smith*

The comparatively small size of the Armenian corpus allowed to take all collocates of the six Armenian culinary verbs into consideration irrespective of their association score (here *Dice*). Nevertheless, it would have been impossible to manually annotate all occurrences of the whole *ARM* corpus. Therefore, the same quantity determination model was also implemented to quantitatively represent the distribution of collocates in the *ARM* corpus (Fig. 3, see also Chapters 2.2 and 3.1).

Wort	Freq	Number	Dice	Model	Designation in 100	Rounded	Zero→1
nush	17	1	0,05787781	0,0446	14,8141265	15	15
surtch	19	2	0,04210527	0,04072864	13,5282344	14	14
hatik	15	3	0,03686636	0,03862175	12,8284198	13	13
ynkujz	23	4	0,02649007	0,03719333	12,35396	12	12
hac	11	5	0,01970443	0,03612183	11,9980586	12	12
chor	7	6	0,01713062	0,03526932	11,7148905	12	12
karag	15	7	0,00759013	0,03456424	11,4806957	11	11
aljur	16	8	0,00424829	0,03396488	11,2816146	11	11
							100

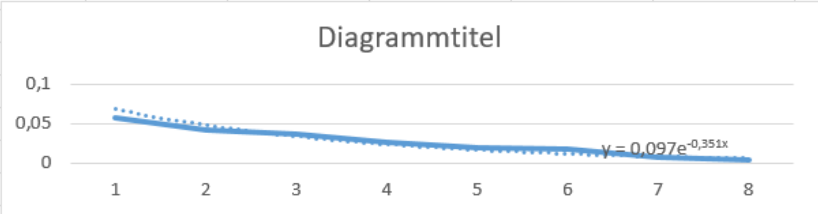


Fig. 3. Annotation quantity determination model for the verb *bovel* in the *ARM* corpus

To maintain consistency with the German and English annotated corpora, 200 occurrences were annotated for more general verbs (hypernyms), and 100 occurrences for more specific verbs (hyponyms). However, due to insufficient occurrences as envisaged by the aforementioned model, the annotation quantity varied but slightly from the initially determined amount. As a result, overall 862 context examples extracted from Armenian culinary texts were annotated for further intralinguistic analysis.

4.1 Neutralization (Generalization) of Annotation Parameters and Reduction of Annotated Occurrences

Sentences	Verb	Ingredient	Substance	Dish	Utensil	Heat intensity	Manner	RES	Heat source	Corpus
Յավի մսի տապակելուց հետո հենց նույն թավայի մեջ հալել 1 ճաշի գդալ կարագ: Ավելացնել սունկը եւ 2-3 րոպե տապակել , մինչեւ դրանք շագանակագույն դառնալ:	տապակել	սունկ	կարագ	N_A	թավա	N_A	N_A	շագանակագույն	N_A	ARM
Սունկը լվանալ, խոնավ սպունգով մաքրել եւ մեծ-մեծ կտրատել: Փոքրիկ սնկերը կարելի է հենց այդպես էլ թողնել: Բեկոնը կտրատել միջին չափի կտորներով, տապակել չոր թավայի մեջ այնքան, մինչեւ խրթխրթան դառնա:	տապակել	սունկ	N_A	N_A	թավա	N_A	N_A	խրթխրթան	N_A	ARM
Յուրաքանչյուր բլիթի կենտրոնում մի փոքր լցոն դնել եւ ձեւավորել կարկանդակներ: Տապակել կարկանդակները երկու կողմերից, միջինից մեծ ջերմության վրա՝ մեծ քանակությամբ յուղ ավելացնելով, մինչեւ ոսկեգույն դառնա:	տապակել	N_A	յուղ	կարկանդա	N_A	միջին ջերմություն	երկու կողմից	ոսկեգույն	N_A	ARM
կարագի կեսը, տապակել սոխը, չկարմրեցնել: Ավելացնել սունկը եւ տապակել 2-3 րոպե: Սոխն ու սունկը հանել թավայից: Թավայի մեջ գցել մնացած կարագը, տաքացնել, մինչեւ փրփուր հայտնվի: Ուժեղ կրակով 4-5 րոպե տապակել միսը, որ բոլոր կողմերից կարմրի: Վրան լցնել սոխն ու սունկը: Ավելացնել մաղադանոսը, սերուցքը, եփել մի քանի րոպե: Մատուցվում է հենց թավայի մեջ, կարելի է մատուցելուց առաջ վրան կոնյակ լցնել եւ այրել:	տապակել	սունկ	բուսայուղ	N_A	թավա	ուժեղ կրակ	N_A	N_A	կրակ	ARM
Սոխը մաքրել, մանր կտրատել, կարելի է դիմել գործիքի օգնությանը: Ավելացնել եգիպտացորենը իր հյութի հետ, աղ անել: Կարելի է մի քիչ ջուր ավելացնել: Ալյուր ավելացնել, խառնել: Ստացվում է պինդ խմոր: Թավան տաքացնել, գդալով խմորից վերցնել, կարմրացնել ձեթի մեջ: Սունկը տապակել սոխի հետ մի քիչ ջրով մոտ 20 րոպե, աղ, պղպեղ անել, բլենդերի մեջ հարել: Ավելացնել կանաչ սոխը, մատուցել:	տապակել	սունկ	ձեթ	N_A	թավա	N_A	N_A	N_A	N_A	ARM
թավայի մեջ 2-3 ճաշի գդալ բուսայուղով: Աղ, պղպեղ անել: Պղպեղները կիսել, սերմատուփը հեռացնել, 1-2 րոպե գցել եռացած ջրի մեջ, հանել եւ հեռացնել կեղեւը: Յավի մեջ տեղադրել նախ պղպեղները, ապա սունկը: Սնկով լցնել նաեւ թեփիկներն ու բդիկները: Յավի փորը կարել: Կարելի է կարել սովորական թելով: Յավի վրա մայոնեզ կամ բուսայուղ քսել, եփել 200-210 աստիճան ջեռոցում 40 րոպե, պարբերաբար վրան լցնել առաջացած հյութը: Յավը գեղեցիկ կեղեւ կունենա: Թելը հեռացնել, կտրատել եւ մատուցել:	տապակել	սունկ	բուսայուղ	N_A	թավա	N_A	N_A	N_A	N_A	ARM
Մեկ-մեկ ավելացնել ձվերը, հնարավոր է, որ երկու ձուն բավական լինի: Ավելացնել 1 ճաշի գդալ ալյուր կամ սպիտակաձավար: Եթե նախընտրում եք խրթխրթան կոտլետներ, կարելի է պաքսիմատ ավելացնել խճողակին: Պատրաստել կոտլետներ, թավալել պաքսիմատի մեջ, տապակել թավայի մեջ մինչեւ կեղեւը ոսկեգույն դառնա: Մատուցել չիլաների կամ խավարտի հետ: Կարելի է մատուցել սոուսի հետ: 150-200 գրամ մանրացված սունկը տապակել, ավելացնել 3-4 ճաշի գդալ թթվասեր, եռացնել, 10-15 րոպե շոգեխաշել: Սոուսը պատրաստ է:	տապակել	սունկ	N_A	N_A	թավա	N_A	N_A	խրթխրթան	N_A	ARM

Table 1. (Partial) annotation of the *ARM* corpus concordances occurrences of the verb *տապակել [tapakel]*

Utilizing the output of the aforementioned model, a targeted amount of examples was extracted from the *ARM* corpus via the *Concordance* function in *Sketch Engine*. As illustrated in Table 1, the extracted occurrences were then annotated according to eight parameters: *Ingredient, Substance*, *Dish*, *Utensil*, *Heat intensity*, *Manner*, *Resultative* (adjectives), and *Heat source*. Chapters 2.2 and 3.1 of this work elaborate in detail the aforementioned annotation parameters and the criteria employed.

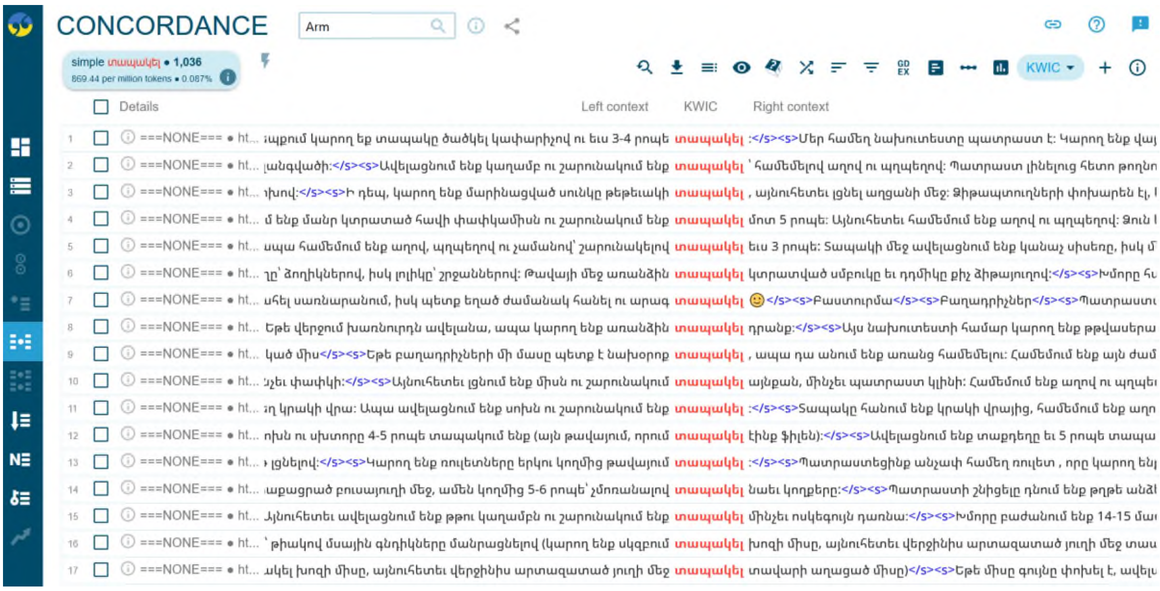


Fig. 4. Concordance lines of the Armenian culinary verb *տապակել [tapakel]* in *ARM* sub-corpus using *Sketch Engine*

As presented in Table 1 above, the annotation starts with the context examples and ends with the name of the corpus, while the verb and the eight annotation parameters are in between the two. However, the Armenian intra-interlinguistic analysis was also based on six of the initially eight annotation parameters—*Substance*, *Utensil*, *Heat intensity*, *Manner*, *Resultative* (adjectives), and *Heat source*[121]—as the values of the *Ingredient* and *Dish* were too heterogeneous to yield significant correlations. Semantically similar values of the aforementioned six annotation parameters were generalized (neutralized) for gaining more significant results and ensure legibility. For instance, *բուսական յուղ [busakan jugh]*, *բուսայուղ [busajugh]*, *ձիթայուղ [dzitajugh]*, *ձիթապտղի ձեթ [dzitatpghi dzet]*, and *արևածաղկի ձեթ [arevatsaghki dzet]* being synonyms or cohyponyms of *oil* were neutralized under the hypernym, i.e. *ձեթ [dzet]* (oil).[122] Despite the fact that the generalization contributed considerably to the acquisition of more informative results while maintaining the fundamental semantics of the annotation, it was necessary to reduce the parameter values in order to achieve a more comprehensive analysis of Armenian culinary collocations. Therefore, the values with very a low number of occurrences were excluded from further statistical analysis, reducing the number of initially annotated examples from 862 to 779.[123] All intralinguistic analysis of the Armenian six cooking verbs, i.e. *Correspondence analysis* (CA), *Multiple Correspondence analysis* (MCA), *Context-Conditional Correlation Graphs* (CCCG), and *Conditional Inference Trees* (CIT) were conducted based on 779 annotated contextual examples.

[121] The Armenian *Heat source* value {կրակ} (fire) does not necessarily mean *Feuer (dt.)* or *fireplace (en.)*. It means rather an abstract word indicating some type of heat source, e.g. stove, gas.

[122] For the full list of all Armenian neutralization (generalization) of all parameter values of the, see Appendix 9.

[123] The same criteria used for reducing the annotated parameter values in the German and English examples were also applied to the Armenian data. On one hand, the significant difference in the number of preceding and succeeding collocates was considered; on the other, maximizing coverage of all frequencies was prioritized, ensuring minimal loss for further analysis and visualization. For instance, the *Heat source* parameter was reduced to the value {krak} (fire) as the difference in frequencies between {krak} and the succeeding collocate {vararan} (hearth) is large (60 annotated occurrences). The parameter *Substance*, for example, was reduced to the value {sous}(sauce) covering still 97% of the annotated occurrences. The values for the parameters *Heat intensity* and *Utensil* were 100% kept for the analysis since they did not meet the aforementioned reduction criteria.

4.2 *Correspondence Analysis* (CA) vs. *Multiple-Correspondence Analysis* (MCA) of the Armenian Culinary Verbs

Correspondence Analysis (CA) contributed to the visualization of the distribution of the Armenian culinary verbs with regard to relevant parameters. As noted at the beginning of Chapter 1, one of the main objectives of this research was to provide a detailed description of German, English, and Armenian culinary verbs through a corpus-based collocational analysis, treating the verbs as nodes and the annotated parameter values as their collocates. In *CA* analysis of the annotated occurrences extracted from the *ARM* corpus, the six Armenian verbs were considered as dependent while the parameter values as independent variables. *CA* illustrates correlations between the independent and dependent variables. *Multiple Correspondence Analysis* (MCA), on the contrary, first of all demonstrates clusters of either strongly correlated independent variables or similar word profiles. The verbs are then distributed around the identified clusters as qualitative supplementary variables.[124]

Both the *CA* and the *MCA* were based on the initially annotated overall 862 contextual examples which were reduced to 779 as mentioned above. At this point, however, it is worth reiterating the conventions for interpreting an *M(CA*) graph: strong correlations between the verb (dependent variable) and the corresponding parameter value(s) (independent variables) are indicated by the sharpness of the imaginary angle they form—provided that both elements are positioned farther from the origin of the two-dimensional graph. (The circles and lines illustrating these angles were added by me to facilitate interpretation.)

[124] For a thorough explanation of the M(CA) analysis, see Chapter 2 on the examples of the German corpus description as well as (Alberti 2013).

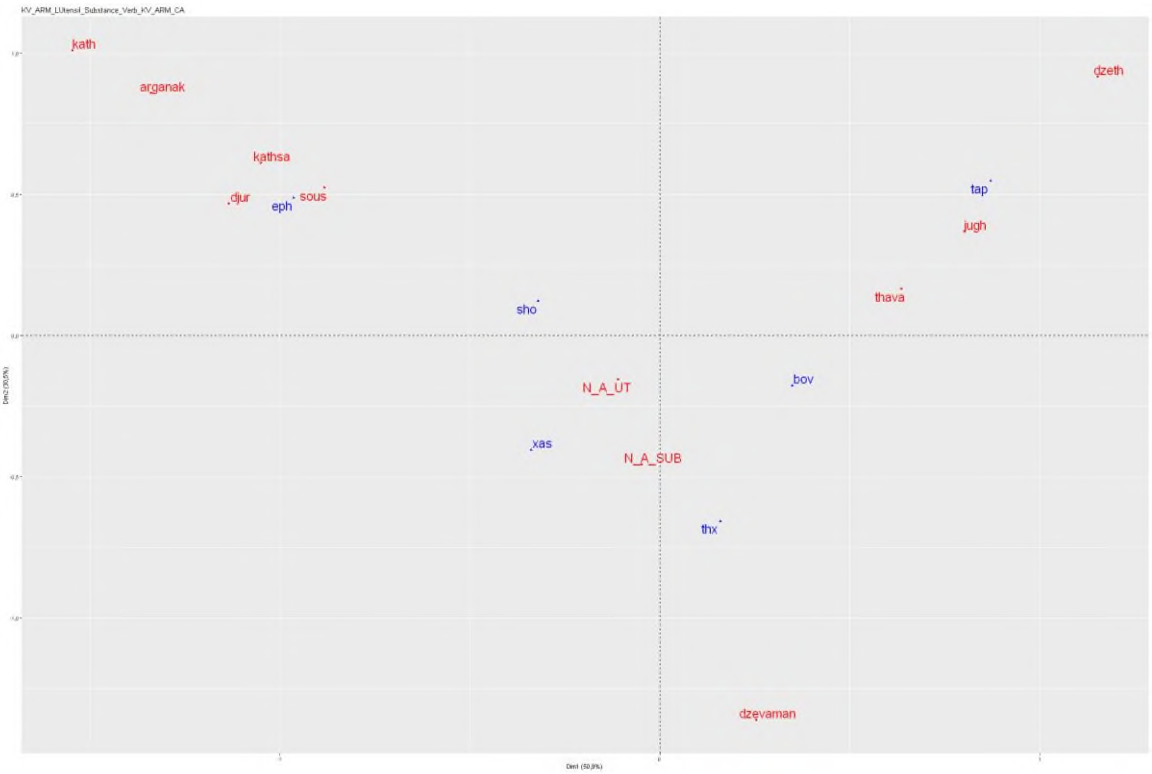


Fig. 1a. CA: Armenian culinary verbs (parameters *Substance* and *Utensil*)

The *CA* demonstrates strong correlations between the dependent variables (culinary verbs marked in blue) and the independent ones (*Substance* and *Utensil* parameters values marked in red) (Fig. 1a). The Armenian culinary verb *ephel* (partially also *shogexashel*) correlates strongly with *water*-like *Substance* parameter values such as {djur, sous, arganak, kath} (water, sauce, stock/broth, milk) as well as *Utensil* parameter one {kathsa} (pot) (examples 1-3).

1. Կարտոֆիլը *կաթսայի* մեջ աղ արած *ջրով* մի քանի րոպե *եփել*, որպեսզի կտորները դառնան փափուկ, բայց չփլվեն: *(ARM)*

   [In a *pot*, cook* the potatoes in boiling salted *water* until the pieces are soft, however, do not fall apart].

2. Բանջարեղենը լցնել *արգանակով կաթսայի* մեջ, *եփեք* 5 րոպե: *(ARM)*
   [*Cook*/boil** the vegetables in *broth/stock* for 5 minutes].

3. Դեղնուցների վրա ավելացնել շաքարավազը և լցնել *կաթի* մեջ: Խառնել, դնել կրակի վրա և *եփել* մինչև մի փոքր թանձրանա: *(ARM)*
   [Add the sugar onto the egg yolks and pour into the *milk*. Stir and *boil*/cook** until it's a bit thicker].

Among other strong correlations between the dependent and independent variables, the verb *tapakel* (to fry) is observed as to correlate strongly with {thava}(pan) x {jugh, dzeth} (ghee, oil) of the *Utensil* and *Substance* parameter values respectively (examples 4-6). [125] and *tapakel* being the strongest one judging from both its distance from the intersection as well as the sharpness of the angle they form.

4. Տաքացնել *արևածաղկի ձեթը թավայի* մեջ և *տապակել* կարտոֆիլը մինչև ոսկե դարչնագույնը: *(ARM)*
   [Heat the *sunflower oil* in a *pan* and *fry* the potatoes until golden brown].

5. *Թավայի* մեջ առանձին *տապակել* կտրատված սմբուկը և դդմիկը քիչ *ձիթայուղով*: (ARM)
   [Separately, in a *pan fry* the egg-plant slices and courgettes with littel *olive oil*].

6. Տաքացնել յուղը *թավայի* մեջ և *տապակել* չոր սմբուկի կտորները 3-5 րոպե միջին ջերմության վրա, ավելացնելով աղ երկու կողմերից: *(ARM)*
   [Heat the *ghee* in a *pan*, dry the egg-plant slices and *fry* them over medium heat for 3-5 minutes by adding salt on both sides].

The Armenian culinary verb *thxel* (to bake) correlates strongly with the *Utensil* and *Substance* parameter values {dzevaman} (baking (form) and {N_A_SUB} respectively. Interesting is the distribution of the *Substance* and *Utensil* parameter values {N_A_SUB}x{N_A_UT} correlating strongly with both *xashel* (to boil*) and *thxel* (to bake). However, the strong correlation between the *Substance* parameter value {N_A_SUB} with the verbs *xashel* and *thxel* denotes two different things. The verb *xashel* implicates {djur} (water) as prototypical cooking substance inherent in the meaning of the verb. Therefore, water does not not have to be mentioned in the context so that *no values* of *Substance* parameter are verbalized (example 7). The strong correlation between the verb *thxel* and {N_A_SUB}, on the contrary, corresponds to the *absence of any cooking substance* (example 8).

7. *Խաշել* ճակնդեղը, կարտոֆիլը, գազարը: *(ARM)*
   [*Boil**, *cook** the beetroot, potato and the carrot]. (No mention of any cooking substance as water is *implicated as prototypical cooking substance* for the verb *xashel*).

[125] The *Substance* value {dzeth} (oil) was neutralized and included such variables as {dzithaptghi dzeth} (olive oil), {dzitajugh} (olive oil), {busakan jugh} (vegetable oil), {arevatsaghki dzeth} (sunflower oil), etc. For a full list of neutralization of the Armenian annotation parameter values see Appendix 8.

8. *Թխել* խմորը նախապես տաքացված ջեռոցում մինչև 20-25 րոպե: *(ARM)*
[*Bake** the dough in the preheated oven for 20-25 mintes]. (No verbalization of any substance parameter values denoting its *absence*).

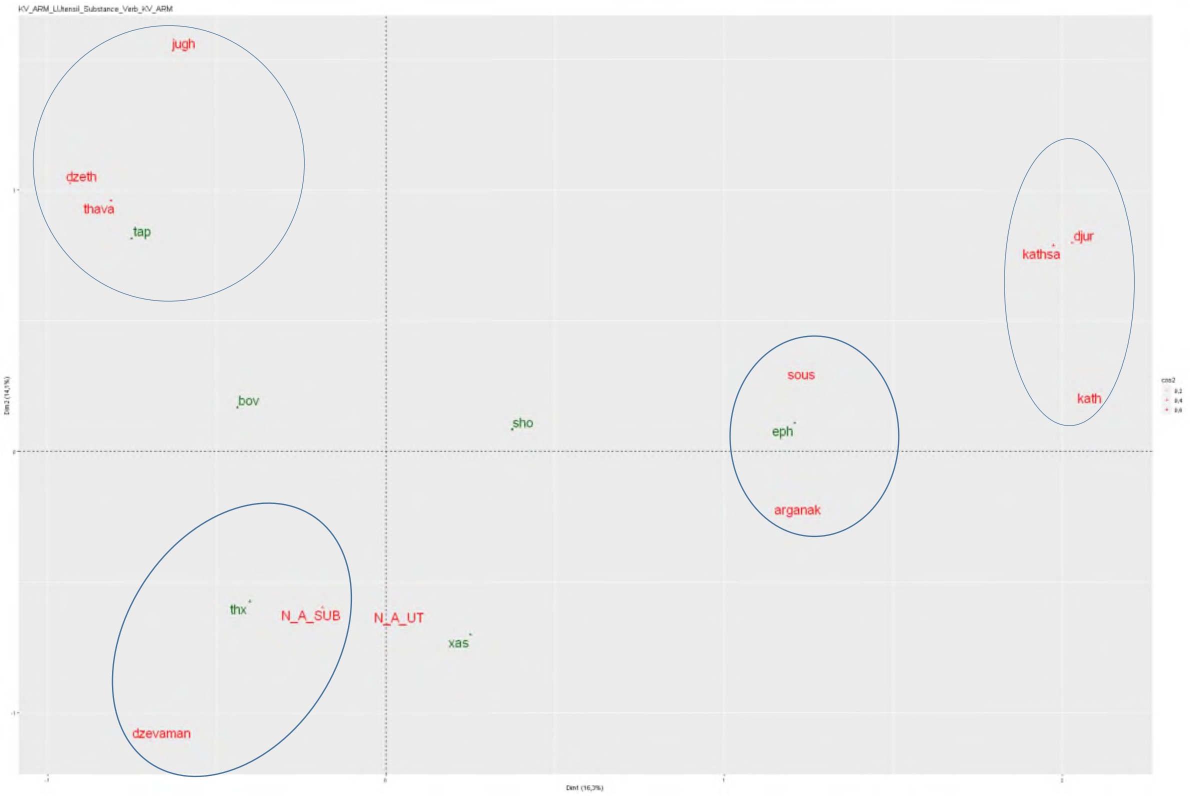


Fig. 1b. MCA: Armenian culinary verbs (parameters *Substance* and *Utensil*)

The distribution on the *MCA* graph (Fig. 1b) mirror the *CA* above (Fig. 1a) in that it splits the graph into three parts grouping three distinctive clusters around the parameter *Substance*. The cluster at the upper-left corner illustrates strong correlation between *oil*-like cooking substances and {thava} (pan), the one in the lower-left side exhibits strong correlation between the *absence* or *no verbalization of any substance* and {thava} (baking form) leaving the right side to *water*-like substance and {kathsa} (pot).

The *Substance* parameter values {sous, arganak} (sauce, broth/stock) appear together on this *MCA* graph (Fig. 1b) due to their profile similarities, i.e. having the same syntagmatic distribution. The cluster right to it, illustrates strong correlation between the *Substance* and *Utensil* parameter values {jur}x{kathsa} (water; pot) as well as {kath}x{kathsa} (milk; pot) where {jur, kath} have similar profiles (example 1 above). The same way, the *Substance* and

Utensil parameter values {jugh}x{thava} (ghee; pan) as well as {dzeth}x{thava} (oil; pan) correlate strongly forming a cluster on the upper-left side of this *MCA* graph with {jugh}x{dzeth} (ghee, oil) having profile similarities (examples 4-6 above). The *Substance* and *Utensil* parameter values {N_A_SUB}x{dzevaman} (baking dish, baking form) form another cluster of strongly correlated independent parameter values. {Jugh} (ghee) and {dzevaman} (baking dish/baking form) values of the *Substance* and *Utensil* parameters respectively are displayed here as outliers due to overestimation of relevant parameters in *MCA*. Even though the dependent variables, e.g. the verbs *ephel* (to cook*, boil*), *tapakel* (to fry) and *thxel* (to bake) appear respectively near the right upper-left and lower-left clusters on this MCA graph (Fig. 1b) as qualitative supplement variables, the distribution of the independent variables would still be the same if we were to exclude the verbs from the analysis.

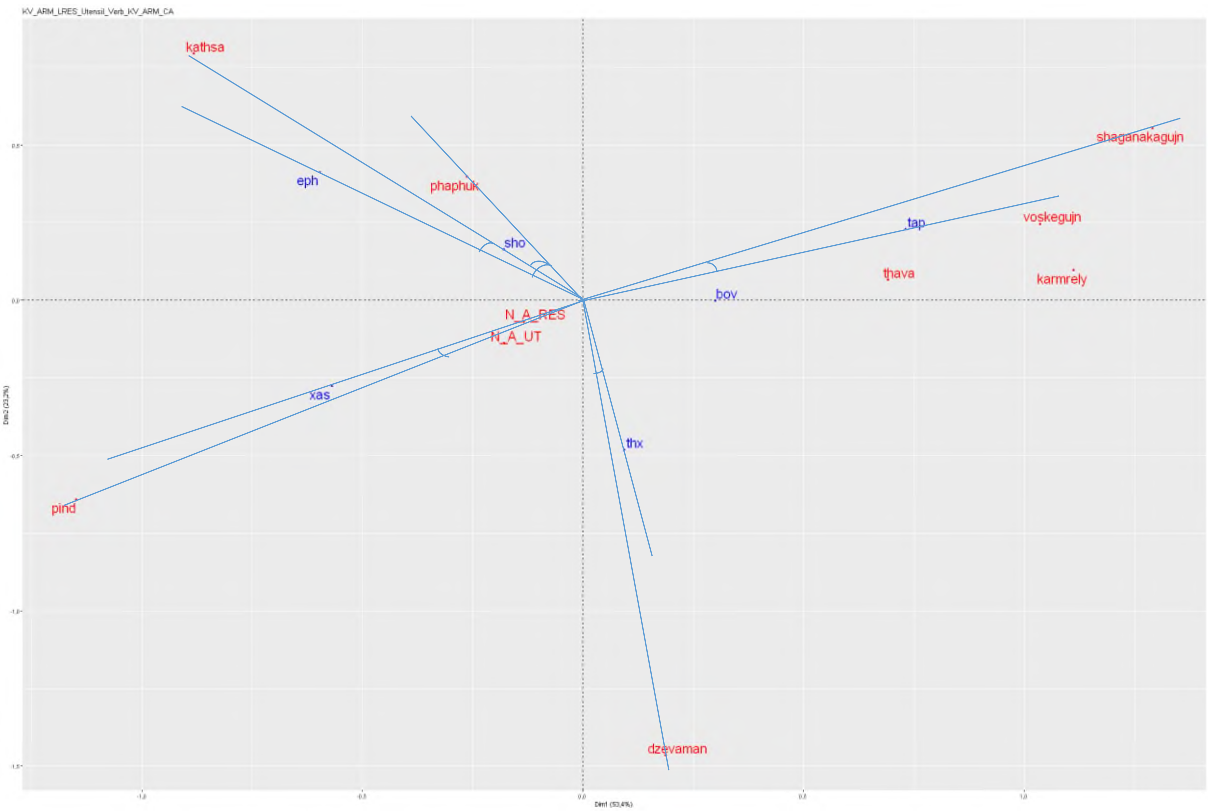


Fig. 2. CA: Armenian culinary verbs (parameters *Utensil* and *Resultative* (adjectives))

The *CA* graph (Fig. 2) illustrates the distribution of the Armenian culinary verbs with relation to the *Utensil* and *Resultative* (adjectives) parameters. Among strong correlations between the independent and dependent variables, the verbs *ephel* (to cook in water*, boil*) and *shogexashel* (to cook with little water*, steam*, braise*) correlate strongly with the values {khatsa} (pot) and {phaphuk} (soft) of the *Utensil* and *Resultative* (adjectives) parameters respectively (example 9-11). However, the *Utensil* parameter value {khatsa} “pulls” the verb *ephel* much farther from the intersection than the *Resultative* (adjectives) value {phaphuk} and forms a much sharper angle with the former than the latter by illustrating a stronger correlation with {khatsa} (pot) than with {phaphuk} (soft).

9. Շաքարավազը լցնում ենք *կաթսայի* մեջ, ավելացնում ենք ջուրը և դնում թույլ կրակի վրա ու *եփում* մինչև շաքարավազը լուծվի ջրում: *(ARM)*
   [We pour the sugar into a *pot*, add the water and put* on low heat and *cook*/boil** till the sugar dissolves in water]

10. Կարտոֆիլը *կաթսայի* մեջ աղ արած ջրով մի քանի րոպե *եփել*, որպեսզի կտորները դառնան *փափուկ*, բայց չփլվեն: *(ARM)*
    [In a *pot boil*/cook** the potatoes in boiling salted water, so that the pieces become *soft* but do not fall apart].

11. Կրակն իջեցնել և *շոգեխաշել* մինչև բանջարեղենի *փափկելը*: (ARM)
    [Reduce the heat and *braise*/steam** until the vegetables are *soft*].

The verb *shogexashel*, on the contrary, is observed as to strongly correlate with more the *Resultative* (adjectives) parameter value {phaphuk} (soft) (example 12) than the *Utensil* one {khatsa} (pot). The verb *xashel* correlates strongly with the *Resultative* (adjectives) parameter value {pind} (hard) while *thxel* does with the *Utensil* one {dzevaman} (baking (form) (example 13).

12. Ձուն պինդ խաշել, սառեցնել, օղակ-օղակ կտրատել: *(ARM)*
    [Hard *boil** the eggs, let them cool down and cut into slices].

13. Յուղաթղթապատ *ձևամանի* մեջ շարել խմորները և *թխել* 200 աստիճան տաքացրած ջեռոցում 10 րոպե: *(ARM)*
    [Put the pieces of dough on buttered *baking form* and *bake* in oven at 200 degrees for 10 minutes]

The Armenian culinary verb *tapakel* is “pulled” by the values {shaganakagujn, voskegujn, karmrely} (brown, golden, until brown/red) x {thava} (pan) of the *Resultative* (adjectives) and

Utensil parameters respectively (examples 14-16), thus illustrating strong correlations between them.

14. Ավելացնել սունկը և 2-3 րոպե *տապակել* , մինչև դրանք *շագանակագույն* դառնալ: *(ARM)*
[Add the mushroom and *fry* for 2-3 minutes until *brown*].

15. *Թավայի* մեջ բուսայուղն ու կարագը խառնել, միսը տապակել, մինչև *ոսկեգունի*: (ARM)
[Mix the oil and butter in a *pan* and fry the meat until *golden*].

16. *Տապակել* կոտլետները *թավայի* մեջ մինչև կարմրելը, ապա դրանք դնել ջեռոցի խոր տապակի մեջ: *(ARM)*
[*Fry* the patties in a *pan* until brown/red* and then place them in a deep oven form].

Even though the verb *bovel* (roast*) is placed nearer the intersection than *tapakel* (to fry), *bovel* also correlates strongly with the *Utensil* and *Resultative* (adjectives) parameter values {thava} (pan) x {shaganakagujn, voskegujn, karmrely} (brown, golden, until brown/red) respectively (examples 17-18).

17. *Թավայի* մեջ կարագով բովել ալյուրը, երբ *ոսկեգույն* դառնա, տաք կաթ ավելացնել:
[In a pan *roast** the flour with butter. Add warm milk as soon as the flour gets *golden*]. *(ARM)*

18. Յուղը հալեցնել *թավայի* մեջ, ալյուրը լցնել տաք յուղի մեջ: Անընդհատ խառնելով բովել, մինչև ստացվի *ոսկեգույն* անուշաբույր զանգված: *(ARM)*
[Melt the butter in a pan, add the flour in hot ghee. *Roast** by constantly stirring until it becomes golden and smells good*/nice*].

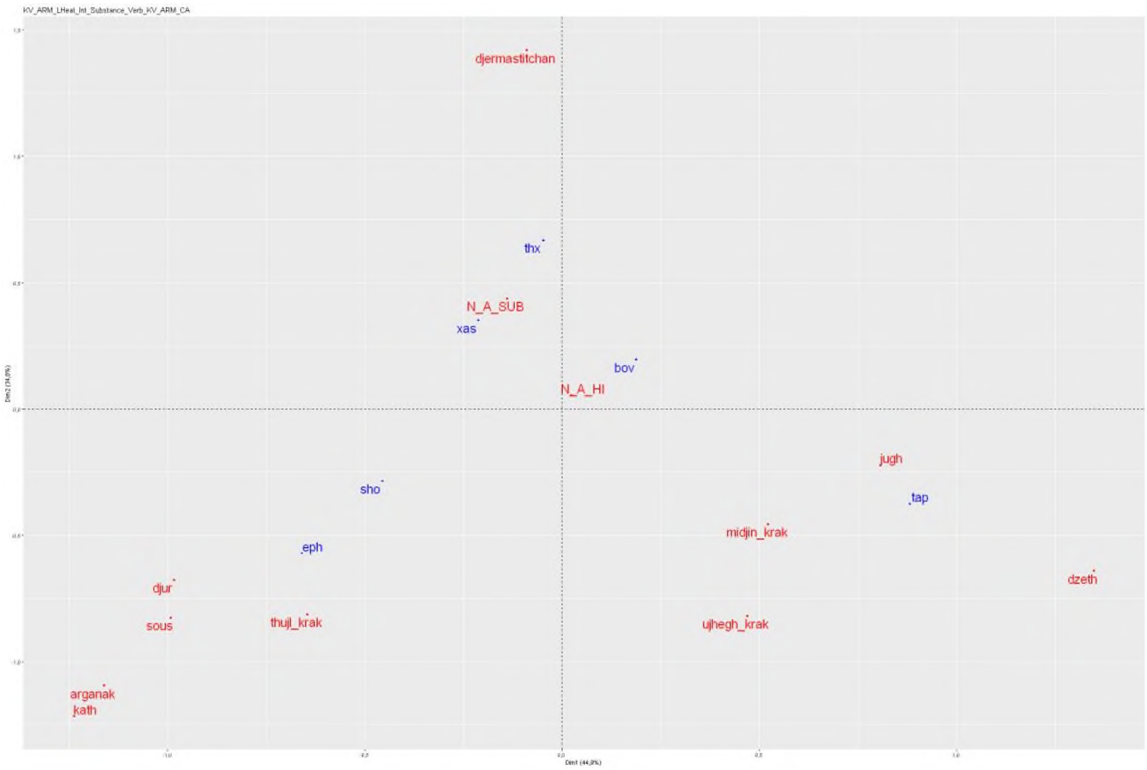


Fig. 3a. CA: Armenian culinary verbs (parameters *Substance* and *Heat intensity*)

Among the strong correlations illustrated on the *CA* graph (Fig. 3a), the verbs *ephel* (to cook*/boil*) and *shogexashel* (to steam*/braise*) correlate strongly with both {djur, sous, arganak, kath} (water, sauce, broth/stock, milk) values of the *Substance* and the {thul krak} (low heat) of the *Heat intensity* parameters (examples 19-23).

19. Լցնել հավի վրա սառը *ջուր*, բերել եռման և *եփել* 30 րոպե: *(ARM)*
[Pour cold *water* on the chicken, bring to boiling and *cook** for 30 minutes].

20. *Արգանակը* հասցնում ենք եռման աստիճանի, ավելացնում ենք կարտոֆիլն ու 10-15 րոպե *եփում* ենք մինչև կարտոֆիլը փափկի: *(ARM)*
[Bring the *stock* to boiling, add the potatoes and *cook** for 10-15 minutes until soft].

21. Զանգվածը խառնելով ավելացնենք 1,5 բաժակ կամ 300 գրամ *կաթ*, քիչ-քիչ լցնել և դնել գազօջախին *մարմանդ կրակի*[126] վրա *եփելու*: *(ARM)*
[Stirring add 1,5 cups or 300ml of *milk* into the mass and put on the stove to cook* *over low heat*].

[126] The *Heat intensity* parameter values were neutralized and {thul krak}(low heat) included also {marmand krak} (small* fire).

22. Թռչնամիսը կիսել կամ կտորների վերածել, խորը կաթսայի մեջ մի փոքր *ջրով շոգեխաշել*: (ARM)
[Slice the poulty into pieces and *braise*/cook*/steam** in a deep pot with little *water*].

23. *Սոուսը* լցնել կոտլետների վրա, *շոգեխաշել* մոտ կես ժամ *մարմանդ կրակի* վրա: (ARM)
[Pout the *sauce* onto the patties and *braise** for about half an hour *over low heat*].

The next strong correlation exposed on the *CA* (Fig.3a) graph of the Armenian annotated data is between the verb *tapakel* (to fry) and the values {jugh, dzeth} (ghee, oil) as well as {midjin krak} (medium heat) (partially also {ujegh krak} (high heat)) of the *Substance* and *Heat intensity* parameters respectively (example 24).

24. Թավայի մեջ *բուսայուղն* ու *կարագը* խառնել, միսը տապակել *միջին կրակի վրա*, մինչև ոսկեգունի: *(ARM)*
[Mix the vegetable *oil* and the *butter* in a pan, *fry** the meat on low heat until it becomes golden].

The Armenian culinary verb *thxel* correlates strongly with the *Heat intensity* parameter value {djermastichan}[127](temperature/degree) placed farther from the intersection and forming a very sharp angle with it (example 25). The verb *thxel* and *xashel* correlate with the {N_A_SUB} value of the *Substance* parameter.

25. Ջեռոցի տապակը ծածկել յուղաթղթով և խմորը հավասարապես տարածել: *Թխել* 4-5 րոպե 180 °C *ջերմաստիճանում*: (ARM)
[Cover the baking tray with parchment paper and spread the dough evenly. *Bake* for 4-5 minutes in oven at *180°C*].

[127] The *Heat intensity* parameter value {djermastichan} (temperature/degree) could not be generalized (neutralized) neither into any of the three heat intensities — low, medium, and high – as {djermastichan}warries throughout the Armenian recipes in the temperature range of 80°-250° making it impossible to objectively define the intensity of the heat. {Djermastichan} is a verb-specific parameter value correlating strongly with exclusively the Armenian culinary verb *thxel* (to bake) specifically mentioning the exact temperature of baking.

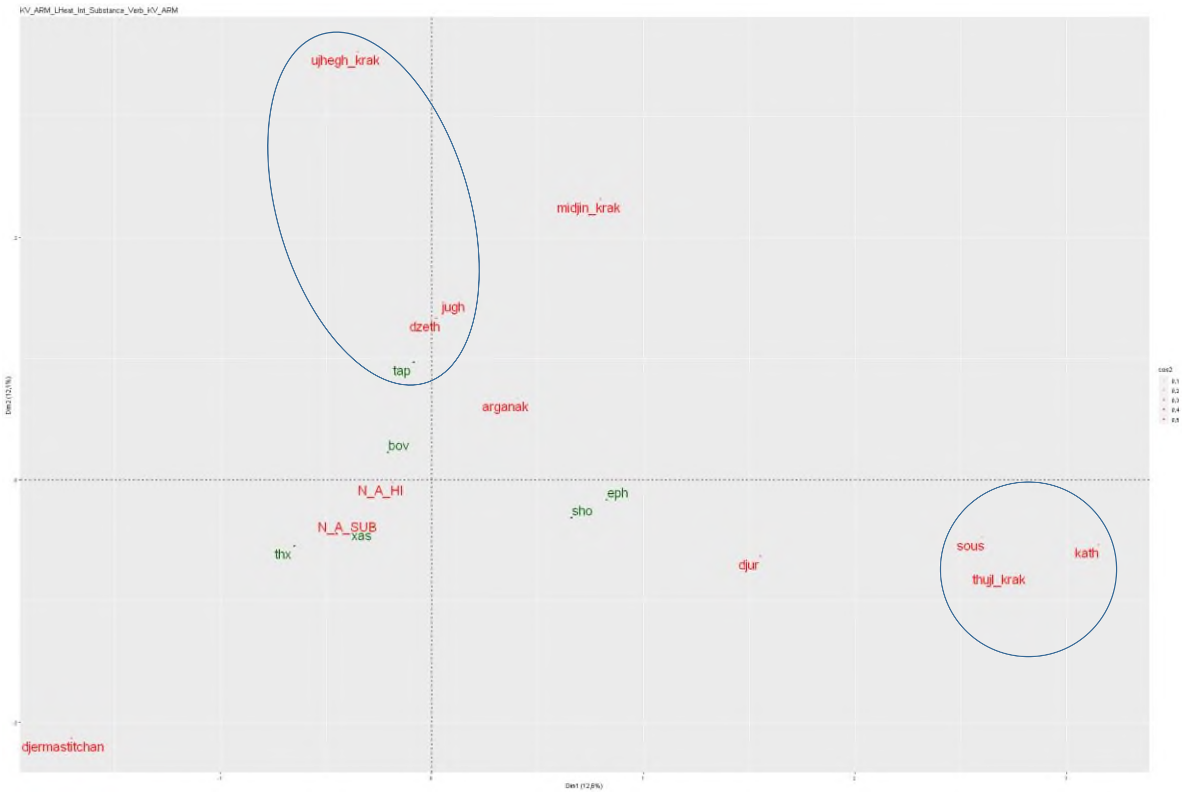


Fig. 3b. MCA: Armenian culinary verbs (parameters *Substance* and *Heat intensity*)

The *MCA* (Fig.3b) illustrates a cluster of strongly correlated *Substance* and *Heat intensity* parameters values {sous, kath} (sauce, water) x {thujl_krak} (low heat) showing their co-occurrence in the same contexts. The upper-right side of the *MCA* graph (fig. 3b), identifies another cluster of the values {ujegh krak} (high heat) x {jugh, dzeth} (ghee, oil) due to their strong correlation with {jugh, dzeth} having similar profiles. {Ujegh krak} is overestimated generating an outlier and being placed much farther from the intersection (example 26). The verb *tapakel* is near the cluster as a qualitative supplement variable.

26. Կարտոֆիլը լցնել բուսական *յուղով* տապակի մեջ, և այն տապակել *բարձր կրակի* վրա: (ARM)
    [Add the potatoes in a pan with the *ghee* and *fry over high heat*].

The *Heat intensity* parameter value {djermastitchan} (temperature/degree) is an outlier for the verb *thxel* (to bake) (example 25 above).

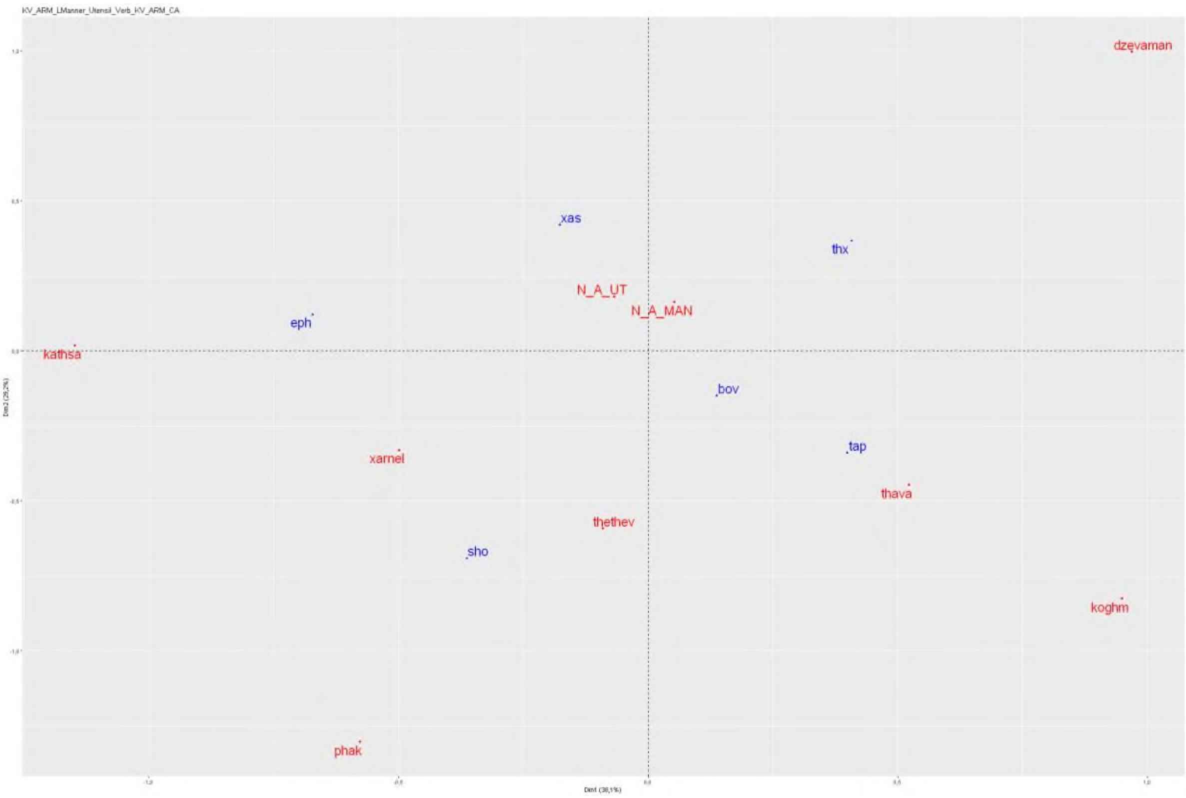


Fig. 4a. CA: Armenian culinary verbs (parameters *Utensil* and *Manner*)

Among the most salient correlations between the dependent and independent variables illustrated on the *CA* graph (Fig. 4a) is the correlation of the verb *thxel* (to bake) and the *Utensil* parameter {dzevaman} (baking form) placed at the farther end of the right-hand corner of this scatterplot. The verb *tapakel* (to fry) creates a strong correlation with the *Manner* parameter value {koghm} (side) and the *Utensil* parameter {thava} (pan). *Shogexashel* (to steam*, to cook in little water*, to braise*) correlates with the *Manner* parameter values {thethev} (light), {phak} (closed), and {xanel} (stir). The latter also "pulls" the verb ephel down by creating a strong correlation with it togher with the *Utensil* parameter value {kathsa} (pot).

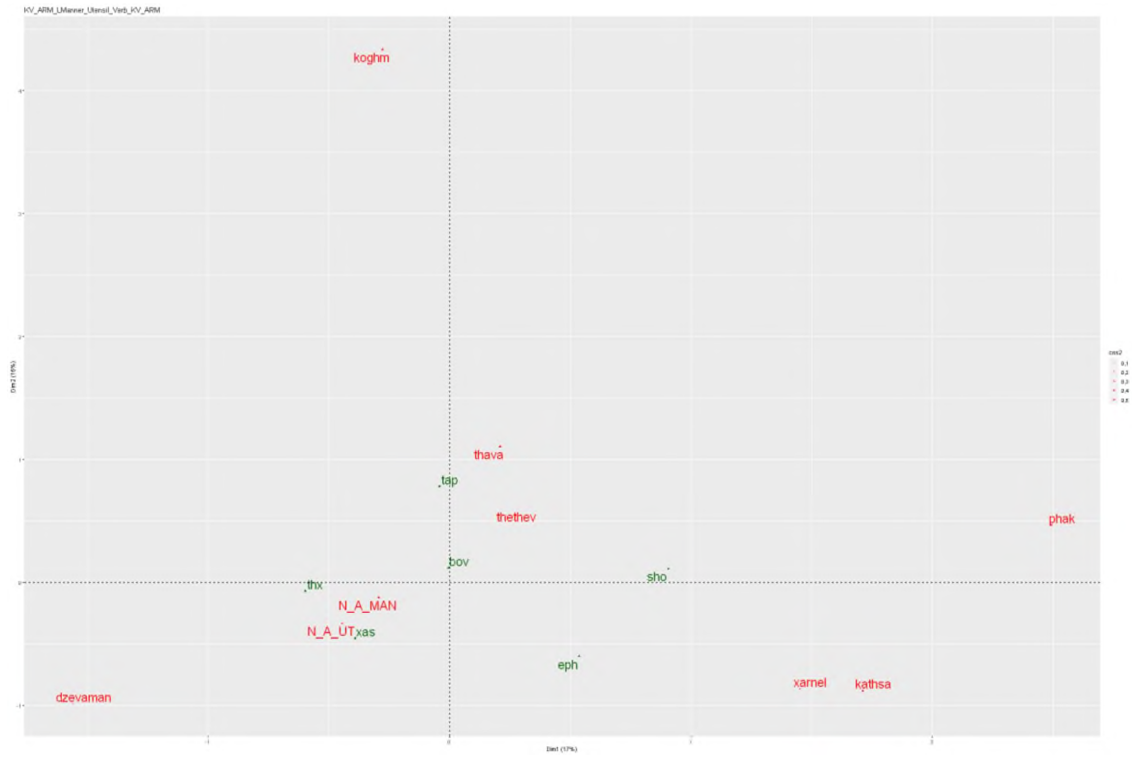


Fig. 4b. MCA: Armenian culinary verbs (parameters *Utensil* and *Manner*)

Fig. 4b illustrates three outliers as a result of overestimation of relevant parameters irrespective of its low number of occurrences due to *MCA* visualization peculiarities. *Utensil* parameter values {koghm} (side_s) on the upper-left, {dzevaman} (baking form) on the lower-left and the *Manner* parameter value {phak} (covered) are illustrated as outliers. A small clusters of strongly correlated independent variables, i.e.{xarnel}x{kathsa} (stir; pot)[128] is displayed on the lower-right side of the *MCA* graph. The *Manner* parameter {phak} (covered) could also be included in this cluster due to its profile similarity with {xarnel} (stir). The verbs *ephel* and *shogexashel* are near the cluster as qualitative supplement variables (example 27-28).

27. Բոլոր բաղադրիչները լավ խառնել, ավելացնել 100 մլ ջուր և եփել ցածր կրակի վրա, *կափարիչը ծածկած շոգեխաշել* մոտ 35 րոպե: *(ARM)*
[Mix all the ingredients, add 100ml of water and *cook*/boil** over low heat by *steaming*/braising** it for about 35 minutes with the *lid closed*].

[128] The *Utensil* parameter value {kathsa} (pot) is positioned on the X axis of the CA graph as it is mutually pulled by both the verb *ephel* and *shogexashel*, even though it is displayed as to strongly correlate only with *ephel* (Fig. 4a).

28. Լցնել կոլոլակների վրա այս սոուսը, *փակել կափարիչով* և *շոգեխաշել* դրանք 40 րոպե: *(ARM)*
[Pour the sauce on the meat balls, *cover it/put the lid on* and *braise*/steam** them for 40 minutes].

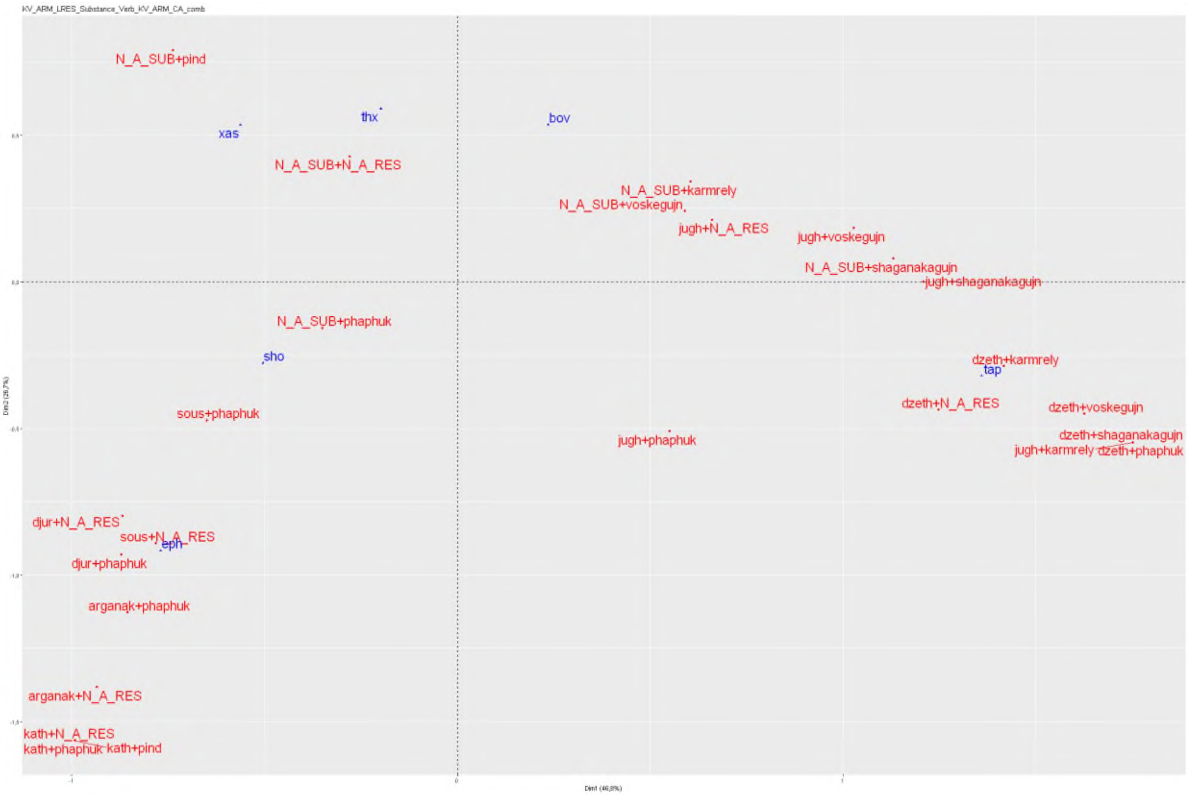


Fig. 5a. CA with concatenated *Resultative* (adjectives) and *Substance* parameters

CA with concatenated parameters exposes strong correlations between pairs of selected parameter values and the verbs allowing to also highlight the specific contexts the verb is used in. For instance, the *CA* with concatenated parameters (Fig. 5a) exhibits strong correlations between the combinations of the *Substance* and *Resultative* (adjectives) parameter values {N_A_SUB} x {pind} (hard) and the verb *xashel* (example 29) as well as {N_A_SUB} x {N_A_RES} and *thxel* (example 30).

29. Ձվերը *պինդ խաշել*, կտրատել սոխը, վարունգը և նեխուրը փոքր խորանարդիկներով: *(ARM)*
[Hard boil the eggs, dice the onion, cucumber and celery].

30. Թխել ջեռոցում 10 րոպե 200 աստիճանում: *(ARM)*
[Bake in the oven for 10 minutes at 200°].

The Armenian culinary verb *bovel* creates a strong correlation with {N_A_SUB} x {karmrely} (become red*, brown), {N_A_SUB}x {voskegujn} (golden), {jugh} (ghee) x {N_A_RES}, {N_A_SUB}x{shaganakagujn} (brown), and {jugh} (ghee) x{shaganakagujn} (brown) (31-34).

31. 5-7 րոպե չոր տապակի վրա *բովել* ընկույզը, որ մի փոքր *շագանակագույն* դառնան: *(ARM)*
    [*Roast* the walnuts in a pan without oil until it is light *brown*].

32. Ալյուրը *բովել*, այնուհետև ավելացնել *յուղը*, շարունակել բովել, մինչև ալյուրը *ոսկեգունի*: (ARM)
    [*Roast* the flour, add the ghee, continue roasting until *golden*].

33. Սոուսի համար նախապատրաստված քունջութը *բովել*, մինչև որ ստանա բաց *դարչնագույն* երանգ: *(ARM)*
    [*Roast* the sesame seeds for the sauce until it gets light *brown*].

34. Թավայի մեջ կարագով *բովել* ալյուրը, երբ *ոսկեգույն* դառնա, տաք կաթ ավելացնել: *(ARM)*
    [*Roast* the flour in a pan with butter until *golden* and add warm milk].

The *CA* below (Fig. 5b), however, exhibits a strong correlation between the verb *bovel* with the *Substance* parameter value {jugh} (fat) judging from the distance of the two from the intersection and the sharpness of the angle they form together, whereas the *Resultative* (adjectives) {voskegujn} (golden) is “pulled” by the verb *tapakel* (to fry) therefore generating a slightly weaker correlation with *bovel* (to roast*).

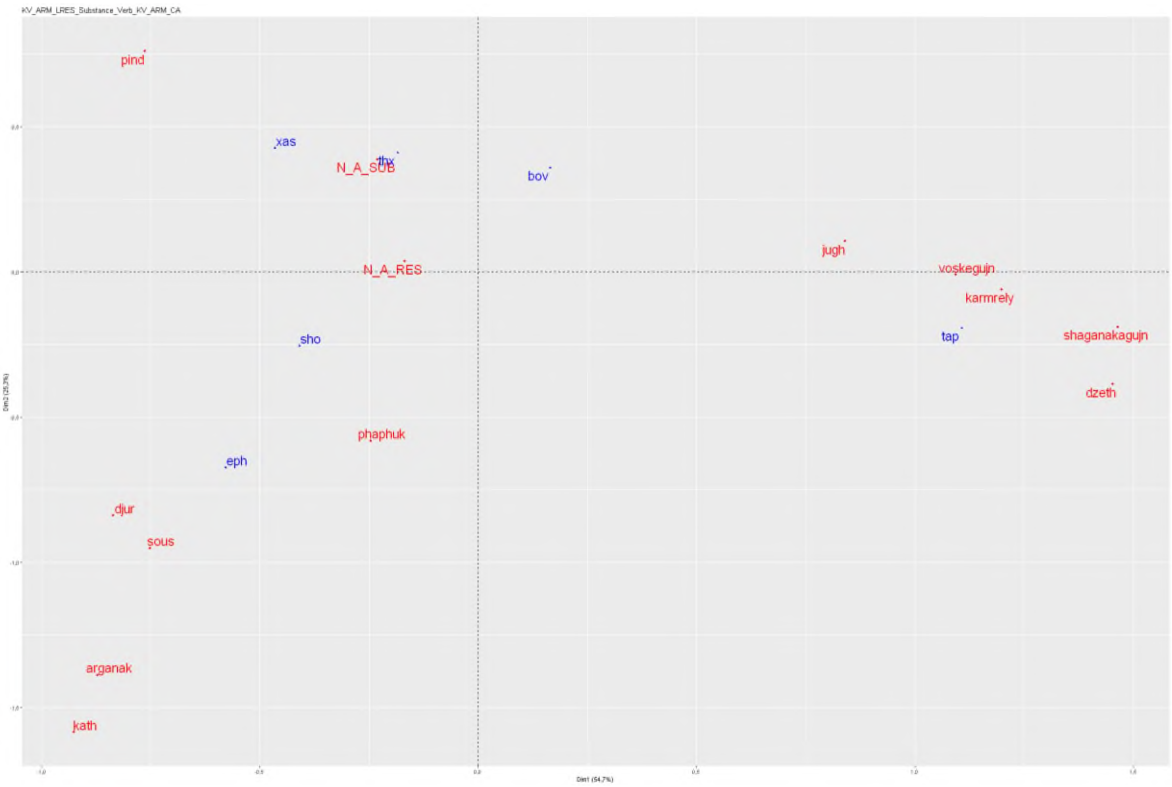


Fig. 5b. CA: Armenian culinary verbs (parameters *Resultative* (adjectives) and *Substance*)

If the *CA* (Fig. 5b) exposes the verb *tapakel* (to fry) to correlate with the *Substance* parameter value {dzeth} (oil) and *Resultative* (adjectives) ones {karmrely, shaganakagujn, voskegujn} (until red*/brown, brown, golden), the *CA* with concatenated parameters (Fig. 5a) reveals strong correlations between {jugh} (ghee) x {karmrely} (becoming red*, brown), {dzeth} (oil) x {phaphuk}(soft), and {jugh} (ghee) x{phaphuk} (soft) too. The verb *bovel* is illustrated on the *CA* graph (Fig.5b) as to strongly correlate with both the *Substance* parameter values {jugh} (fat) and {N_A_SUB} denoting *absence of any cooking substance.* The verbs *shogexashel* and *ephel* correlate strongly with the *Resultative* (adjectives) parameter value {phaphuk} (soft) as well as {djur, sous, arganak, kath} (water, sauce, broth/stock, milk) values of the *Substance* parameter (examples 35-38).

35. Մի քիչ *եռջուր* լցնել, կրակն իջեցնել, *շոգեխաշել* 5-7 րոպե՝ մինչև խնձորը *փափկի*, բայց ձևը չկորցնի: *(ARM)*
    [Add a little boiling water, reduce the heat, steam*/sauté*/braise* for 5-7 minutes until the apples are soft but have preserved their shape].

36. *Եփել* մինչև մսի *փափկելը*: *(ARM)*
    [*Cook*/boil** until the meat becomes *soft*].

37. Շաքարավազը խառնել կակաոյի հետ, լցնել տաք *կաթը* և *եփել*, մինչև շաքարավազը լուծվի: *(ARM)*
[Mix the sugar with cocoa powder, add warm *milk* and *cook*/boil** until the sugar dissolves].

38. Ջուր ավելացնել և *եփել* մինչև թանձրանա, պարբերաբար խառնելով: *(ARM)*
[Add water and boil*/cook* until it thickens by constantly stirring].

Nevertheless, the *CA* with concatenated parameters (Fig. 5a), exposes also strong correlations between the combinations of *Substance* and *Resultative* (adjectives) parameter values {N_A_SUB} x {phaphuk} (soft), and {sous} x {phaphuk}(sauce; soft) and the verb *shogexashel* (to braise*, to steam*, to cook in little water*) and *ephel* (to cook*, to boil*). The Armenian culinary verb *ephel* is displayed on the *CA* with concatenated parameters (Fig. 5a) as to also correlate strongly with the {djur} (water) x {N_A_RES}, {sous} (sauce) x{N_A_RES}, {djur} x {phaphuk} (water; soft), {arganak} x {phaphuk} (stock/broth; soft), {arganak} (stock/broth) x {N_A_RES} as well as {kath} x {pind} (milk; hard), {kath} (milk) x {N_A_RES}, and {kath} x {phaphuk} (milk; soft) combinations of the *Substance* and *Resultative* (adjectives) parameter values. The *CA* (Fig. 5b) also illustrates strong correlations between the Armenian verb *xashel* (to boil*) with the {pind} (hard) and {N_A_SUB} values of the *Resultative* (adjectives) and *Substance* sparameter values respectively even though {N_A_SUB} correlates with the verb *thxel* (to bake) as well.[129]

[129] Strong correlations between the verbs *xashel* as well as *thxel* with the {N_A_SUB} value of the *Substance* parameter was also discussed in more details at the beginning of this chapter (see Fig. 1a)

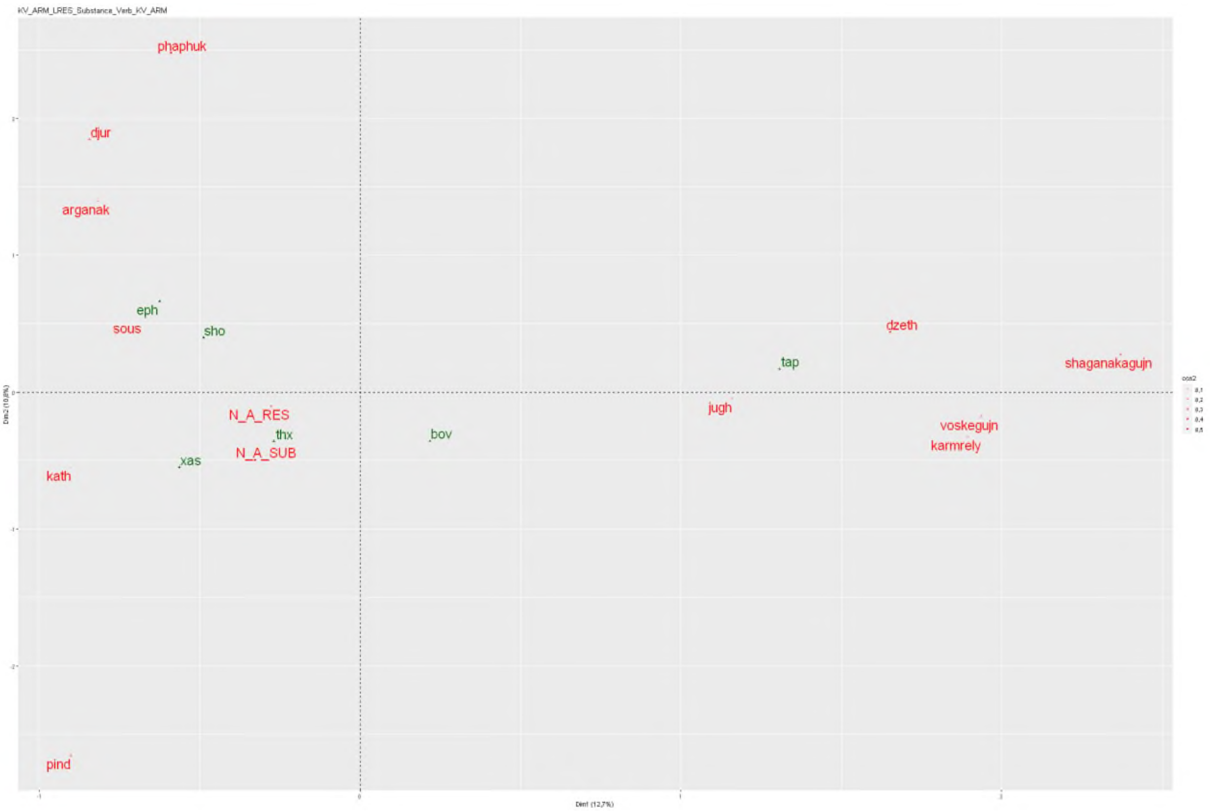


Fig. 5c. MCA: Armenian culinary verbs (parameters *Resultative* (adjectives) and *Substance*)

The *MCA* (Fig. 5c) identifies clusters of either strongly correlated independent variables or variables with profile similarities. In the clusters of the *Substance* parameter values {jugh, dzeth} (ghee, oil) as well as the *Resultative* (adjectives) ones {shaganakagujn, voskegujn, karmrely} (brown, golden, become red*), the independent variables have the same syntagmatic distribution stating their profile similarities with {shaganakagujn} creating an outlier. However, the aforementioned values of the *Resultative* (adjectives) and *Substance* parameters could form a larger cluster of strongly correlated independent variables stating their co-occurrence in similar contexts. The verb *tapakel* (to fry) is placed near the large cluster as qualitative supplement variable (examples 39-41).

39. Տապակել *բուսայուղով*, մինչև *կարմրի*: (ARM)
[Fry with oil until it gets red*/until brown].

40. Տաքացնել *արևածաղկի ձեթը* թավայի մեջ և տապակել կարտոֆիլը մինչև ոսկե *դարչնագույնը*: (ARM)
[Heat the *sunflower oil* in a pan and fry the potatoes until *brown*].

41. Հավի միսը կտրատել փոքր կտորներով, տապակել քիչ քանակությամբ *բուսական յուղով*, մինչև առաջանա թեթև *ոսկեգույն* կեղև: (ARM)
[Slice the chicken into small pieces and *fry* them in some *oil* until it gets *light golden*].

This *MCA* graph (Fig. 5c) illustrates another large cluster of strongly correlated *Resultative* (adjectives) and *Substance* parameter values, i.e. {phaphuk} (soft) x {djur, arganak, sous} (water, broth/stock, sauce) – {phaphuk} being overrepresented as a relevant parameter irrespective of its fewer frequencies in the annotated *ARM* corpus. The verbs *shogexashel* and *ephel* are distributed around the cluster (examples 42-43) as qualitative supplement variable. The *Substance* parameter value {kath} (milk) could also be included in this cluster due to its profile similarities with {djur, arganak, sous} (water, broth/stock, sauce), however, creates a strong correlation with the *Resultative* (adjectives) parameter value {pind}(hard) and is place on the lower-right side of the *MCA* graph (Fig. 5c) – {pind} generating an outlier (Fig. 5c).

42. Կարտոֆիլը կաթսայի մեջ աղ արած *ջրով* մի քանի րոպե *եփել*, որպեսզի կտորները դառնան *փափուկ*, բայց չփլվեն: *(ARM)*
[*Boil*/cook** the potatoes in a *pot* in boiling salted *water* so that the pieces become *soft* but not fall apart].

43. *Արգանակը* հասցնում ենք եռման աստիճանի, ավելացնում ենք կարտոֆիլն ու 10-15 րոպե *եփում* ենք մինչև կարտոֆիլը *փափկի*: *(ARM)*
[Bring the *stock/broth* to boiling, pour it onto the potatoes and *cook*/boil** it for 10-15 minutes until they become *soft*].

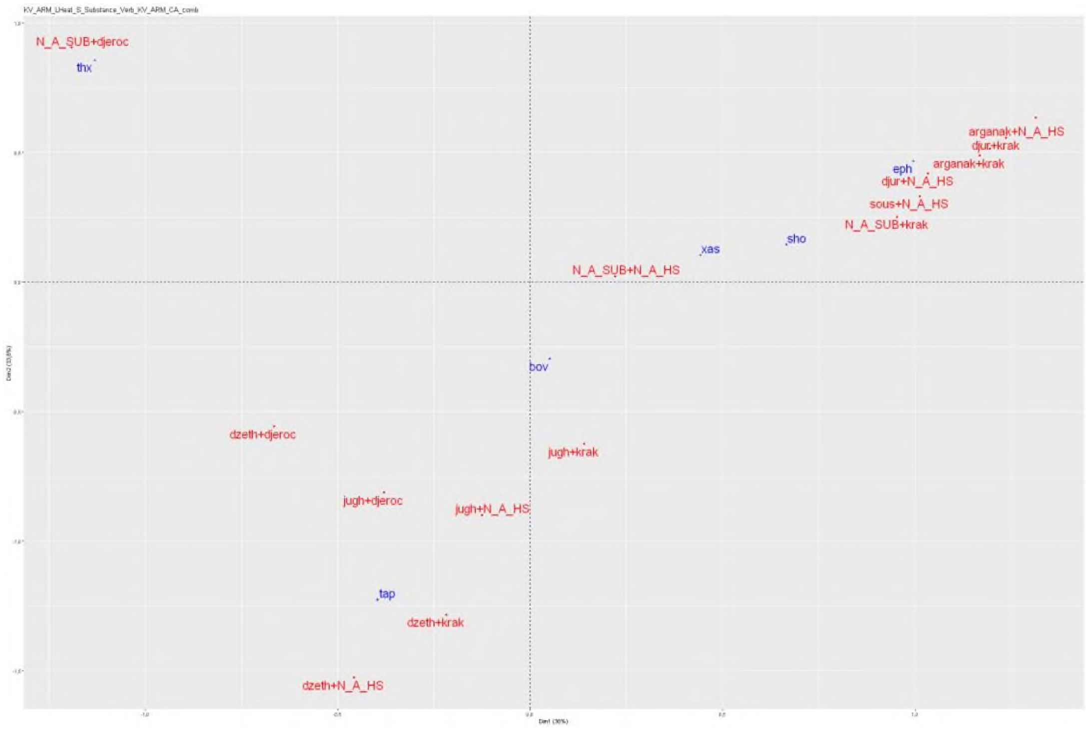


Fig. 6a. CA with concatenated parameters (*Substance* and *Heat source*)

The effect of the parameter *Substance* on the distribution of the Armenian culinary verbs has already been elaborated upon above (Fig. 3a), nevertheless, in pairs with the parameter *Heat intensity* in *CA* (Fig. 3a) and MCA (Fig. 3b) Taking into account that *CA* with concatenated parameters provides a more detailed image of the correlations between the independent and dependent variables, it is worth considering the parameter *Substance* paired with *Heat source*. If the *CA* with concatenated parameters display more precisely strong correlations between the Armenian culinary verbs and parameter combinations (Fig.6a), the *MCA* graph below (Fig. 6b) present clusters of strongly correlated *Substance* and *Heat source* parameter values.

The strong correlations in the *CA* with concatenated parameters (Fig. 6a) are determined by mutually exclusive values of the *Substance* parameter [+Water], [+Fat], and [N_A_SUB]. *(no substance at all).* In addition, the distribution of the verbs on the *CA* with concatenated parameters (Fig. 6a) is impacted by three main parameter values, i.e. {N_A_HS}, {krak} (fire) and {djeroc} (oven). One of the strongest correlations between the *Substance* and *Heat source* parameter with concatenated parameters (Fig. 6a) is between the Armenian culinary verb *thxel* (to bake) and the values {N_A_SUB} x {djeroc} (oven) placed at the upper-left corner of the plot (examples 44-45).

44. Խմորեղենը հարկավոր է միշտ *թխել ջեռոցի* մեջտեղի դարակում, եթե բաղադրատոմսում հատուկ նշված չի կոնկրետ դարակը: *(ARM)*
[You should always *bake* the cake in the middle shelf of the *oven* if a specific shelf is not mentioned in the recipe].

45. Խմորը լցնել ձևամանի մեջ և այն 45–55 րոպե *թխել ջեռոցում*: *(ARM)*
[Pour the dough in a baking form and *bake* it in *oven* for 45-55 minutes].

The verb *ephel* creates a strong correlation with all possible combinations of the *Substance* parameter values {arganak} (stock/broth), {djur} (water), {sous} (sauce), and {kath} (milk) all denoting *liquid* and the aforementioned three main *Heat source* ones: {N_A_HS}, {krak}, and {djeroc}[130](examples 46-47).

46. Այս ապուրը կարելի է *եփել* ինչպես մսի, այնպես էլ բանջարեղենի *արգանակով*: *(ARM)*
[You can *cook** this soup both with meat *broth* and vegetable *stock*].

47. Սառը *ջուր* լցնել, *եփել* այնքան, մինչև ոսպի հատիկները փափկեն: *(ARM)*
[Add cold water, *cook*/boil** until the lentil until it becomes *soft*].

The verb *tapakel* (to fry) correlates strongly with the *Substance* parameter values of [+Fat], i.e. {dzeth} (oil) and {jugh} (ghee) in different combinations with the three main *Heat source* values {krak, djeroc, N_A_HS} (example 48).

48. Հավի միսը լցնում ենք *բուսայուղով* լի տապակի մեջ ու մոտ կես րոպե տապակում ուժեղ *կրակի* վրա: *(ARM)*
[Add the chicken pieces to a pan filled with vegetable oil and fry for about half a minute *over high heat*].

[130] For the sake of saving space, all parameter combinations correlating with *ephel* were given in the text, however, the CA pairs graph displays all of them (see Fig.6a).

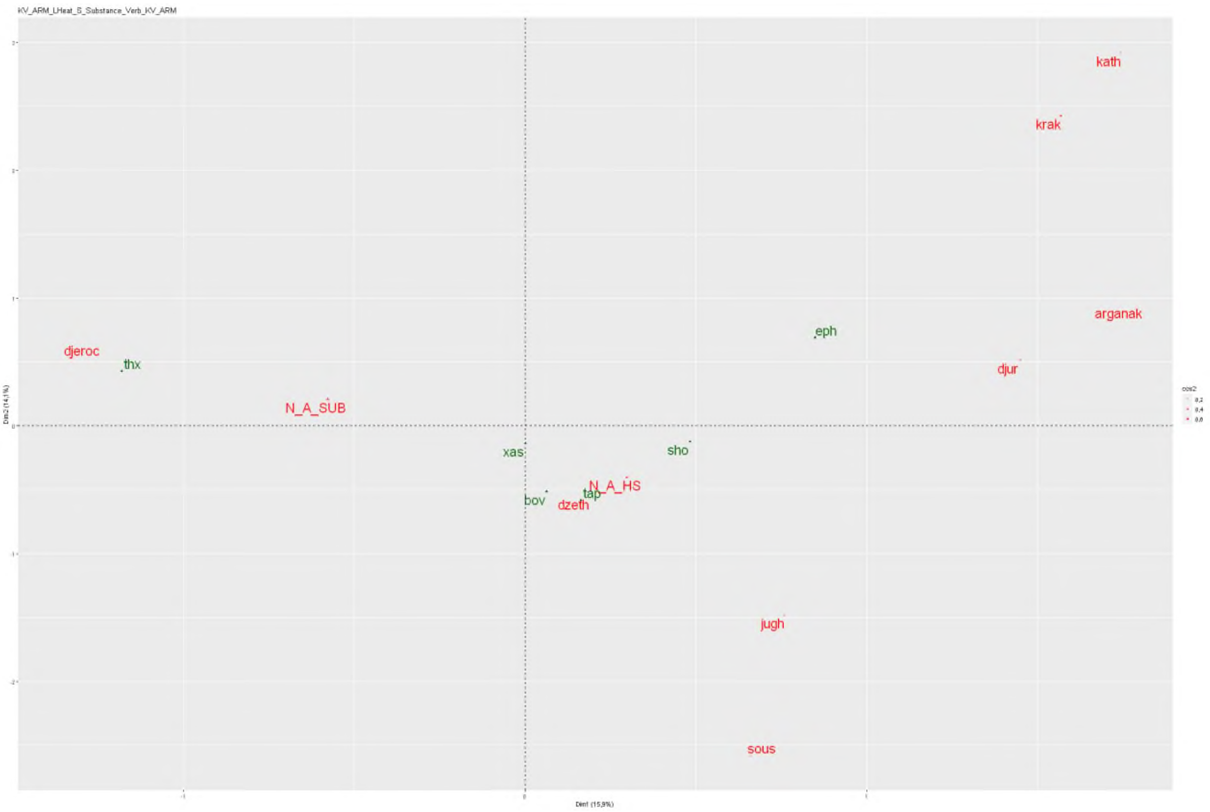


Fig. 6b. MCA: Armenian culinary verbs (parameters *Substance* and *Heat source*)

The *MCA* graph (Fig. 6b) distinguishes two clusters placed on upper- and lower-right sides of the graph plot, strongly correlated values of the *Substance* and *Hear source* parameters {kath} (milk) x{krak} (fire*/heat) (examples 49) as well as {arganak} (stock/broth) and {djur}(water) having profile similarities.

49. Ավելացնել *կաթի* մնացած քանակությունը: Եփել միջին *կրակի* վրա ՝ անընդհատ խառնելով: *(ARM)*
    [Add the rest of the *milk* and *boil*/cook** over medium *heat* by constantly stirring it].

Moreover, {kath} (milk) is overestimated and is positioned at the upper-right corner of the graph, whereas the *Substance* parameter values {arganak} (stock/broth) and {djur}(water) are pulled down by {N_A_HS}.

The positioning of the mutually exclusive *Substance* parameter values {jugh} (fat) and {sous} (sauce) as a small cluster is due to their strong correlation (separately from each other) with the *Heat source* parameter value {N_A_HS}, denoting *non-verbalization* of any heat source. The positioning of the {N_A_HS} much nearer to the intersection is due to its large quantity in the overall *ARM* corpus while the *Substance* parameter value {sous} (sauce) is overrepresented generating an outlier on the *MCA* graph (Fig. 6b).

4.3 *Context-Conditional Correlation Graphs* (CCCG) Visualization of the Armenian Data Analysis

CCCG visualization method generated with the *qgraph* package of the programming language *R* allows to illustrate all strong correlations between the Armenian culinary verbs and the annotation parameters at the same time and within one graph. At this respect, *CCCG* resembles the *Mosaic-Plot* visualization of contingency table analysis where vertical and horizontal columns marked blue, red or grey demonstrate strong correlations, anti-correlations and absence of correlation (neutral) between the verbs and the annotation parameter values[132]. *CCCG* also allowes to compare the verbs in pairs. Chapter 2.4 on the German and Chapter 3.3 on the English *CCCG* visualization of data analysis elaborate in detail on both the interpretation of this type of data representation and the calculations behind it. Therefore, they will not be discussed here in detail. It is worth restating, however, that the more edges the verbs are connected to each other, in more parameters they differ. Consequently, less edges potentially indicate the semantic closeness of the given verbs. Nevertheless, less edges could also be the result of the comparison between verbs both having non-verbalized values in the same parameters. The numbers next to the parameter values indicate the deviation of the observed values from the expected values, assuming an independent distribution of the parameters across verbs. These deviations are expressed in standard deviation units (corresponding to the square root of the expected value). Deviations exceeding twice the standard deviation are generally considered statistically significant. To insure conformity, the Armenian graph plots are also based on the deviation over 2.5 times of the standard deviation as were the German and English ones. The following *CCCG* visualization is based on the initially annotated 862 context examples which have been later reduced to 779 ones (see Chapter 4.1).[133]

[132] For a more detailed interpretation of *Mosaic-Plot* visualization see Chapter 2.4.1 of this work as well as Friendly 1994.

4.3.1 CCCG of the Armenian culinary verb *bovel*

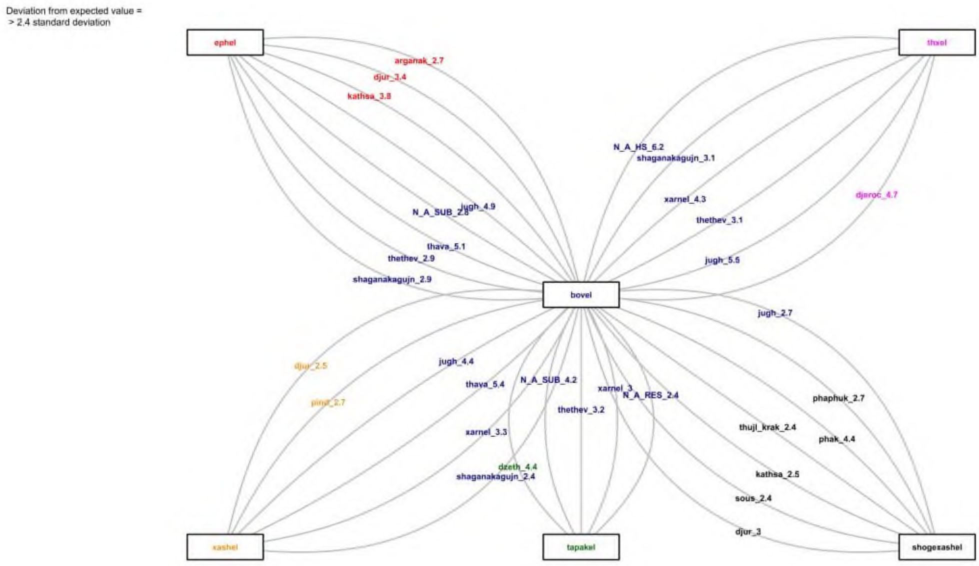


Fig. 1. *CCCG* analysis of the Armenian culinary verb *bovel* (to roast)

As mentioned above, deviations exceeding twice the standard deviation are in general statistically significant and could serve as a comprehensive reference to see all parameter values having impacted the distribution of the verbs in pairs. Up to this point in this work, all *CCCGs* have been generated based on the 2.5 times the standard deviation for the sake of legibility and with the intention to profoundly elaborate on the most significant parameters and its separate value. Nevertheless, for certain verb pairs there is a need to lower this threshold and observes distribution of parameters with lower than 2.5 times the standard deviation.

In comparing the verb *bovel* (to roast) to the verb *ephel* (to cook*, to boil*), the connecting edges denote their semantic similarity only in parameters *Heat intensity* and *Heat source* and differences in the other four annotation parameters. Among them, the *Substance* parameter value {djur} (water) and {arganak} (broth) create a strong correlation with the verb *ephel* while *bovel* does with {jugh} (oil) and {N_A_SUB} restating that *ephel* presupposes some type of *liquid*-like substance. *Bovel* implicated either the *absence of any cooking substance*, for instance, in a dry pan or only with some kind of *fat*, e.g. butter. Next, the *Utensil* parameter values {kathsa} (pot) and {thava} (pan) oppose the verb *ephel* to *bovel.* The process of *bovel* is typically done in a shallow utensil such as {thava} (pan) while *ephel* rather in a deeper one, i.e. {kathsa} (pot), even though the values could be to some contexts interchanged (example 1).

1. *Թավայի* մեջ *կարագով բովել* ալյուրը, երբ ոսկեգույն դառնա, տաք կաթ ավելացնել: *(ARM)*
   [In a *pan roast* the flour with *butter* until it becomes golden and add the warm milk].

The parameter *Resultative* (adjectives) has also a statistically significant impact on the distribution of the occurrences for this pair of verbs, differentiating *bovel* from *ephel*. *Bovel* creates a strong correlation with {voskegujn} (golden) and {shaganakagujn} (brown) (examples 1 and 2).

2. Ալյուրը *բովել* թեժ կրակի վրա, մինչև այն դառնա բաց *շագանակագույն*: *(ARM)*
   [*Roast* the flour over high heat until it turns *light brown*].

The semantic difference between *bovel* and *ephel* in the parameter *Manner* is observed in *bovel* correlating strongly with {thethev} (light) (example 3) (Fig. 1, partial Table 1.1).

3. Քնջութը լավ կլինի *թեթև բովել*, ավելի համեղ կլինի: *(ARM)*
   [It would be better to *lightly roast** the sesame seeds; it would taste better].

5	**bovel**	ephel	**shaganakagujn_2.9**	0.436801243704289
6	bovel	ephel	thethev_2.9	0.436801243704289
7	**bovel**	ephel	**thava_5.1**	0.398938687254965
9	bovel	ephel	N_A_SUB_2.8	0.440862127030585
10	bovel	ephel	jugh_4.9	0.401345680884024
8	**ephel**	bovel	**kathsa_3.8**	0.416370937939784
11	ephel	bovel	djur_3.4	0.423987071209447
12	ephel	bovel	arganak_2.7	0.442747702141477

Table (partial) 1.1. Actual distribution of the annotated occurrences in the verb pair *bovel-ephel*

In pairs *bovel* (to roast)-*shogexashel* (to steam*, to braise*, to cook in little water*) and *bovel-xashel* (to boil*), the parameter *Substance* with its values denoting *liquid* – {arganak} (stock/broth), {sous} (sauce), {kath} (milk), and {djur} (water) – correlate with *xashel* and *shogexashel*, while {jugh} (fat, ghee) with *bovel*. For instance, in comparing the verbs *bovel* and *ephel*, the co-occurrences of the *Substance* parameter value {jugh} (fat, ghee) is overrepresented for *bovel* 6.7 times the standard deviation (example 1; 4). In the abovementioned three pairs the verb *bovel* opposes *shogexashel*, *xashel*, and *ephel* in *fat* vs. *liquid* contrast (partial Tables 1.1, 1.2 and 1.3). Semantically the verb pairs *bovel-xashel* and *bovel-shogexashel* are closer only in parameters *Heat intensity* and *Heat source,* mostly because the overwhelming majority of the annotated examples in these two parameters were not verbalized.

4. Թավայի մեջ *բուսայուղ* լցնել, ավելացնել ալյուրն ու *բովել*, մինչև մի քիչ գույնը փոխի: *(ARM)*
   [In a *pan roast** the flour with some vegetable *oil* until it starts to change the color].

2	**bovel**	xashel	**xarnel_3.3**	0.425608653737254
3	**bovel**	xashel	**thava_5.4**	0.396471822965778
4	**bovel**	xashel	**jugh_4.4**	0.406916126841555
1	xashel	bovel	pind_2.7	0.442322663335028
23	shogexashel	bovel	phaphuk_2.7	0.443309194943083
24	shogexashel	bovel	phak_4.4	0.406501523421995
25	shogexashel	bovel	kathsa_2.5	0.449960788492695
27	**shogexashel**	bovel	**djur_3**	0.432709761243843
26	**bovel**	shogexashel	**jugh_2.7**	0.443803619796016

Table (partial) 1.2. Actual distribution of the annotated occurrences in the verb pairs *bovel-xashel* and *bovel-shogexashel*

In the verb pair *bovel-tapakel* (to fry) the semantic difference between the verbs being compared is observed in the parameters *Manner* and *Substance*. In the remaining four annotation parameters – *Resultative* (adjectives), *Utensil*, *Heat intensity* and *Heat source* – *bovel* and *tapakel* could be considered as semantically close. For instance, in the verb pair *bovel-tapakel*, the co-occurrences of {dzeth} (oil) with *tapakel* are superior to the expected value by 4.4 times the standard deviation under the hypothesis of uniform independent distribution. *Bovel*, on the contrary, correlates strongly with *no substance* or *non-verbalized Substance* parameter values (cf. examples 5 and 6) (see Fig. 1, partial Table 1.3).

16	bovel	tapakel	N_A_RES_2.4	0.444739323156838
17	bovel	tapakel	xarnel_3	0.4269125919601
18	bovel	tapakel	thethev_3.2	0.422577024271692
19	**bovel**	tapakel	**N_A_SUB_4.2**	0.404822024860202
20	**tapakel**	bovel	**dzeth_4.4**	0.402504436054964

Table (partial) 1.3. Actual distribution of the annotated occurrences in the verb pair *bovel-tapakel*

5. 5-7 րոպե *չոր տապակի* վրա բովել ընկույզը, որ մի փոքր շագանակագույն դառնան (միջին ջերմության վրա, *առանց յուղի*). *(ARM)*
   [*Roast* the walnuts in *a dry* pan* until they become a bit (light) brown (on medium heat *without fat*)].

6. Սուրճի հատիկները թավայի մեջ *բովել*, անընդհատ խառնել փայտե գդալով: *(ARM)*
   [*Roast** the coffee beans in a pan by stirring them constantly with a wooden spatula/spoon].

In general, it could be stated that while the process of *tapakel* implies a *fat*-like cooking substance, *bovel* excludes any substance and often supposes a dry surface or utensil. In addition, the manner of *bovel* distinguishes it from *tapakel* by correlating strongly with {xarnel} (to stir) and {thethev} (lightly).

In comparing the verbs *bovel* and *thxel* (to bake), the two verbs are semantically close only in parameters *Heat intensity* and *Utensil* and different in the rest of the annotation parameters. For instance, while *thxel* correlates strongly with the *Heat source* parameter value {djeroc} (oven), no specific heat source is verbalized in the annotated examples with *bovel*. Under the hypothesis of a uniform independent distribution**,** the expected values for the co-occurrence of the *Heat source* parameter {djeroc} (oven) with the verbs *bovel* and *thxel* are 50 and 86 examples, respectively. These values are calculated based on the total number of annotated instances for each verb within the *ARM* sub-corpus and their overall frequencies of co-occurrence with {djeroc} (oven). In the observed distribution of {djeroc} (oven) with *bovel* and *thxel*, the number of co-occurrences of {djeroc} (oven) with *thxel* surpasses the expected value by 4.7 times the standard deviation: out of 137 overall annotated examples of {djeroc} (oven), 131 ones are with the verbs *thxel*. The semantic difference in this verb pairs is also strongly impacted by the *Resultative* (adjectives), *Manner* (example 7), and *Substance* parameters (see Fig. 1, partial Table 1.3).

20	**bovel**	thxel	**N_A_HS_6.2**	0.388428280704407
22	bovel	thxel	shaganakagujn_3.1	0.426346566300922
23	bovel	thxel	xarnel_4.3	0.405550286531516
24	bovel	thxel	thethev_3.1	0.426346566300922
25	bovel	thxel	jugh_5.5	0.393342133178567
21	**thxel**	bovel	**djeroc_4.7**	0.400664103490904

Table (partial) 1.4. Actual distribution of the annotated occurrences in the verb pair *bovel-thxel*

7. Նախապես *բովել* ընկույզները միջին կրակի վրա, *անընդհատ խառնելով*, որպեսզի դրանք ձեռք բերեն հարուստ համ: *(ARM)*
   [*Roast** the walnuts beforehand on medium heat *by constantly stirring* so that they become rich in flavour].

Summarizing our observations, it could be stated that in comparing *bovel* (to roast) to the other verbs, the most significant parameters affecting the distribution are *Substance* and *Manner*. In pairs *bovel-shogexashel* (to cook with little water*, steam*, braise*), *bovel-xashel* (to boil*), *bovel-ephel* (to cook*, to boil*), the verb bovel sematically differes from the verbs being compared in {jugh} (fat, ghee) vs. {water} (water) mutually exclusive components while in verbs pairs *bovel-tapakel* and *bovel-thxel,* the semantic difference lies in *no substance* vs. {dzeth} (oil), {jugh} (fat, ghee) vs. *no substance* features. The parameter *Manner* differentiates the verb *bovel* from all the above verbs compared in pairs.

4.3.2 CCCG of the Armenian Culinary Verb *shogexashel*

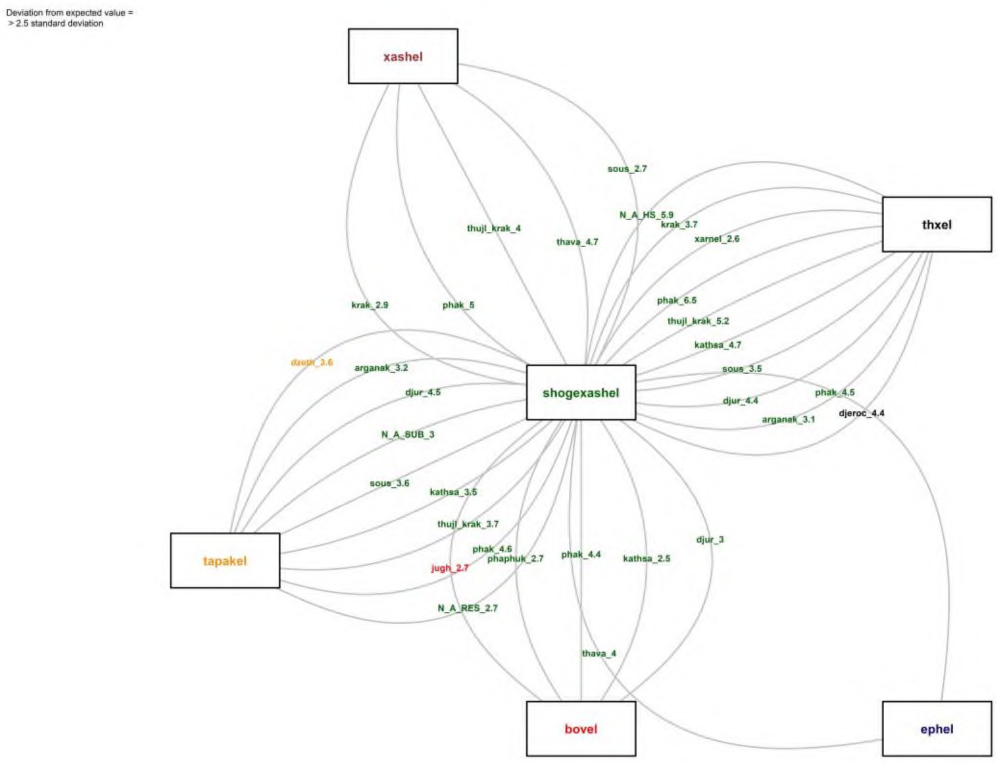


Fig. 2. *CCCG* analysis of the Armenian culinary verb *shogexashel*

The number of edges connecting the verb *shogexashel* (to steam*, to steam*, to cook with little water*) to the other five Armenian culinary verbs shows its statistically significant semantical difference in a number of parameters. The paramters *Manner*, *Substance* and *Heat intensity* have the most significant impact on the distribution of the annotated examples. For instance, in the distribution of the *Manner* parameter values for the verbs *shogexashel* and *xashel* (to boil*, to cook*), the number of occurrences of {*phak*} (closed) with *shogexashel* is exceeds the expected value by the 5 times the standard deviation under the hypothesis of uniform independent distribution (example 8) (Fig.2, partial Table 2.1). There are statistically no significant differences between the verbs *shogexashel* and *xashel* in the six annotation parameters except for the *Resultative* (adjectives), indicating their semantic similarity in this parameter.

8. *Կափարիչով ծածկված* տապակի մեջ *շոգեխաշել* մարմանդ կրակի վրա, այնքան ժամանակ, մինչև հավի միսը քնքուշ ու փափուկ դառնա: *(ARM)*

[Steam*, braise* in a pan with the *lid on/closed* over low heat so long till the chicken is soft and tender].

1	shogexashel	xashel	krak_2.9	0.436618909822003
2	**shogexashel**	xashel	**phak**_5	0.400009450902646
3	shogexashel	xashel	thujl_krak_4	0.412839961886284
4	shogexashel	xashel	thava_4.7	0.402714589795575
5	shogexashel	xashel	sous_2.7	0.441304347826087

Table (partial) 2.1. Actual distribution of the annotated occurrences in the verb pair *shogexashel-xashel*

In the verb pair *shogexashel* (to steam*, to steam*, to cook with little water*)-*thxel* (to bake), the distribution is affected by all of the six annotation parameters. This is one of the rare verb pairs, where also the parameter *Heat intensity* impacts the distribution. Particularly in comparing the verbs *shogexashel* and *thxel*, the observed distribution of the *Heat intensity* parameter value {thujl_krak} (low heat) co-occurring *shogexashel* is superior to the expected value by 5.2 times the standard deviation (example 9) (Fig. 1, partial Table 2.2).

9. Լավ խառնել և 5-10 րոպե *շոգեխաշել* այս խառնուրդը *ցածր ջերմության վրա*: *(ARM)*
   [Mix*, stir well and *braise*/steam** this mixture for 5-10 minutes *over low heat*].

22	**shogexashel**	thxel	**N_A_HS_5.9**	0.392529398774837
23	shogexashel	thxel	krak_3.7	0.416830222945757
25	shogexashel	thxel	xarnel_2.6	0.444512207666307
26	**shogexashel**	thxel	**phak_6.5**	0.388584447207736
27	**shogexashel**	thxel	**thujl_krak_5.2**	0.398483739536704
28	shogexashel	thxel	kathsa_4.7	0.40283393026136
29	shogexashel	thxel	sous_3.5	0.42044524034848
30	shogexashel	thxel	djur_4.4	0.406481847267062
31	shogexashel	thxel	arganak_3.1	0.429877394432965
24	**thxel**	shogexashel	**djeroc_4.4**	0.407390275956025

Table (partial) 2.2. Actual distribution of the annotated occurrences in the verb pair *shogexashel-thxel*

The parameter *Heat source* differentiates *shogexashel* from *thxel* with the latter correlating strongly with {djeroc} (oven) while the vast majority of the annotated examples *shogexashel* have no verbalized values in this parameter. The verb *thxel* (to bake) correlates strongly with *no substance* (more clear from the examples that it is not merely *non-verbalization* but rather the *absence of any cooking substance*) as opposed to *shogexashel* correlating with *liquid*-like substances, e.g. {sous} (sauce), {djur} (water), and {arganak} (broth/stock) (examples 10-11, Fig. 2., Table (partial) 2.2). A more detailed semantic analysis of the verb *thxel* is followed, i.e. Fig. 3.

10. Լոբին մաքրել, լցնել մսի վրա, մի փոքր մսի *արգանակ* ավելացնել և *շոգեխաշել* թույլ կրակի վրա: *(ARM)*
    [Clean the green beans, add them to the meat, pour some *stock/broth* and *braise** over low heat].

11. Լցնել հավի ազդրերը *սոուսի* մեջ: Կափարիչով ծածկված տապակի մեջ *շոգեխաշել* մարմանդ կրակի վրա: *(ARM)*
    [Add the chicken drumsticks into the *sauce* and braise* them in a pan with the lid on over low heat].

The verb *shogexashel* semantically differs from *tapakel* (to fry) in all of the six annotation parameters. Nevertheless, *Substance* seems to have the most significant effect on the distribution. Here too, strong correlations with mutually exclusive cooking substances (*water* vs. *fat*) is observed (example 12) (Fig. 2, partial Table 2.3). *Tapakel* correlates with some type of *fat/oil*-like substance whereas *shogexashel* creates a strong correlation with some *water*-like ones. A more detailed comparison of the verb *tapakel* with the rest of the Armenian verbs in Chapter 4.3.4.

12. Տապակի մեջ շարել լյարդի կտորները, խնձորի շերտերը, մի քիչ *եռջուր* լցնել, կրակն իջեցնել, *շոգեխաշել* 5-7 րոպե՝ մինչև խնձորը փափկի, բայց ձևը չկորցնի: *(ARM)*
    [In a pan put in layers the pieces of the liver, the apple slices, a bit of butter, a little bit of *boiling water*, lower the heat and *braise*/steam** for 5-7 minutes until the apples are soft but have not lost their shape].

8	shogexashel	tapakel	N_A_RES_2.7	0.443896085563916
9	shogexashel	tapakel	phak_4.6	0.404229098080874
10	shogexashel	tapakel	thujl_krak_3.7	0.418255065286364
11	shogexashel	tapakel	kathsa_3.5	0.420563161540358
12	**shogexashel**	tapakel	**sous_3.6**	0.419895497872448
13	shogexashel	tapakel	N_A_SUB_3	0.433184968578421
15	**shogexashel**	tapakel	**djur_4.5**	0.406041072696433

16	**shogexashel**	tapakel	**arganak_3.2**	0.429254045057232
14	**tapakel**	shogexashel	**dzeth_3.6**	0.420289635346212

Table (partial) 2.3. Actual distribution of the annotated occurrences in the verb pair *shogexashel-tapakel*

The verb pair *bovel-shogexashel* was thoroughly described at the beginning of this chapter (see Fig. 1, partial Table 1.2).

4.3.3 CCCG of the Armenian Culinary Verb *thxel*

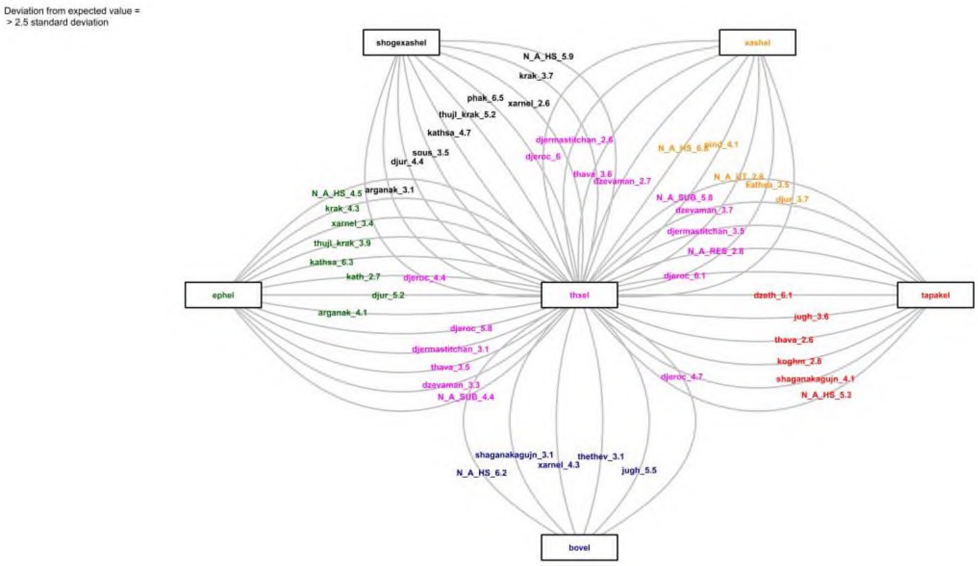


Fig. 3. CCCG analysis of the Armenian culinary verb *thxel*

The number of edges connecting *thxel* (to bake) to the rest of the verbs illustrates the semantic difference between the verb *thxel* and the rest of the six Armenian culinary verbs. The impact of the parameter *Heat source*, particularly its value {djeroc} (oven), in distinguishing *thxel* from *bovel* and *shogexashel* was discussed above and therefore will be omitted here (see Fig.2, partial Tables 1.4, 2.2).

However, the *Heat source* parameter value {djeroc} (oven) differentiates *thxel* from all the other five Armenian culinary verbs in the rest of the verb pairs *thxel-xashel*, *thxel-ephel,* and *thxel-tapakel* with high deviations from the expected value (example 13) (Fig. 3, Tables (partial) 3.1, 3.2 and 3.3).

13. Խմորը 10 հավասար մասերի ենք բաժանում և հաջորդաբար յուրաքանչյուրը 10-12 րոպե *թխում*՝ 180 աստիճան տաքացրած *ջեռոցում*: *(ARM)*

    [Divide the dough into 10 equal pieces and *bake* each of them separately in the preheated oven at 180°C].

The difference between *thxel* (to bake) and the other verbs in comparison is also significantly influenced by the parameter *Utensil*. *Thxel* and only *bovel* are semantically closer in the *Utensil* parameter. While *thxel* correlates with somewhat shallow utensils, often intended to be also put in the oven, e.g. {thava} (pan) and {dzevaman} (baking dish) in the pairs *thxel-shogexashel, thxel-xashel,* and *thxel-ephel,* the verbs in comparison tend to correlate with {kathsa} (pot), a much deeper vessel suitable for boiling*, cooking* dishes with some type of *liquid* (examples 14-15) (Fig. 3, partial Table 3.1).

14. Բլիթները կարելի է *թխել* թե *թավայի*, թե ջեռոցի մեջ: *(ARM)*
    [You can bake the pancakes either in a *pan* or in the oven].
15. Խմորը լցնել *ձևամանի* մեջ և այն 45–55 րոպե *թխել ջեռոցում*: *(ARM)*
    [Pour the dough in the *baking form* and *bake* it in the *oven* for 45-55 minutes].

10	ephel	thxel	N_A_HS_4.5	0.405882371811625
11	ephel	thxel	krak_4.3	0.408754955082849
13	ephel	thxel	xarnel_3.4	0.423829227703863
14	ephel	thxel	thujl_krak_3.9	0.414362828528158
17	**ephel**	thxel	**kathsa_6.3**	0.389612582635891
20	ephel	thxel	kath_2.7	0.442899740941345
21	ephel	thxel	djur_5.2	0.398515307425883
22	ephel	thxel	arganak_4.1	0.410817156415072
12	thxel	ephel	djeroc_5.8	0.392967674977425
15	thxel	ephel	djermastitchan_3.1	0.429445783893952
16	**thxel**	ephel	**thava_3.5**	0.421629699530875
18	**thxel**	ephel	**dzevaman_3.3**	0.424902203370792
19	thxel	ephel	N_A_SUB_4.4	0.407335178595008
2	thxel	xashel	djeroc_6	0.391414686587849
4	thxel	xashel	djermastitchan_2.6	0.446757448192469
5	thxel	xashel	thava_3.8	0.415110659176808
8	**thxel**	xashel	**dzevaman_2.7**	0.441223796996268
1	xashel	thxel	N_A_HS_6.8	0.386909214662559
3	**xashel**	thxel	**pind_4.1**	0.411587994796513
6	xashel	thxel	N_A_UT_2.8	0.440541836039975
7	**xashel**	thxel	**kathsa_3.5**	0.421747842654501
9	xashel	thxel	djur_3.7	0.417113961363327

Table (partial) 3.1. Actual distribution of the annotated occurrences in the verb pairs *thxel-ephel* and *thxel-xashel*

The *Utensil* parameter value {thava} (pan), however, in the verb pair *thxel-tapakel,* correlates strongly with *tapakel* (to fry) while {dzevaman} (baking dish, baking form) is overrepresented for *thxel*.

The semantic difference in the verb pairs *thxel-xashel, thxel-ephel, thxel-shogexashel,* and *thxel-tapakel* is also observed in the parameter *Substance*. The verbs *xashel, ephel,* and *shogexashel* correlate strongly with some type of *water*-like cooking substance, for instance {arganak} (broth) or {djur} (water) while *thxel* correlates with {N_A_SUB}. In comparing the verbs *thxel* and *tapakel*, the number co-occurrences of {dzeth} (oil) and {jugh} (fat) is overrepresented for the verb *tapakel* by respectively the 6.1 times and 3.6 times the standard deviation (Fig. 3, Table (partial) 3.3). The verb *thxel* semantically differs from *xashel* and *tapakel* in the *Resultative* (adjectives) parameter as well[134]. For instance, *xashel* correlates strongly with {pind} (hard) (example 16) (Fig.3, partial Table 3.1).

16. Ձուն *պինդ խաշել*, սառեցնել, օղակ-օղակ կտրատել: *(ARM)*
    [Hard boil* the egg, let it cool down, slice in rings].

[134] For a more detailed comparison of the verb pair *tapakel-thxel*, see the CCCG analysis of the verb *tapakel* Chapter 4.3.4.

4.3.4 CCCG of the Armenian culinary verb *tapakel*

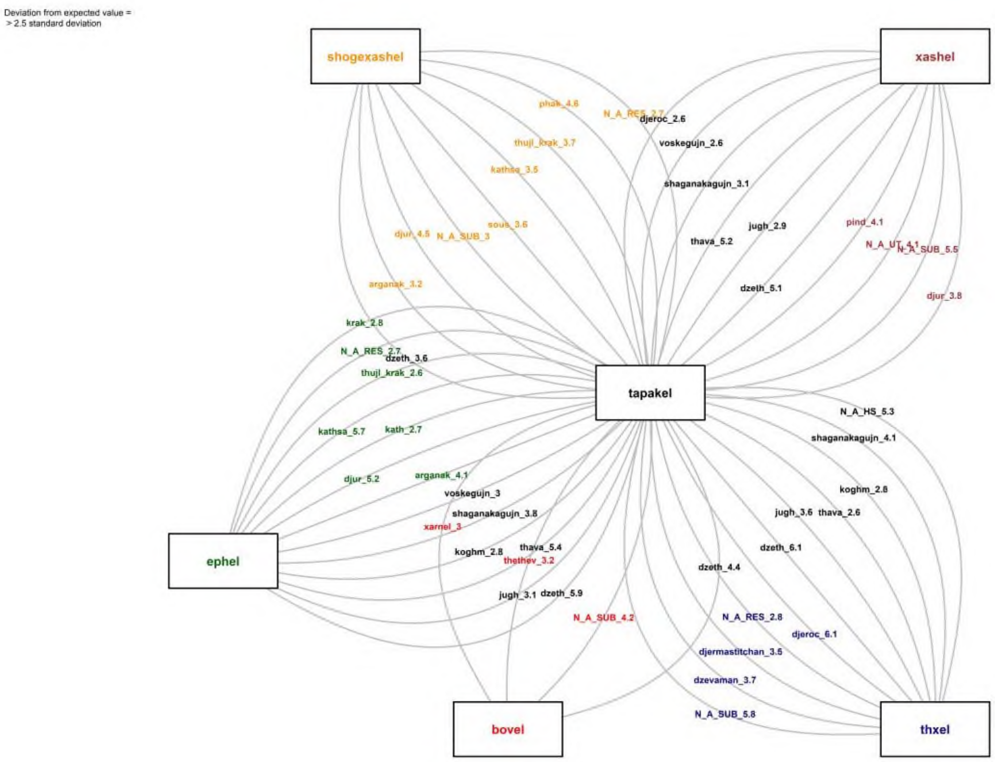


Fig. 4. CCCG analysis of the Armenian culinary verb *tapakel*

CCCG analyses of the verbs *shogexashel*, *thxel*, and *bovel* were discussed at the beginning of this chapter (see Fig.1; 2; 3), therefore, they will not be elaborated upon here. In the verb pair *tapakel-xashel*, the semantic differences are observed in all of the six annotation parameters judging from the edges connected them (Fig. 4, partial Tables 4.1 and 4.2).

14	tapakel	xashel	djeroc_2.6	0.493552352411229
15	**tapakel**	xashel	**voskegujn_2.6**	0.493552352411229
16	**tapakel**	xashel	**shaganakagujn_3.1**	0.471323881416638
18	tapakel	xashel	thava_5.2	0.422137771321351
21	**tapakel**	xashel	**jugh_2.9**	0.481044769330936
22	**tapakel**	xashel	**dzeth_5.1**	0.422877665258326
17	xashel	tapakel	pind_4.1	0.441617174958655
19	xashel	tapakel	N_A_UT_4.1	0.441578883491001
20	xashel	tapakel	N_A_SUB_5.5	0.418068723143537

23	**xashel**	tapakel	**djur_3.8**	0.449837501784366

Table (partial) 4.1. Actual distribution of the annotated occurrences in the verb pair *tapakel-xashel*

From the six annotation parameters having significantly influenced the distribution of the occurrences when comparing *tapakel* to the rest of the Armenian culinary verbs, the parameters *Substance*, *Manner* and *Resultative* (adjectives) are probably the most decisive ones, without neglecting the semantic difference between the compared verb pairs in the parameters *Utensil, Heat source,* and *Heat intensity* (Fig. 4, partial Table 4.1). For example, in the verb pairs *tapakel–xashel* and *tapakel–ephel*, the semantic distinction within the parameter *Substance* can be characterized by the mutually exclusive features [+oil], [–water] for *tapakel*, and [–oil], [+water] for *xashel* and *ephel*. These distinctions reflect contrasting features with *fat*-like cooking substances—{dzeth} (oil) and/or {jugh} —in the case of *tapakel*, vs. *liquid*-like substances — {djur} (water), {kath} (milk), {arganak} (broth) — in the case of *xashel* and *ephel* (see examples 17–19, Fig. 4, and partial Tables 4.1 and 4.2).

17. Տաքացրեք *յուղը տապակի* մեջ և *տապակել* սոխը *միջին ջերմության վրա*, խառնելով մինչև ոսկե դարչնագույնը, 5-7 րոպե: *(ARM)*
    [Heat the *ghee* in a *pan* and *fry*/sauté** the onion *over medium heat* until golden brown].

18. Մեծ *թավայի* մեջ տաքացնել *բուսայուղը* և *կարագի* կեսը, *տապակել* սոխը: Ավելացնել սունկը և *տապակել* 2-3 րոպե: *(ARM)*
    [Heat the vegetable *oil* and the half of the *butter* in a large *pan* and *fry*/sauté** the onion. Add the mushroom and *fry* for 2-3 minutes].

19. Բանջարեղենը լցնել *արգանակով* կաթսայի մեջ, եփեք 5 րոպե: *(ARM)*
    [Put the vegetables in a pot with *broth*/stock, cook* for 5 minutes].

35	ephel	tapakel	krak_2.8	0.484833278561441
38	ephel	tapakel	N_A_RES_2.7	0.489256939184715
40	ephel	tapakel	thujl_krak_2.6	0.493187399644391
42	**ephel**	tapakel	**kathsa_5.7**	0.415409913849531
43	ephel	tapakel	kath_2.7	0.488145308633981
46	**ephel**	tapakel	**djur_5.2**	0.42214403452484
47	ephel	tapakel	arganak_4.1	0.440437333388322
36	tapakel	ephel	voskegujn_3	0.475945094616391
37	tapakel	ephel	shaganakagujn_3.8	0.449580327634948
39	tapakel	ephel	koghm_2.8	0.485154581637768
41	**tapakel**	ephel	**thava_5.4**	0.419052774463168
44	**tapakel**	ephel	**jugh_3.1**	0.470298557805624
45	**tapakel**	ephel	**dzeth_5.9**	0.414072383954256

Table (partial) 4.2. Actual distribution of the annotated occurrences in *tapakel-ephel* verb pair

As for the semantic difference in the aforementioned verb pairs *tapakel-xashel* and *tapakel-ephel* in the parameter *Resultative* (adjectives), *tapakel* correlates with {shaganakagujn} (brown) and {voskegujn} (golden) while *xashel* correlates with {pind} (hard) (cf. examples 20-21) (Fig.4, partial Tables 4.1 and 4.2).

20. Ավելացնել սունկը և 2-3 րոպե *տապակել*, մինչև դրանք *շագանակագույն* դառնան: (ARM)
    [Add the mushroom and *fry*/sauté** for 2-3 minutes until they become *brown*].
21. Ձվերը *պինդ խաշել*: (ARM)
    [*Hard boil** the eggs].

4.4 *Conditional Inference Trees* (CIT) of the Armenian Culinary Verbs

CIT model was introduced in Chapters 2.5. and 3.4 based on the German and English *CIT* models). *CIT* model of the Armenian culinary verbs was generated by the *partykit* package of the programming language *R* for statistical computation. Being a tree-structured model of data analysis, *CIT* model is based on repeated partitioning of data into inner nodes determined by p-values until the splitting is statistically insignificant (see Gries 2009). Thus, lower p-values reflect the real difference in the split nodes and the non-accidental character of data selection regarding the distribution of verbs in the given node. In the case *CIT* analysis of the Armenian culinary verbs, lower p-values indicate the significance of the annotation parameters affecting on the distribution in the inner neighboring nodes. For instance, the p-value under the parameter *Resultative* (adjectives) splitting the nodes 6 and 7 shows that the likelihood of a merely accidental difference in the distribution between the aforementioned nodes is smaller than 0,1% (Fig. 1).

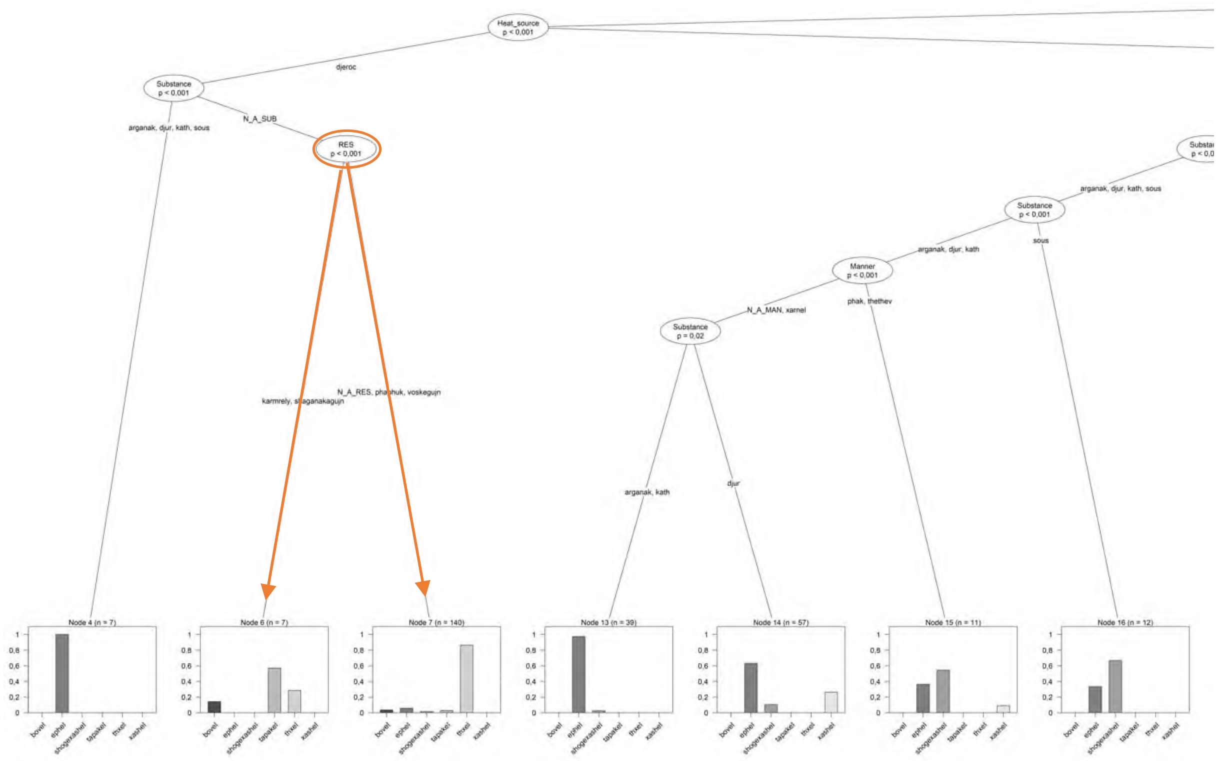


Fig. 1. Partial *CIT*. Distribution model of the Armenian culinary verbs in context[135]

[135] This is a partial CIT model of Armenian culinary verbs. The complete model, along with grouped nodes based on the number of correct predictions for the most frequent verb within identified context clusters, is provided in Appendix 4.

Analogous to both the English and German *CIT* models, *CIT* representation of Armenian culinary verbs identifies clusters of possible context types where the most frequently verb is singled out – also visually prominent in the form of the highest column at the bottom of the *CIT* graphic. The identified context types are determined by certain combinations of the annotation parameters, viz. *Substance, Utensil, Manner, Resultative* (adjectives), *Heat intensity* and *Heat source*. The initially annotated 862 contexts examples extracted from the *ARM* corpus which were reduced to 779 (see Chapter 4.1) served as the basis for generating the *CIT* model of the Armenian cooking verbs. As an outcome of the partitioning of this *CIT* model, 25 inner and 26 terminal nodes were identified representing the parameter determined clusters where the quantitatively dominant verb stood out. The six Armenian culinary verbs in clusters are organized alphabetically starting with *bovel* and followed by *ephel, shogexashel, tapakel, thxel*, and *xashel* (Fig.1).

The overall frequency of the dominant verbs in the identified context clusters is ≈69% which means that applying a certain parameter combination as heuristic to "guess" an adequate lexical choice in producing Armenian culinary texts by choosing the most frequent verb in the identified clusters resulted in an average "error rate" of ≈31%. However, "error rate" does not denote that the verb choice is incorrect but rather showed the degree of deviation from the most frequently used verb in context. For instance, the model suggested semantically synonymous verbs or hypernyms instead of more a specific verb. By excluding the node 26 with 157 annotated occurrences (52.9% error rate) as most of the annotation parameters here were not verbalized – making a context based prediction impossible – the overall error rate of this *CIT* model reduced to as low as 25%. Therefore, this *CIT* model of Armenian culinary verbs represents a significant improvement over a baseline scenario of an independent discrete uniform distribution, where the probability of correctly selecting the appropriate culinary verb from six options was 16.6%, resulting in a random guess error rate of 83.3%.

The results of this *CIT* model – the identified most frequent verb(s) in the given clusters – could also be compared to the distribution of the six Armenian culinary verbs in our manually compiled *ARM* corpus of recipes. Fig. 2 below illustrates the distribution of our six Armenian culinary verbs in the *ARM* corpus according to frequency.

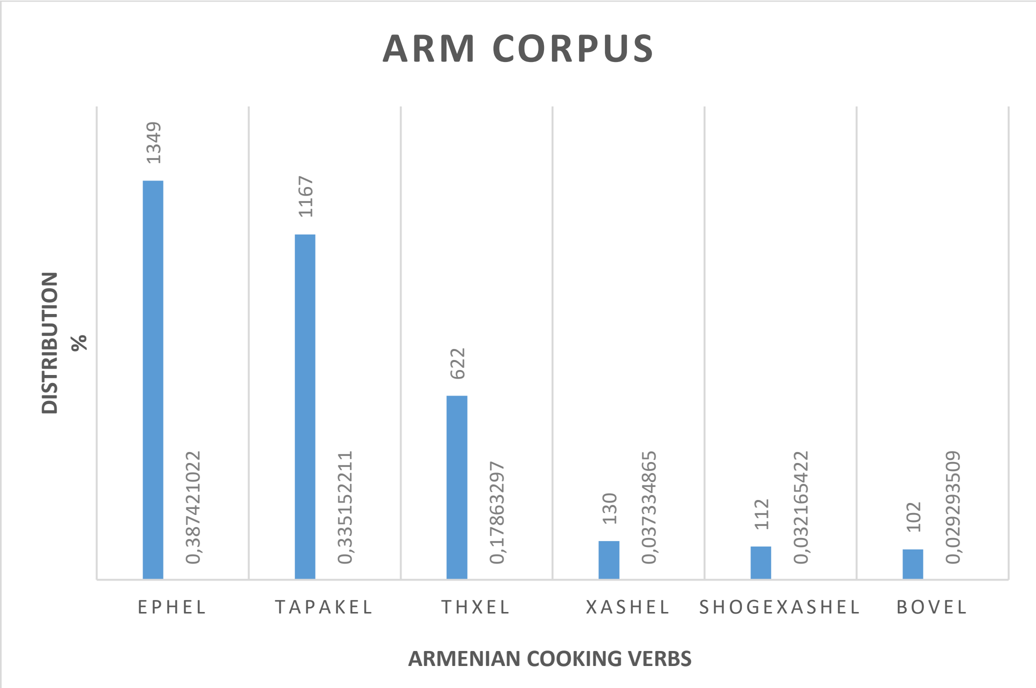


Fig. 2. Distribution of the Armenian culinary verbs in the *ARM* corpus

If every time the most frequent verb in the *ARM* corpus, viz. *ephel*, would be chosen, the probability of a correct prediction would be around 38,5% resulting in 65,5% of incorrect predictions (see Fig. 3 below). Here too, the *CIT* model significantly improved the percentage of "correct predictions", viz. by 31,5%.

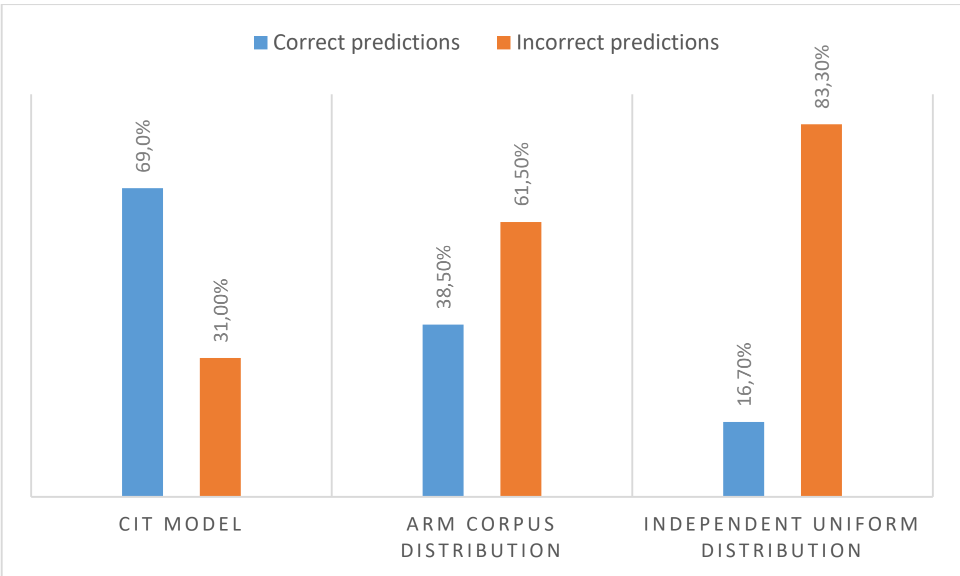
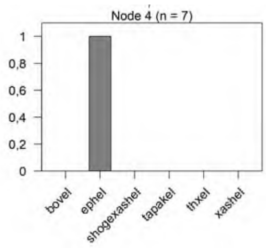
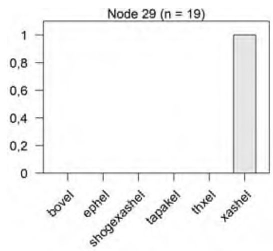


Fig. 3. *CIT* model in comparison with regard to correct and incorrect prediction of the verb

The 26 clusters identified by the *CIT* model as contexts are elaborated upon in three groups organized according to “error rate”. Thus, the first group includes clusters with the lowest error rate of 0% to approximately 1/3. The second group constitutes of clusters with moderate error rate (from 1/3 up to 1/2). The third group of clusters is relatively small and includes 4 clusters with error rates above ½ of the incorrect guesses.

4.4.1 The First Group of Clusters Determined by Context Parameters

The first group of context clusters determined by certain parameter combinations includes context types with the lowest error rate ranging from 0% up to 30%. Based on the excel tables of the clusters generated by the *partykit* package of *R* in addition to the *CIT* graph (Fig.1), the path of each cluster is trackable from top to bottom. It could be rather complicated to track the path without these tables, due to the size and complexity of the distribution on the *CIT* graph. Besides, the accompanying tables allow to observe the distribution in a specific node, identifies the most frequent verb in context, specifies the number of occurrence in the node, indicates the p_value affecting the partitioning in that node, and reveals the error rate – the deviance from the dominant culinary verb in the given cluster.[136] 10 clusters with overall 351 occurrences represent the first group with the lowest error rate. For the sake of legibility, the group will be split into two partial tables (partial Tables 1.1 and 1.2).

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluste/Path
4	eph	7	0	bov = 0, eph = 7, sho = 0, tap = 0, thx = 0, xas = 0		Heat_source (djeroc), Substance (arganak, djur, kath, sous)
29	xas	19	0	bov = 0, eph = 0, sho = 0, tap = 0, thx = 0, xas = 19		Utensil (dzevaman, kathsa, N_A_UT), Substance (N_A_SUB, Manner koghm, N_A_MAN, thethev, xarnel), Heat_source (N_A_HS), RES (pind),

[136] The p-value for each cluster is not included in the tables introducing the cluster groups, however, full information on the portioning of the nodes and their p-values could be found in Appendix 3 of this work. Instead, the tables here include the image of the described cluster cut from the whole *CIT* model to ease the reading as well as provide visualization.

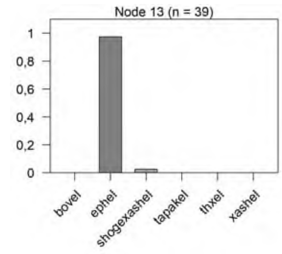
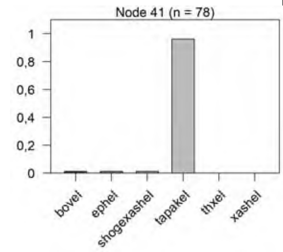
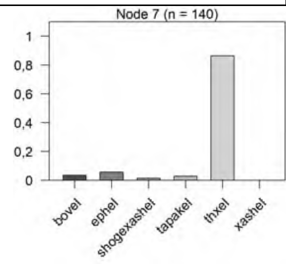
13	eph	39	2,564 1025 6	bov = 0, eph = 38, sho = 1, tap = 0, thx = 0, xas = 0		Heat_source (krak, N_A_HS), Utensil (dzevaman, kathsa, N_A_UT), Manner (N_A_MAN, xarnel), Substance (arganak, kath),
41	tap	78	3,846 1538 5	bov = 1, eph = 1, sho = 1, tap = 75, thx = 0, xas = 0		Utensil (N_A_UT, thava), Substance (dzeth), Heat_source (krak, N_A_HS), Manner (koghm, N_A_MAN),
7	thx	140	13,57 1428 6	bov = 5, eph = 8, sho = 2, tap = 4, thx=121, xas = 0		Heat_source (djeroc), Substance((N_A_SUB), RES (N_A_RES, phaphuk, voskegujn)

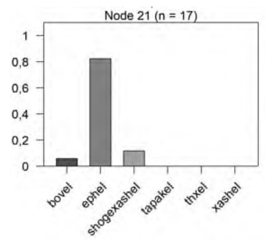
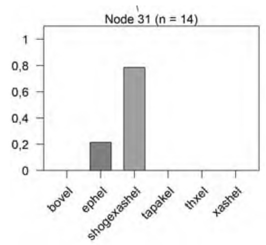
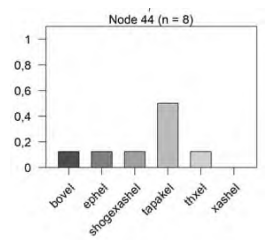
Table (partial) 1.1. The first group of context-parameter determined clusters with the lowest error rate

The first two clusters in this group with (under nodes 4 and 29) with 7 and 19 occurrences each, identify the verbs *ephel* and *xashel* (to boil* and to cook in water*) as the only culinary verb in their respective nodes. Since the clusters are relatively small with all examples exclusively with the identify dominant verbs, the path will not be expanded here in detail.

The next cluster under node 13 with merely 2,5% error rate, singles out the verb *ephel* (to boil*, to cook*) as the most frequently used verb in context determined by the combinations of *Heat source*, *Utensil*, *Manner*, and *Substance* parameters, viz. {krak, N_A_HS} (heat*; fire*)x{dzevaman, kathsa, N_A_UT} (baking form(dish)); pot)x{N_A_MAN; xarnel} (stir)x{arganak, kath} (broth/stock; milk). Node 41 exposes the next context type with overall 78 occurrences where almost all examples are exclusively with the most frequently used verb in context, viz. *tapakel* (to fry), giving rise to as low as only 3,8% deviance in the distribution. The cluster is determined by possible combinations of the parameters *Utensil*, *Substance, Heat source*, and *Manner* with their values {N_A_UT; thava} (pan)x{dzeth} (oil)x{krak, N_A_HS} (heat*; fire*)x{koghm; N_A_MAN} (side_s).

Node 7 represents a considerably large cluster with more than 1/3 of the overall occurrences in the first group determined by particular context parameter combinations of {djeroc} (oven)x{N_A_SUB}x{N_A_RES, phaphuk, voskegujn} (soft; golden) where *thxel* (to bake) with only 13,5% error rate is identified as the most frequently used verb in context.

The next five clusters in the group are relatively small, therefore they are not elaborated in detail, even though the path is provided in partial Table 1.2. However, the ones under the nodes 44, 41 and 48 with occurrences ranging from 9 to 18 occurrences identify *tapakel* (to fry) with 22,2%-30% error rate as the dominant verb in context. The remaining two clusters with 17 and 14 occurrences each are represented under the nodes 21 and 31. Even though these clusters are considerably small, most of the examples are respectively with the verbs *ephel* (to boil*, to cook*) and *shogexashel* (to braise*, to cook with little water*) identify in the context types as the most frequently used verb (partial Table 1.2).

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
21	eph	17	17,6 4705 88	bov = 1, eph = 14, sho = 2, tap = 0, thx = 0, xas = 0		Utensil (dzevaman, kathsa, N_A_UT), Substance (N_A_SUB), Manner (koghm, N_A_MAN, thethev, xarnel), RES (N_A_RES, phaphuk, pind), Heat_source (krak)
31	sho	14	21,4 2857 14	bov = 0, eph = 3, sho = 11, tap = 0, thx = 0, xas = 0		Heat_source(krak, N_A_HS), Utensil (dzevaman, kathsa, N_A_UT), Substance (N_A_SUB), Manner (phak)
44	tap	9	22,2 2222 22	bov = 0, eph = 0, sho = 2, tap = 7, thx = 0, xas = 0		Utensil (N_A_UT, thava), Substance (dzeth), Heat_source (krak, N_A_HS), Manner (phak, xarnel)

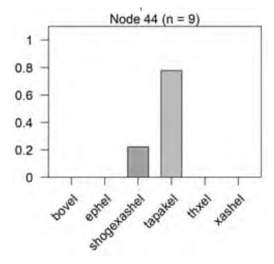
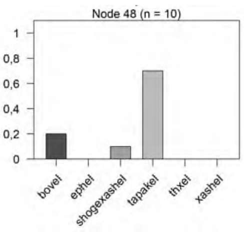
41	tap	18	27,7777778	bov = 0, eph = 2, sho = 0, tap = 13, thx = 3, xas = 0		Utensil (N_A_UT, thava), Substance (dzeth), Heat_source (djeroc)
48	tap	10	30	bov = 2, eph = 0, sho = 1, tap = 7, thx = 0, xas = 0		Utensil (N_A_UT, thava), Substance (jugh), Heat_source (N_A_HS), Heat_intensity (N_A_HI), RES (karmrely, phaphuk, voskegujn)

Table (partial) 1.2. The first group of context-parameter determined clusters with the lowest error rate

4.4.2 The Second Group of Clusters Determined by Context Parameters

This is the largest group of context types comprising 12 clusters, nevertheless, with overall 194 occurrences. Consequently, the clusters determined by certain parameter combinations are relatively small with moderate error rate ranging from 33,3% to 50%.

In the first two clusters under nodes 16 and 36 with 12 occurrences each, the Armenian culinary verb *shogexashel* (to braise*, to steam*, to cook in little water*) is identified as the most frequently used verb in context (partial Table 2.1). The verb *shogexashel* is also identify as the most frequent verb in context in the cluster under node 15, however with 45,5% error rate due to shared parameters values with the rest of the verbs in the distribution. Node 51 exposes a somewhat larger cluster than the aforementioned ones with overall 32 occurrences where the verb *bovel* is selected as the quantitatively dominant verb in context. Since this context type is determined by possible combinations of *Utensil* x *Substance* x *Heat source* x *Heat intensity*, viz. their respective values {N_A_UT, thava} (pan) x {jugh}(oil) x {N_A_HS} x {N_A_RES, shaganakagujn} (brown) the distribution in the cluster is also shared by the verb *tapakel* (to fry) giving rise to 34,3% deviance from the identified dominant verb *bovel* (to roast). The addition of the values {thethev, xarnel} (light/ly, to stir) of the parameter *Manner* to the ones determining the aforementioned cluster (node 51), the verb *bovel* (to roast) is here identified as the most frequently used verb in context cluster under the node 27, however, with 40% error rate and 6 occurrences out of the overall 10 ones. The nodes 19, 35 and 48 display three small clusters with 11 overall occurrences each, where *tapakel* (to fry) with moderate error rate of 36,3%-45,5% is identified as the dominant verb in context determined by their respective specific parameter combinations provided in the path for each cluster. Two quite small clusters under the nodes 6 and 46 will not be considered here due both their size and diversity of distribution, e.g. in the latter one almost all Armenian culinary verbs except for *xashel* (to boil*, to cook in water*) share the distribution.

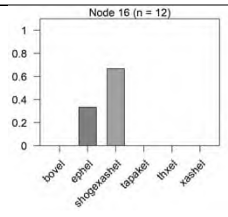
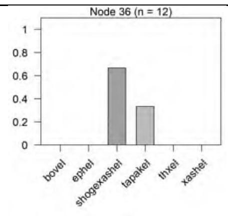
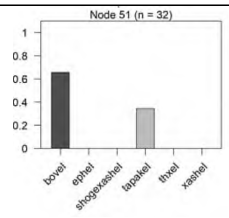
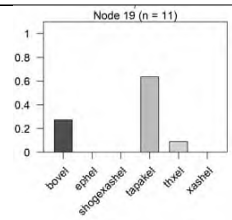
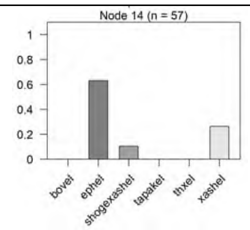
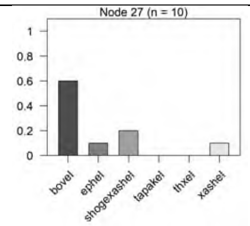
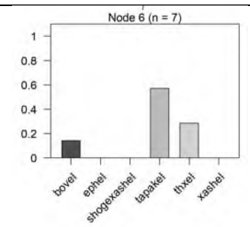
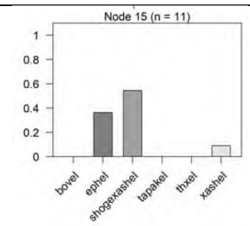
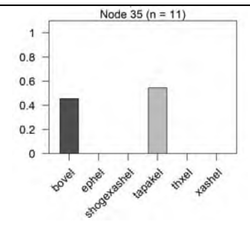
Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster/Path
16	sho	12	33,3333333	bov = 0, eph = 4, sho = 8, tap = 0, thx = 0, xas = 0		Heat_source (krak, N_A_HS), Utensil (dzevaman, kathsa, N_A_UT), Substance (sous)
36	sho	12	33,3333333	bov = 0, eph = 0, sho = 8, tap = 4, thx = 0, xas = 0		Substance (arganak, djur, kath, N_A_SUB, sous), Heat_source (krak, N_A_HS), Utensil (thava), Manner (phak)
51	bov	32	34,375	bov = 21, eph = 0, sho = 0, tap = 11, thx = 0, xas = 0		Utensil (N_A_UT, thava), Substance (jugh), Heat_source (N_A_HS), Heat_intensity (N_A_HI, RES (N_A_RES, shaganakagujn)
19	tap	11	36,3636364	bov = 3, eph = 0, sho = 0, tap = 7, thx = 1, xas = 0		Heat_source (krak, N_A_HS, Utensil (dzevaman, kathsa, N_A_UT), Substance (N_A_SUB), Manner (koghm, N_A_MAN, thethev, xarnel), RES (karmrely, shaganakagujn, voskegujn)

Table (partial) 2.1. The second group of context-parameter determined clusters with moderate error rate

The largest cluster in this group is observed under the node 14 with overall 57 occurrences determined by possible combinations of the context parameter values {krak, N_A_HS} (heat*; fire*)x{dzevaman, kathsa, N_A_UT} (baking form (dish)x{N_A_MAN, xarnel} (to stir)x{djur} (water), *ephel* (to boil*, to cook*) is identified as the most frequent verb with the overwhelming majority of the annotated examples with just 36% error rate. Two semantically closer verbs, viz. *xashel* (to boil*, to cook in water*) and *shogexashel* (to braise*, to cook with little water*, to steam*) also share the distribution in this cluster (partial Table 2.2) (cf. *M(CA)* on the Armenian verbs in Chapter 4.2 of this book).

14	eph	57	36,84 2105 3	bov = 0, eph = 36, sho = 6, tap = 0, thx = 0, xas = 15		Heat_source (krak, N_A_HS), Utensil (dzevaman, kathsa, N_A_UT), Manner (N_A_MAN, xarnel), Substance (djur),
27	bov	10	40	bov = 6, eph = 1, sho = 2, tap = 0, thx = 0, xas = 1		Substance (N_A_SUB, Heat_source (N_A_HS, RES N_A_RES), Utensil (dzevaman, N_A_UT), Manner (thethev, xarnel)
6	tap	7	42,85 7142 9	bov = 1, eph = 0, sho = 0, tap = 4, thx = 2, xas = 0		Heat_source (djeroc, Substance (N_A_SUB, RES karmrely, shaganakagujn)
15	sho	11	45,45 4545 5	bov = 0, eph = 4, sho = 6, tap = 0, thx = 0, xas = 1		Heat_source (krak, N_A_HS, Utensil (dzevaman, kathsa, N_A_UT, Substance (arganak, djur, kath, Manner (phak, thethev)
35	tap	11	45,45 4545 5	bov = 5, eph = 0, sho = 0, tap = 6, thx = 0,		Substance (arganak, djur, kath, N_A_SUB, sous), Heat_source (krak, N_A_HS, Utensil (thava),

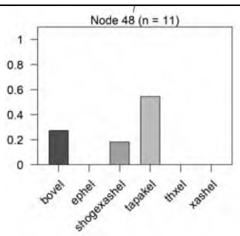
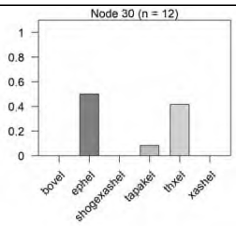
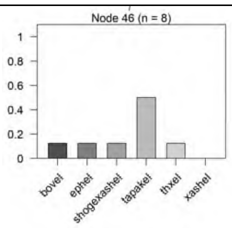
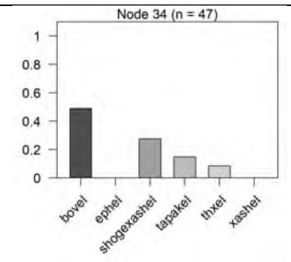
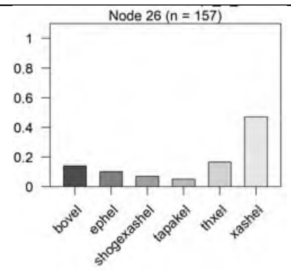
				xas = 0		Manner (koghm, N_A_MAN, thethev, xarnel(, RES (shaganakagujn, voskegujn)
48	tap	11	45,45 4545 5	bov = 3, eph = 0, sho = 2, tap = 6, thx = 0, xas = 0		Utensil (N_A_UT, thava, Substance (jugh), Heat_source (N_A_HS, Heat_intensity (midjin_krak, thujl_krak, ujhegh_krak)
30	eph	12	50	bov = 0, eph = 6, sho = 0, tap = 1, thx = 5, xas = 0		Utensil (dzevaman, kathsa, N_A_UT, Substance (N_A_SUB, Manner (koghm, N_A_MAN, thethev, xarnel), Heat_source (N_A_HS, RES phaphuk),
46	tap	8	50	bov = 1, eph = 1, sho = 1, tap = 4, thx = 1, xas = 0		Utensil (N_A_UT, thava, Substance (jugh), Heat_source (djeroc, krak)

Table (partial) 2.2. The second group of context-parameter determined clusters with moderate error rate

4.4.3 The third group of context clusters determined by parameter combinations

Four clusters with overall 233 occurrences are included in the third group of context types with considerably higher error rate, viz. from 51% up to 66,6%. Irrespective of its error rate, the clusters in this group were worth our attention as it is still amelioration of the 83,3% error rate of the correct verb choice under the hypothesis of discrete uniform independent distribution and the exclusive choice of the most frequent verb in the *ARM* corpus. As mentioned at the beginning of this chapter, the cluster under the node 26 with 52,8% error rate were not taken into consideration due to the absence of verbalized data in all its determining parameters, except for the value {dzevaman} (baking form (dish) of the *Utensil* parameter, making it impossible to draw reasonable conclusions on the distribution of the verbs in this context type. From the other three clusters, the one under node 34 is the largest with overall 47 occurrences where *bovel* (to roast) is identified as the most frequently used verb in context. Due to common parameter values, other Armenian cooking verbs, except for *xashel* and *ephel*, also share the distribution rising the deviation from the dominant verb *bovel* to around 51%. The remaining two clusters, under nodes 38 and 28 are rather small. In the cluster node 38, *thxel* (to bake), however, with 64,2% error rate, is identified as quantitatively the dominant verb in context determined by {dzeth, jugh} (oil, fat*, ghee*, butter*)x{dzevaman, kathsa} (baking form (dish)) values of the *Substance* and *Utensil* parameters. Here too semantically closer verbs share the distribution, e.g. *bovel* (to roast), *tapakel* (to fry). In the last cluster under node 28, *ephel* (to boil*, to cook) is identified – with 66,6% error rate – as the most frequently used verb in context. The verb *xashel* (to boil*, to cook in water*), however, also shares the same amount of occurrences in the distribution as *ephel*, yet *R* gave alphabetical preference to *ephel*.[137]

Node	Prediction	Weight	Error	Distribution	Cluster	Node_cluster_path
34	bov	47	51,063 8298	bov = 23, eph = 0, sho = 13, tap = 7, thx = 4, xas = 0		Substance (arganak, djur, kath, N_A_SUB, sous), Heat_source (krak, N_A_HS), Utensil (thava), Manner (koghm, N_A_MAN, thethev, xarnel), RES (karmrely, N_A_RES, phaphuk),
26	xas	157	52,866 242	bov = 22, eph = 16, sho = 11, tap = 8, thx = 26, xas = 74		Substance (N_A_SUB), Heat_source (N_A_HS, RES (N_A_RES), Utensil (dzevaman,

137 On the semantic closeness of the verbs *ephel* and *xashel*, see the *Context-conditional correlation graph* (CCCG) analysis of the Armenian verbsin Chapter 4.3 of this book.

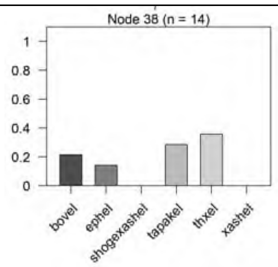
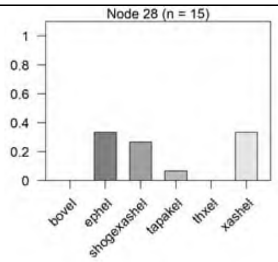
						N_A_UT), Manner (N_A_MAN,
38	thx	14	64,285 7143	bov = 3, eph = 2, sho = 0, tap = 4, thx = 5, xas = 0		Substance (dzeth, jugh), Utensil (dzevaman, kathsa)
28	eph	15	66,666 6667	bov = 0, eph = 5, sho = 4, tap = 1, thx = 0, xas = 5		Substance (N_A_SUB, Manner (koghm, N_A_MAN, thethev, xarnel), Heat_source (N_A_HS, RES (N_A_RES), Utensil (kathsa)

Table (partial) 3.1. The third group of clusters with the highest error rate

Chapter 5 Identifying Possible Translation Equivalents of the Languages in Comparison

One of the objectives of the following chapter is to identify possible equivalents of the culinary verbs in the explored languages based on monolingual analysis of language use in the culinary comparable corpora. Even though the definition of the verbs and the classification of A. Lehrer (1972) served as the basis for the initial componential semantic analysis as a first step for comparison on an abstract level, further description of the culinary lexical fields derived from previously annotated occurrences allowed to draw parallels among the culinary verbs on a conceptual level. For the sake of convenience, the definition of the verbs being compared are provided in footnotes in the next sub-chapter on the *CCCG*, where the verbs are compared in pairs to aid the reader in having both the conventional logic-based semantic closeness and the corpus-based comparison at hand. In order to ensure comparison between the German, English, and Armenian culinary verbs with the objective of yielding possible translation equivalents, the values of the annotation parameters were neutralized according to their conceptual content.[139] More precisely, each annotation parameter was neutralized and generalized to a level of conceptual abstraction at which the observed values exist in all three languages. Thus, the reduction of the occurrences for the translation equivalents was not based on the frequency of the given parameter values but rather on the significance of the latter for the interlinguistic analysis, to guarantee the comparison of all three culinary fields, viz. the culinary verbs in three languages. Therefore, the initial neutralization of parameter values for the intralinguistic analysis was disregarded, followed by a renewed reduction of the parameters based on the original annotated occurrences extracted from respective German, English, and Armenian sub-corpora.[140] As a result, the initial overall 4543 examples in the three language have been reduced to 4189.[141] Fig. 1 below displays the quantity of the initially annotated context examples compared to those after the neutralization and reduction of parameter values serving as the basis for identifying possible translation equivalents.[142]

[139] For the description of the annotation parameters see Chapter 2.1.
[140] Chapters 2, 3 and 4 of this work illustrate the description of the German, English, and Armenian sub-corpora.
[141] See Appendix 10 for the full list of annotated examples after neutralization and reduction of parameters for translation purposes.
[142] The significance of specific values, and not their frequency in the annotated corpora, served as the basis for the reduction of the parameters. Consequently, different parameters were reduced according to different frequencies, which resulted in more annotated examples for further contrastive linguistic analysis than those for the intralinguistic description of culinary fields in the respective languages. For instance, the Armenian *Resultative* (adjectives) parameter values {kisov chap patrast} (half-way through, half-way ready*) and {kisaep} (half-cooked), both denoting not thoroughly cooked food, as well as {al dente} in English, have been left out for

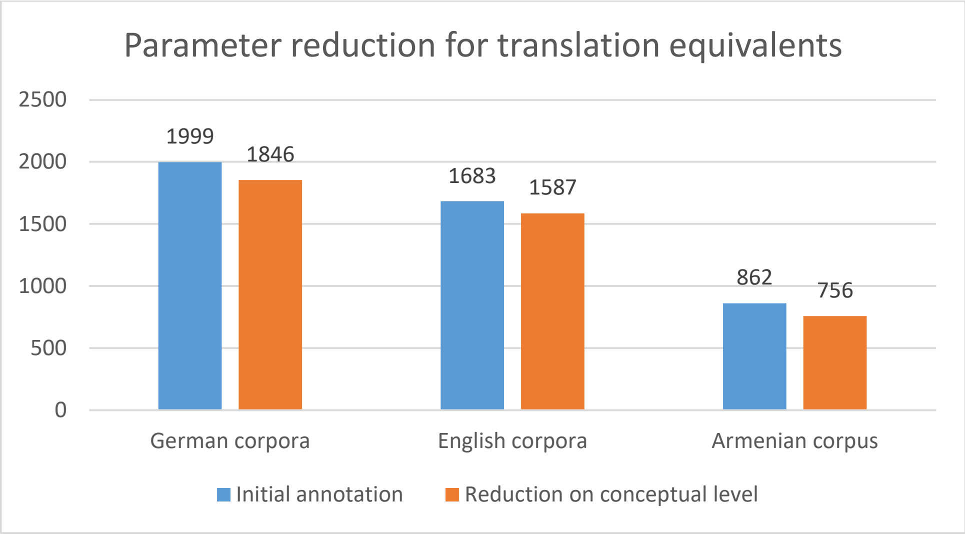


Fig. 1. Parameter reduction on conceptual abstraction level

intralinguistic analysis due to their considerably low frequency, with 4 and 6 occurrences respectively. However, they were retrieved and neutralized with the German {bissfest} (al dente, firm to the bite) under one concept, {al dente}, since this value proved to be significant for comparing the verbs *ephel* (to boil*, to cook) and *shogexashel* (to cook with little water* to steam*, to braise*) *(arm.), braise (en.), kochen* (to boil*, to cook) as well as *schmoren* (to braise) *(dt.)* (see also Fig. 1). As a result, the reduction of examples in German, English, and Armenian ensured more than 98% coverage of the initially annotated and neutralized occurrences, serving as the basis for the contrastive language analysis.

5.1 Visualization of Trilingual Comparison Based on *Correspondence Analysis* (CA)

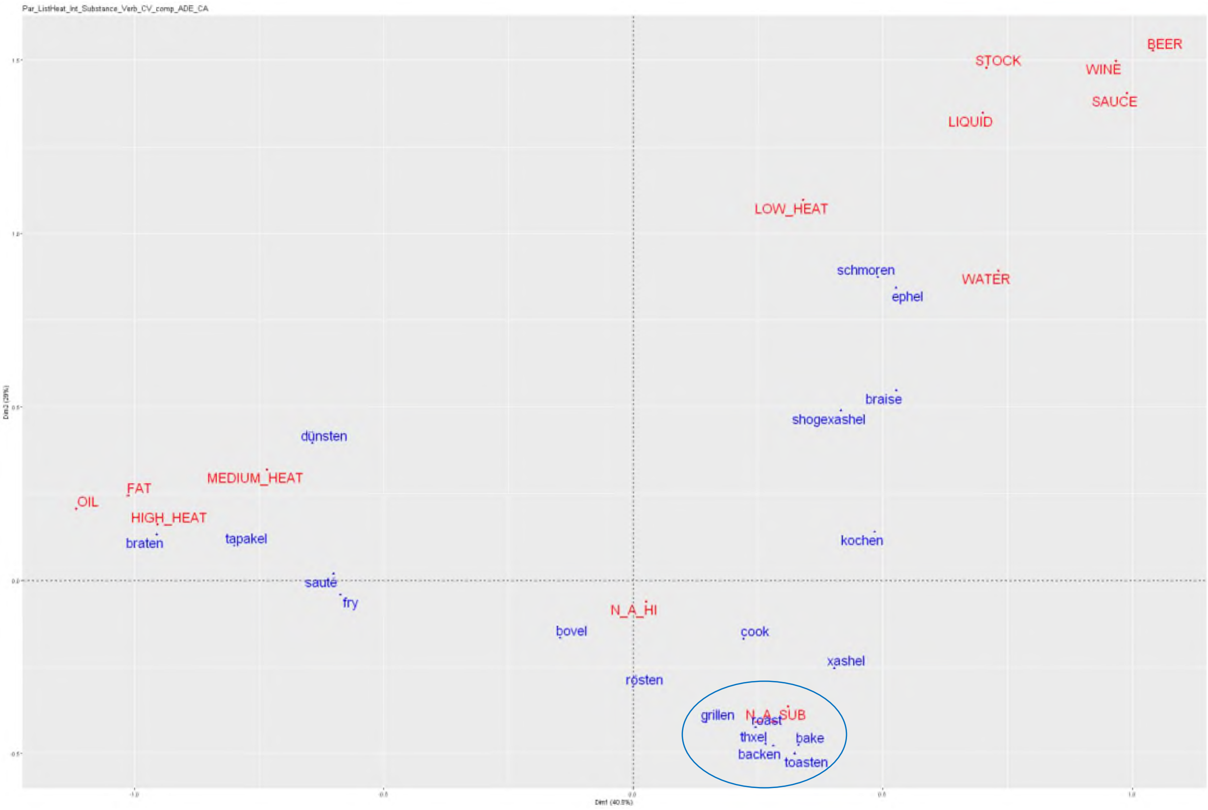


Fig. 2. Possible translation equivalents determined by the parameters *Substance* and *Heat intensity*

Correspondence analysis (CA) illustrates strong correlations between the independent (parameter values, in red) and the dependent (culinary verbs, in blue) variables. The same conventions for the intralinguistic correspondence analysis (see Chapters 2.3, 3.2, 4.2)[143] serve as the basis for identifying strong correlations, viz. the distance of the dependent and independent variables from the intersection, as well as the imaginary angle drawn between them starting from the intersection. Thus, the further the dependent and independent variables are far from the intersection, the sharper angle they form, the stronger the correlation between them. Variables near the intersection indicate the neutral character of the correlation. Therefore, CA contributes to the illustration of the distribution of the culinary verbs in German, English, and Armenian in relation to the given parameter pairs. For instance, the following CA (Fig. 2)

[143] The contrastive language analysis aims at identifying possible translation equivalents for the languages being compared, therefore only *Correspondence Analysis* (CA) shall be considered here since *Multiple Correspondence Analysis* (MCA) illustrates first and foremost strong correlations between the independent variables, viz. annotation parameter values and only then the verbs are distributed to the clusters. This would not by any means contribute to the aforementioned objective.

displays the distribution of the culinary verbs in the aforementioned three languages in relation to the parameters *Substance* and *Heat intensity*. On the upper-right-hand corner of the *CA* graph, the Armenian culinary verb *shogexashel* (to cook with little water*, to braise, to steam) and *ephel* (to cook*, to boil), the German *schmoren* and, partially, *kochen* as well as the English *braise*, create a strong correlation with *liquid*-like cooking substances, e.g. {water, liquid, stock, etc.}, as well as {low heat} from the *Heat intensity* parameter (examples 1–7). In this respect, the aforementioned verbs could serve as possible translation equivalents in culinary contexts with *liquid*-like substances and *low heat* intensity.

1. *Արգանակը* հասցնում ենք եռման աստիճանի, ավելացնում ենք կարտոֆիլն ու 10-15 րոպե *եփում* ենք մինչեւ կարտոֆիլը փափկի: *(ARM)*
   [Bring the *stock/broth* to boil, add the potatoes and *cook** then for 10-15 minutes until they become soft].

2. 20 րոպե *շոգեխաշել ցածր ջերմության վրա*, կափարիչի տակ: *(ARM)*
   [*Braise*/steam** for 20 minutes *over low heat* with the lid on/closed].

3. Mit etwas *Wasser* aufgießen und das Fleisch ca. 90 Minuten *schmoren* lassen. *(detenten13)*
   [Pour some *water* and *braise* the meat for about 90 minutes].

4. Den Deckel auflegen, die Hitze reduzieren und den Fasan *auf kleiner Flamme* etwa 1 Stunde *schmoren* lassen. *(detenten13)*
   [Put the lid on, reduce the heat and *braise** the pheasant *over low heat/flame* for about 1 hour].

5. Den Eintopf *auf kleinem Feuer* ca. 1,5 bis 2 Stunden *schmoren* lassen, bis die Bohnen gar sind. *(detenten13)*
   [*Braise** the one-pot stew *over low heat/fire** for ca. 1,5-2 hours until the beans are ready].

6. Slowly *braised* beef brisket in *red wine* and rosemary is the ultimate special occasion dinner. *(ententen13)*

7. *Braise* the beef brisket in *beer* for two hours. *(ententen13)*

The German verb *braten* (to fry) and the Armenian one *tapakel* (to fry) correlate with the values {oil, fat} x {high heat}[144] from the *Substance* and *Heat intensity* parameter respectively (examples 8–11), thus opposing to *liquid* and *low heat* in cooking. The German verb *dünsten* (to sauté) could also be considered an equivalent to *tapakel* (to fry), but seems to be more strongly correlated on its own with {fat} x {medium heat}. The English verbs *sauté* and *fry* could also be included in this block; however, they are also "pulled" down by the {N_A_HI} from the *Heat intensity* parameter situated near the intersection. The largest block of culinary verbs in the three languages is gathered around the {N_A_SUB} value of the *Substance*, denoting either the absence of any cooking substance or non-verbalized ones,[145] grouping the German *backen*, *grillen*, *toasten*, as well as the English *roast* and *bake*, and the Armenian *thxel* (to bake) in a cluster identifying the aforementioned verbs as possible translation equivalents in the given context (Fig. 2).

8. Մեծ թավայի մեջ տաքացնել *բուսայուղը* եւ *կարագի* կեսը, *տապակել* սոխը, չկարմրեցնել: *(ARM)*
   [In a large pan heat the vegetable *oil* and half of the *butter*, *fry*/sauté** the onion, do not make red*/brown].

9. Կարտոֆիլը լցնել բուսական յուղով տապակի մեջ, եւ այն *տապակել բարձր կրակի վրա*: *(ARM)*
   [Pour the potatoes in a pan with vegetable oil and fry them over high heat].

10. In einer Pfanne etwas *Öl* erhitzen, darin 4 Spiegeleier *braten*. *(detenten13)*
    [In a pan heat some oil and fry 4 eggs sunny-side up].

11. Im Wok oder einer großen Pfanne portionsweise *Öl* erhitzen und das Fleisch *bei starker Hitze braten*. *(detenten13)*
    [In a wok or a large pan heat some oil (portionwise) and fry* the meat over high heat].

[144] The *Heat intensity* parameter value {high heat} was generalized from {bei starker Hitze, bei hoher Hitze, auf höher Flamme} *(dt.)*, {high heat} *(en.)*, and {ուժեղ կրակ, բարձր կրակի վրա} *(arm.)* to ensure the comparison of the culinary fields on the conceptual level in the given languages. For the full list of all parameter generalizations, see Appendix 9.

[145] The differentiation of the N_As in the respective annotation parameters is clarified further in Chapter 5.2. on bilingual comparison based on the *Context-conditional correlation graph* (CCCG) visualization of data analysis. For a small qualitative analysis of the N_As in the *Substance* parameter, differentiated as 1) absence of any cooking substance, 2) non-verbalization of values on the German verb *rösten* (roast), see Chapter 2.4.1 of this work based on the *Mosaic-Plot* visualization method.

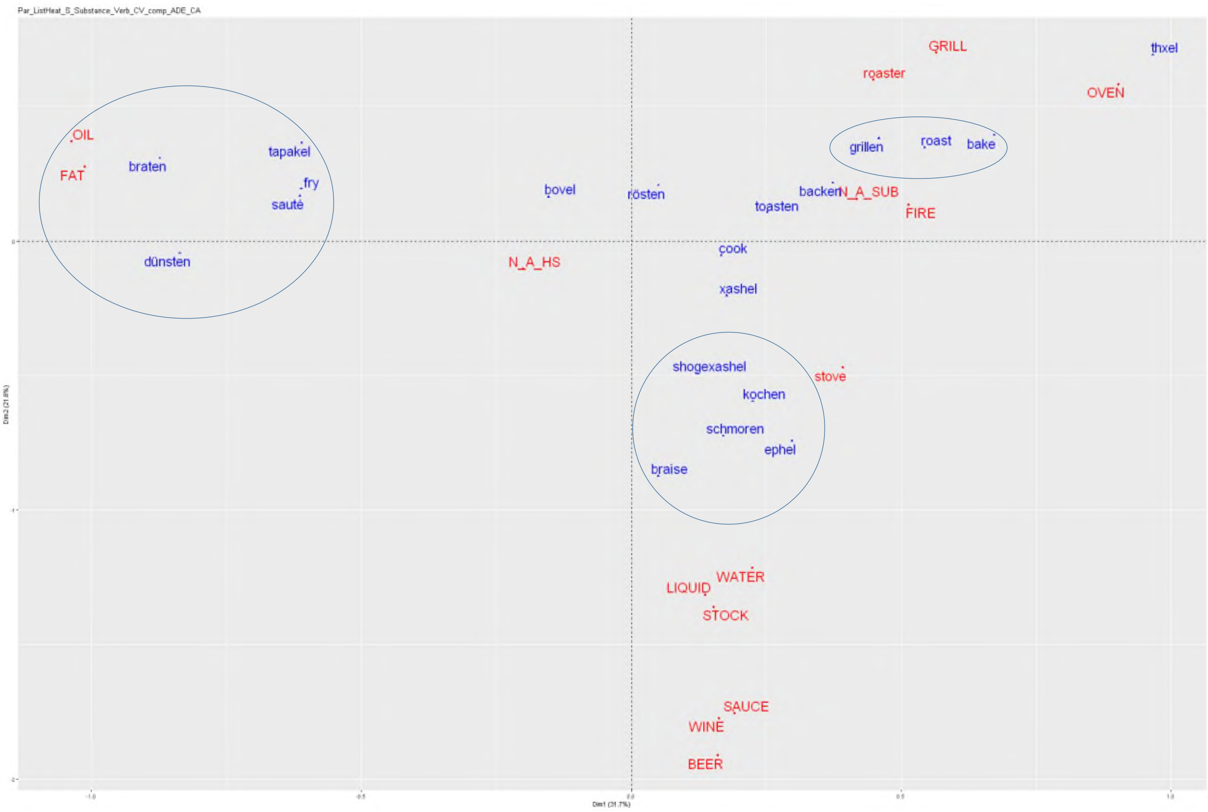


Fig. 3. Possible translation equivalents determined by the parameters *Substance* and *Heat source*

If the preceding *CA* graph (Fig. 2) could not clearly manage to illustrate the possible translation equivalence of the English verbs *sauté* and *fry* for the German *braten* and the Armenian *tapakel* (to fry) due to the *Heat intensity* parameter value {N_A_HI} "pulling" *sauté* and *fry* down toward the horizontal axis and near the intersection, the diagram above (Fig. 3) displays strong correlations between the aforementioned four verbs and the *Substance* parameter values {oil, fat} (examples 12–13).

12. In another frying pan *sauté* onions and carrots in *olive oil*, until golden brown. *(ententen15)*

13. *Fry* the potato slices in *oil* lightly and let them remain crunchy. *(ententen15)*

The Armenian *bovel* (to roast) might also be a possible equivalent to the identified verb group; however, it is "pulled" down and further toward the intersection by both the values {N_A_HS} and {N_A_SUB} from the *Heat source* and *Substance* parameters. Similarly, the German *dünsten* (to sauté) is placed outside the verb group since it is mutually "pulled" by both {oil, fat} and {N_A_HS} (see also Fig. 1). In fact, three distinctive clusters of possible translation equivalent verbs are identified, grouped around a) some kind of *oil* or *fat*, b) some form of *liquid*-like cooking substance, and c) *absence of any cooking substance* or *no verbalization of any substance* (Fig. 3).

Another distinctive group of culinary verbs, viz. the Armenian *shogexashel* (to cook with little water*, to steam*, to braise*) and *ephel* (to boil*, to cook), the German *kochen* (to boil*, to cook) and *schmoren* (to braise), as well as the English *braise*, which could serve as equivalents as they all strongly correlate on the conceptual level with the *Heat source* parameter value {stove}, as well as the *Substance* parameter values {water, liquid, stock, sauce, wine, beer}, which could be roughly generalized as {liquid}. The English *cook* and the Armenian *xashel* could also be included in the aforementioned verb group of possible translation equivalents; however, they are "pulled" upward by the {N_A_SUB} of the *Substance* parameter (cf. examples 14–18).

14. Միջուկը պատրաստելու համար կաղամբը կտրտել, *շոգեխաշել* մի քիչ *ջրի* մեջ, մինչեւ փափկի: *(ARM)*
    [To prepare the stuffing *braise*/sauté** the cabbage with some water until it gets soft].

15. Կարտոֆիլը ավելացնել *արգանակին* եւ եփել 20 րոպե: *(ARM)*
    [Add the potatoes to the *broth* and cook*/boil* for 20 minutes].

16. Unter ständigem Rühren das Risotto nun *kochen* und verdunstete *Geflügelbrühe* immer wieder mit neuer nachgießen. *(detenten13)*
    [Now *cook* the risotto by constantly stirring it by consistently adding poultry *stock* every time it evaporates].

17. Hähnchenbrustfilet auf die Fenchel setzen, 50 ml *Gemüsebrühe* angießen, aufkochen und zudecken 8–10 min. fertig *schmoren*. *(REZ_DE)*
    [Put the chicken fillets on the fennels, pour 50ml vegetable *broth*, bring to boiling, close and *braise* for 8-10 minutes until ready].

18. There are plenty of ways to cook *braised* short ribs, but the classic includes a combination of red *wine*, carrots, celery, onions, garlic, thyme, and canned tomatoes or tomato paste. *(ententen15)*

The last group of verbs serving as possible equivalents are determined by their strong correlations with *no cooking substance*, marked {N_A_SUB}, as well as some type of "air" heating sources.[146] Thus, the German culinary verbs *backen* (to bake), *grillen* (to grill), and *toasten* (to toast), as well as the English *bake* and *roast*, together with the Armenian *thxel* (to bake) form a very sharp angle with the *Heat source* parameter value {oven} and the *Substance* {N_A_SUB} (examples 19–27). *Thxel* is “pulled” by {oven}, and ends up the farthest away from its equivalent counterparts in the plot, since the vast majority (128 out of 163) of the annotated context examples with *thxel* co-occur with {oven}. The German *rösten* (to roast), and partially also the Armenian *bovel* (to roast), are “pulled” by the *Heat source* parameter value {grill, roaster} from the right, as well as the *Substance* ones {oil, fat}, and therefore are placed closed to the intersection.

19. Im auf 180 Grad Celsius vorgeheizten *Backofen* 15-20 Minuten die Pizza *backen*. *(REZ_DE)*
    [*Bake* the pizza for 15-20 minutes in a preheated 180^{o}C *oven*].

20. Damit kann man gleichzeitig Brot *toasten*, Kaffee kochen und Spiegelei braten. *(detenten13)* (no verbalized *Substance* and *Heat source* parameter value)
    [One could parallelly *toast* the bread with it, make coffee and fry eggs].

21. Das Toastbrot *toasten* oder in einer Pfanne ohne Öl braten. *(detenten13)* (shows that the German verb *toasten* is itself without any cooking substance)
    [Toast the bread for toasting* or fry in a pan without oil].

22. Wer knusprige Steaks *grillen* will, sollte einen *Holzkohlegrill* nehmen. *(detenten13)*
    [Who wills to barbecue*/grill* crunchy steaks, should take a charcoal grill].

23. *Roast* the red pepper and garlic in the *oven*. *(ententen15)*

24. Our artisan breads *are baked* in a wood fire brick *oven*. *(ententen15)*

25. The other is to *bake* the potatoes in foil in Convection *oven* 175 °C for 75 minutes. *(ententen15)*

26. Խմորեղենը հարկավոր է միշտ *թխել ջեռոցի* մեջտեղի դարակում: *(ARM)*
    [You should always *bake* the cake* in the middle shelf of the *oven*].

27. Խմորը լցնել ձեւամանի մեջ եւ այն 45−55 րոպե *թխել ջեռոցում*: *(ARM)*
    [Pour the dough into the baking form and bake for 45-55 minutes in oven].

[146] For the differentiation of food being *cooked*, *roasted*, and *rotted* depending on [air+] or [air-] as well as [water+] and [water-], see *“Le triangle culinaire”* (Lévi-Strauss 1965:17), republished in *Food & History* 2 (2004:7-20). Lehrer then develops the triangle by adding *oil* to the aforementioned substances and fits the English verbs in it (1972:169). See also Chapter 1.2 of this work.

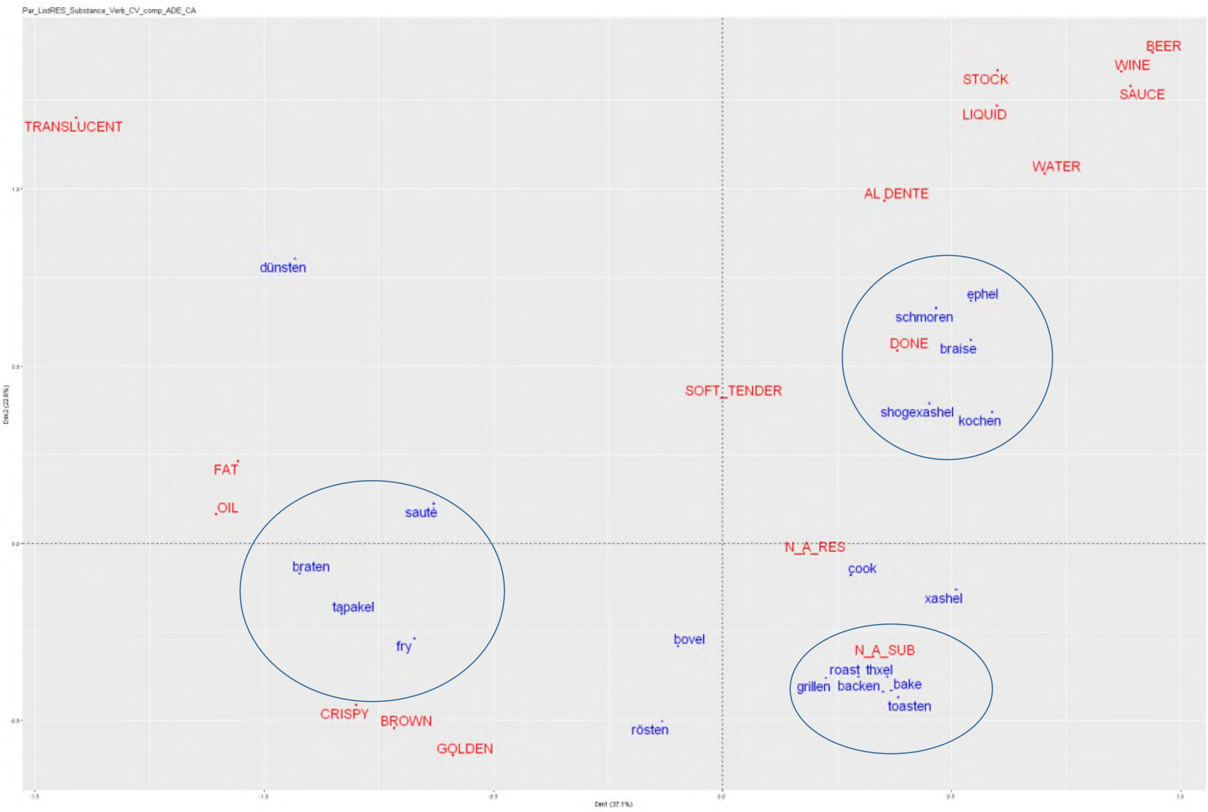


Fig. 4. Possible translation equivalents determined by the parameters *Substance* and *Resultative* (adjectives)

Depending on the distribution of the *Substance* and *Resultative* (adjectives) parameter values, at least three groups (clusters) of translation equivalents are observed in Fig. 4. The first large cluster of potential equivalents of German, English, and Armenian culinary verbs, located in the upper right-hand corner of the graph, strongly correlates with cooking substances denoting some type of *liquid*, e.g. {water, stock} (cf. Fig. 3), as well as with the {done}[147] and {al dente} values of the *Resultative* (adjectives) parameter. In the lower right-hand corner, the next large cluster of culinary verbs is "pulled" away from the intersection by the *Substance* and *Resultative* (adjectives) parameter values {N_A_SUB} and, partially, {N_A_RES}. The English *cook* and the Armenian *xashel* are displayed considerably apart from the cluster as there are strongly "pulled" by {N_A_RES} and, to a lesser extent, by {N_A_SUB}. Thus, *cook* and *xashel* (to cook, to boil in water*) could also be considered as a small cluster of English and Armenian translation equivalents. The left-hand side of the following graph distinguishes clusters of possible verb equivalents depending on the distribution of the *Resultative* (adjectives) and

[147] As most of the values of all six annotation parameters were generalized on the conceptual level to ensure comparison among the verbs, {done} value from the *Resultative* (adjectives) parameter also comprises such entries as {fertig} (de.), {minchev pathrast linely} (until it is ready) (arm.), {thoroughly cooked}, {cooked through} (en.), etc.

Substance parameter values. The German *braten*, the Armenian *tapakel*, as well as the English *fry* and *sauté* form a trilingual cluster of translation equivalents correlating strongly with the *Substance* parameter value {fat, oil}. The identified verbs in the cluster also strongly correlate with the *Resultative* (adjectives) parameter values {crispy, brown, golden}, which "pull" them down and away from the German verb *dünsten*. The latter also strongly correlates with the *Substance* parameter value {fat, oil}; however, its correlation with the {translucent} value from the *Resultative* (adjectives) parameter is much stronger, forming an imaginary line from the intersection between them. This means that, out of the overall 155 co-occurrences of {translucent} and the Armenian, English, and German verbs, 95 are with the German verb *dünsten*. The English verb *sauté*, for instance, co-occurs with {translucent} 34 times. The Armenian *bovel* and the German *rösten* are also correlated with {crispy, brown, golden}; however, they are "pulled" by the *Substance* parameter value {N_A_SUB} further towards the right, thus parting from the cluster due to *the absence of any cooking substance and/or non-verbalized values*. Therefore, they are located between the other two groups of verbs, closer to the vertical axis.

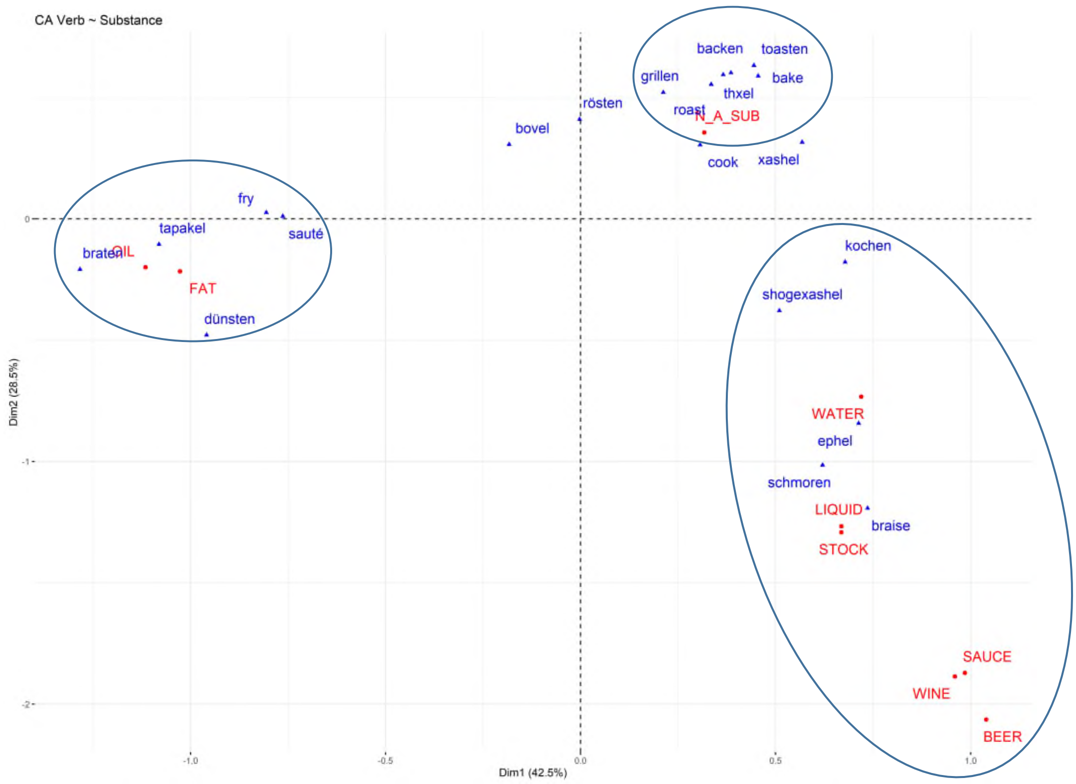


Fig. 5. Possible translation equivalents determined by the parameter *Substance*

So far the *CA* of the German, Armenian, and English culinary verbs was carried out and illustrated based on correlations between two parameters, for instance Fig. 3 and Fig. 4 above. Since the parameter *Substance* proved decisive in differentiating verbs in both monolingual and contrastive linguistic analyses, especially for identifying potential translation equivalents, this *CA* (Fig. 5) aims to illustrate the possible changes in the distribution of the verbs driven by only one parameter. Here, the analysis reveals three distinct clusters of potential translation equivalents, based on strong correlations between the dependent (verbs in blue) and the independent variables (values marked in red), i.e. grouped around {fat, oil} on the right-hand side of the plot, as well as *water*-like cooking substances on the lower left-hand side. The upper right-hand side of the *CA* plot shows potential trilingual translation equivalents gathered around the *Substance* parameter value {N_A_SUB}, denoting either absence of verbalized values in this parameter or absence of any cooking substance. This is a rather large cluster; however, due to the specific character of certain verbs (e.g. implying *water* in their encyclopedic definition, therefore not verbalized and consequently annotated as {N_A_SUB}), the Armenian *xashel* is also near this cluster. The German verb *dünsten* is here undoubtedly identified as a potential translation equivalent in the cluster around {fat, oil}, together with the German *braten*, the Armenian *tapakel*, as well as the English *sauté* and *fry* (Fig. 5). The *CA* above (Fig. 5) restates the Armenian *shogexashel*, *ephel*, the English *braise*, as well as the German verbs *kochen* and *schmoren* to be potential translation equivalents gathered around water-like cooking substances (cf. Fig 4).

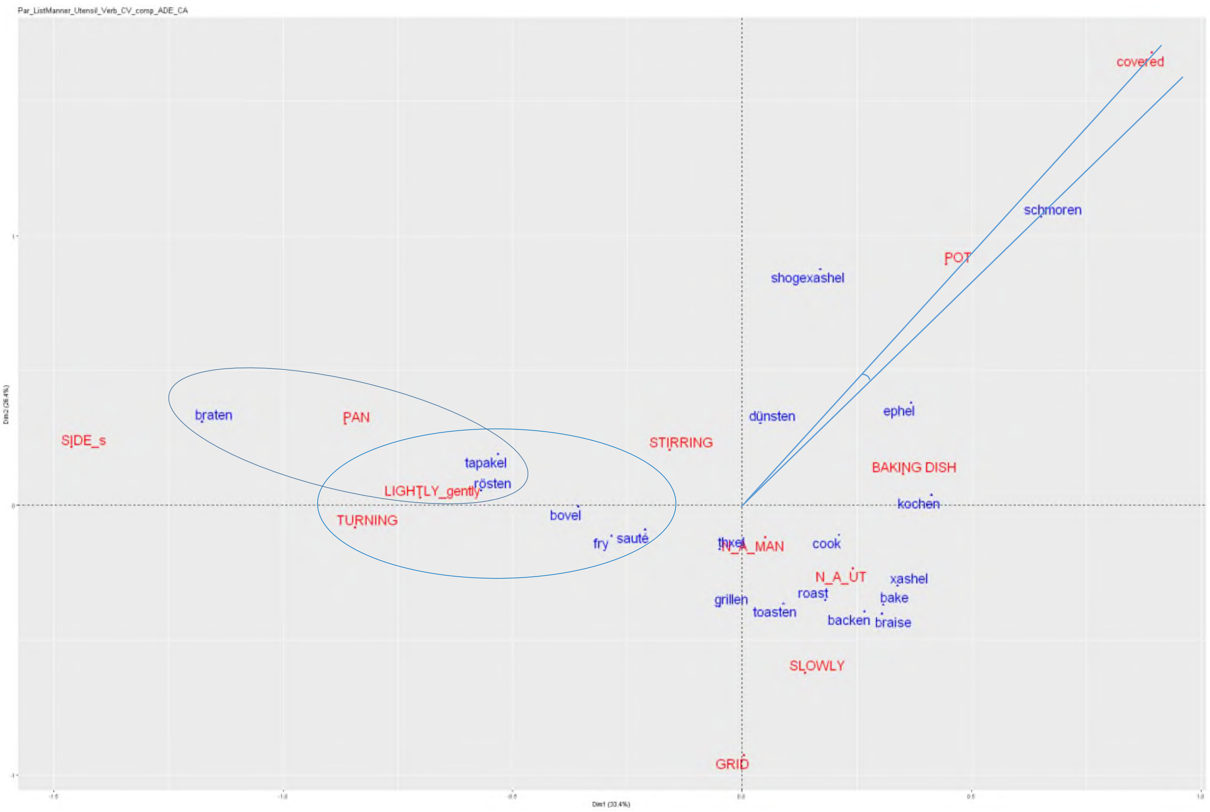


Fig. 6. Possible translation equivalents determined by the parameters *Utensil* and *Manner*

Fig. 6 illustrates the distribution of the German, English, and Armenian culinary verbs according to the parameters *Utensil* and *Manner* with respect to their translation potential.

On the left-hand side of the *CA* graph (Fig. 6), a small group of possible equivalent verbs is distinguished, viz. the German verbs *braten* and *rösten*, as well as the Armenian *tapakel*, strongly correlating with the values {lightly_gently} and {side_s},[148] as well as {pan} from the *Manner* and *Utensil* parameters respectively. However, *bovel* (arm.) together with *sauté* and *fry* (en.) could also be integrated into the group to form a larger cluster of possible equivalents in the three languages (cf. Fig. 7 below). The distribution of these three verbs is affected by their strong correlation with the *Utensil* {pan} and the *Manner* parameter values {turning, lightly_gently}, pulling them away from the group. On the right-hand side of the CA graph, a somewhat scattered group of possible translation equivalents, viz. *schmoren* (dt.), *shogexashel*

[148] The *Manner* parameter value {side_s} was generalized to include values such as {on both sides}, {on one side}, {on two sides}, {on the other side} (en.), {երկու կողմից} (arm.), {auf jeder Seite}, {von beiden Seiten}, {beideseitig}, {rundherum} (dt.), enabling trilingual comparison of culinary verbs on a conceptual level. For a full list of generalization, see Appendix 9.

and *ephel* (arm.), is identified, strongly correlating with the {pot} and {covered} values from the *Utensil* and *Manner* parameters respectively.[149]

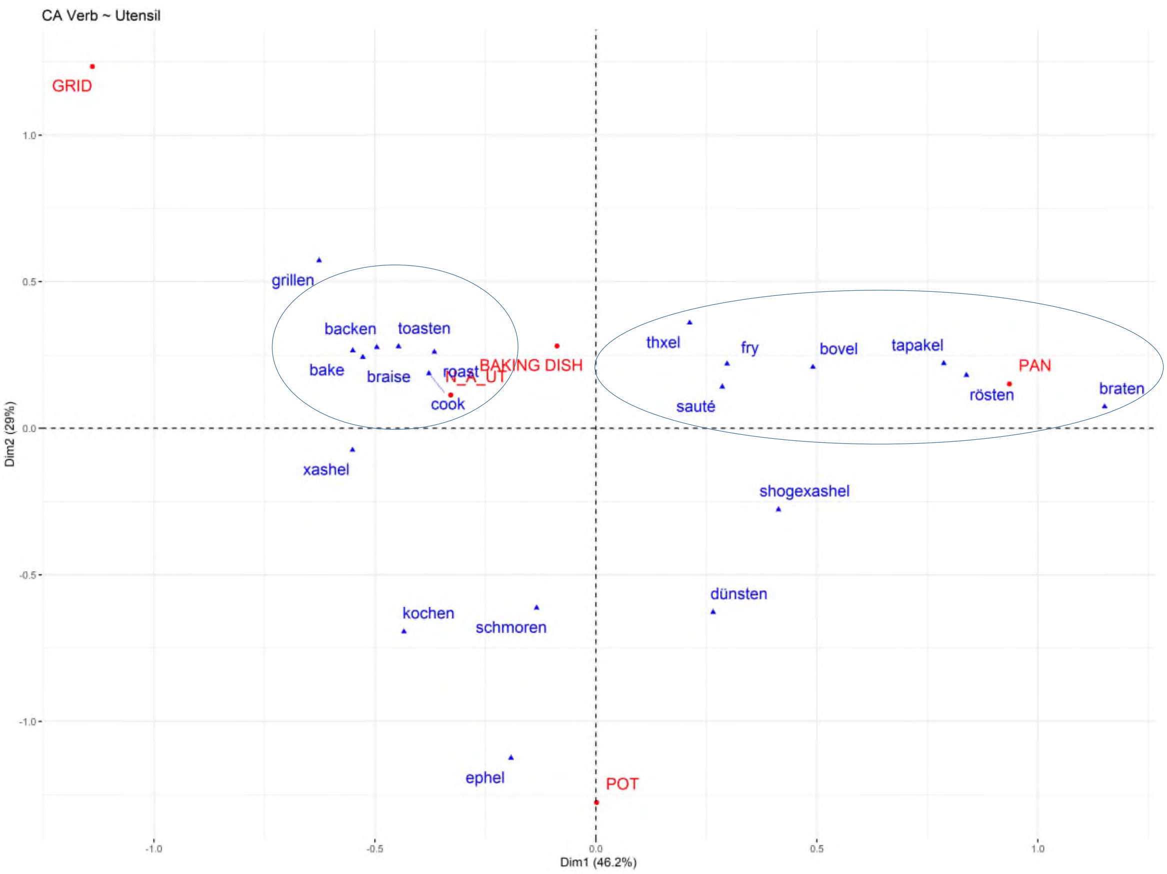


Fig. 7. Possible translation equivalents determined by the parameter *Utensil*

By excluding the parameter *Manner* from the *CA* of the eight German and six Armenian and English verbs, larger clusters of potential translation equivalents are identified gathered around a certain utensil (Fig. 7). The upper right-hand side of the *CA* plot identifies a considerably large cluster of potential trilingual equivalents strongly correlating with the *Utensil* parameter value {pan}. The lower side of the *CA* graph could be divided into two clusters of potential equivalents (on the left-hand and right-hand corners) or into a larger scattered cluster with a strong correlation with the *Utensil* parameter value {pan}. The upper right-hand side of the graph (Fig. 7) groups the identified verbs around {N_A_UT},[150] denoting either no verbalized

[149] The distribution of the German culinary verb *schmoren*, placed rather apart from the other verbs in the group, is due to its strong correlation with the *Manner* parameter value {covered} ({zugedeckt}) before generalization on the conceptual level, forming a sharp angle with it.
[150] Partially also the *Utensil* parameter value {baking dish}, which is, however, somewhat near the intersection.

values in this parameter or absence of any utensil in the process, e.g. *roast* (en.) over the open fire. The German verb *grillen* is torn away from the cluster due to its considerably strong correlation exclusively with the *Utensil* parameter value {grid}. Out of the total 26 co-occurrences of {grid} and the Armenian, German, and English verbs, 22 are with *grillen*.

5.2 Bilingual Comparison Based on *Context-Conditional Correlation Graph* (CCCG) Visualization of Data Analysis

As stated in Chapters 2.4, 3.3 and 4.3, the *Context-Conditional Correlation Graph* is another visualization technique based on statistical evaluation of data, suitable for both intralinguistic and contrastive analysis. In the former case, all relevant parameters could be evaluated in one graph, ensuring the comparison of the culinary verbs in pairs within the respective language. In the case of interlinguistic contrastive analysis, the *CCCG* aided in displaying the observed potential translation equivalents by comparing culinary verbs in bilingual pairs. Moreover, due to the *CCCG*'s visualization of the differences between the verbs being compared, it was possible to identify those verbs that are far from being potential translation equivalents, viz., with respect to which parameters the verb in the target language would be an inappropriate lexical choice for the verb in the source language, generating severe translation mistakes. Thus, the *CCCGs* illustrate general tendencies in specific pairs of verbs, representing German-Armenian, Armenian-German, English-Armenian, Armenian-English, German-English, English-German comparisons, which can contribute to making correct decisions in individual cases, namely the presence or absence of which specific parameter values that favor a particular verb as a potential translation equivalent or, rather, exclude it.

In order to illustrate the predominant tendencies in the distribution and to ensure legibility, only deviations exceeding 3.0 times the standard deviation were taken into consideration, thus displaying only the parameters with a significant impact on the distribution. In order to avoid redundancy, one general (hypernym) and one specific (hyponym) verb in each language pair will be considered. In the case of more general verbs – for instance, *backen* (dt.) (bake) or *cook* (en.) or *ephel* (arm.) (to cook, to boil*) – in many parameters, no values were verbalized and were therefore annotated as N_As. However, the manual semantic annotation did not allow to differentiate whether the specific parameter value was not verbalized because it was already implicated, e.g. by the encyclopedic definition of the verb[151], or the N_As in particular parameters denoted the absence of respective values. For instance, no substance should be used for specific verbs, viz. *bovel* (arm.) (to roast), *rösten* (to roast), *grillen* (dt.) *(to grill)*, *roast* (en.). Consequently, when it came to the comparison of the verbs, the absence of any verbalized

[151] For instance, the German verb *backen* is defined in the *"Wörterbuch zu Verbvalenz"* as 1) *etwas im Backofen herstellen;* 2) *sich irgendwielange im Backofen befinden*, therefore the overwhelming majority of the occurrences of *backen* are annotated as N_As in the parameter *Heat source* as *Ofen* is such a strong correlate of *backen* that it is already obviously assumed. The valency of the verb *backen* could be traced under <https://grammis.ids-mannheim.de verbvalenz/400207/>

values annotated as N_As in specific parameters in some cases resulted in semantically opposite verbs having fewer connected edges on *CCCG*, thus suggesting their semantic closeness.[152] Such instances were specifically taken into consideration, e.g. accompanied by *CCCGs* of certain verb pairs illustrating deviations in distribution exceeding 2 or even 2.5 times the standard deviation. The contextual examples as well cast light on the aforementioned problematic cases of the *CCCG* vizualisation method.

[152] For all *CCCG* Excel-tables based on language pairs for translations purposes see Appendix 11. Verbs being compared are connected to each other with edges. Low number of edges denote the semantic closeness of the two verbs. Consequently, high number of edges indicate the difference between these verbs in many parameters.

5.2.1 CCCG Visualization of Language Pairs: *German-Armenian*

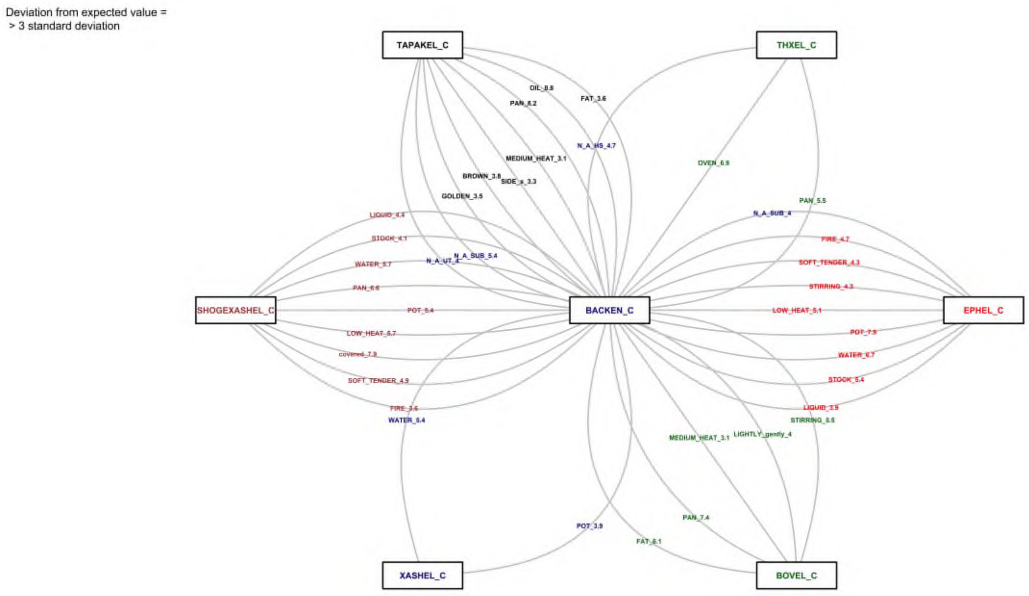


Fig. 1. Potential translation equivalents based on the *German-Armenian* language pair

The *CCCG* above (Fig. 1) compares the German verb *backen* to the six Armenian culinary verbs on the conceptual level (therefore marked as *backen_C*), with the objective to identify potential translation equivalents as one of the outcomes of interlinguistic comparison, as well as to single out specific parameters hindering the translation potential of a specific verb pair. Judging from the edges connecting the German verb *backen* to the Armenian verbs, it is most semantically close to *xashel* (to boil*, to cook in water*) and *thxel* (to bake). However, as mentioned at the beginning of this chapter, due to non-verbalized values in a number of annotation parameters, the *CCCG* visualization might generate fewer edges connecting two semantically opposite verbs. A closer and a more detailed look at the manual semantic annotation of the extracted occurrences sheds light on why *xashel* (arm.) is illustrated as being semantically closer to *backen* (dt.) (to bake), despite their differences. Out of the overall annotated examples of *xashel* and *backen*, with 96 and 245 occurrences respectively, the overwhelming majority of parameters, viz. *Heat intensity*, *Manner*, *Resultative* (adjectives), and *Heat source,* show no verbalized value. As a result, these parameters do not show statistically significant differences between the two verbs, which is why they appear semantically closer in the *CCCG* and may be interpreted as potential translation equivalents. The parameters

Substance and *Utensil*, however, have been decisive in differentiating the verb *xashel* from *backen*. Even though the *Substance* parameter value {water} co-occurs with the Armenian verb *xashel* only 16 times, it is still massively overrepresented – 5.4 times the standard deviation – as it never co-occurs with the German verb *backen* (examples 1–2). Furthermore, in the distribution of the *Utensil* parameter values when comparing the verbs *xashel* and *backen*, the occurrences of {pot} with *xashel* exceed the expected value by 3.9 times the standard deviation (example 2).

1. Եռացող աղ արած *ջրի* մեջ *խաշել* ծաղկակաղամբը 3-5 րոպե: *(ARM)*
   [*Boil*/cook** the cauliflower in boiling salted *water* for 3-5 minutes].

2. *Ջուրը* լցնել *կաթսայի* մեջ, բերել եռման, ավելացնել աղը եւ 10 րոպե *խաշել* սմբուկը, մինչեւ փափկի: *(ARM)*
   [Pour the water into a *pot*, bring to boiling, add salt and *cook** the egg plants until they are soft].

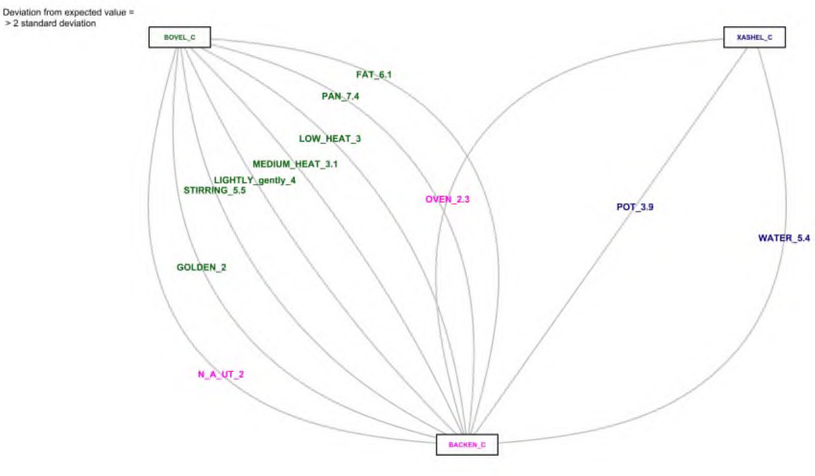


Fig. 1a. Comparison of the verb *backen* with *bovel* and *xashel*

A separate *CCCG* illustrating the comparison of the German culinary verb *backen* with the Armenian *bovel* and *xashel*,[153] displays statistically significant difference in the distribution of the *Heat source* parameter as well. In comparing *xashel* and *backen*, the co-occurrences of {oven} with *backen* are overrepresented by a factor of 2.3 times the standard deviation.

The German *backen* (to bake) and the Armenian *thxel* (to bake) might potentially be translation equivalents as they differ only in the parameters *Utensil* and *Heat source*. Even though the verb

[153] The verb *bovel* is included in this CCCG comparing in pairs *bovel-backen* as a second potential equivalent to *backen* after *thxel* (to bake).

backen shows a strong correlation with the {N_A_HS} value from the parameter *Heat source*, this reflects the case where the value (e.g., {oven}) is not verbalized as it is already implicated, for instance, by the verb's prototypical or defining semantic features and encyclopedic definition.[154] Thus, irrespective of the strong correlation of the *Heat source* parameter value {oven} with the Armenian verb *thxel*, marked by a considerably higher deviation value, the verbs *thxel* and *backen* are potentially semantically closer in this parameter as well, since many of the non-verbalized values here could reasonably be annotated as {oven} (example 3). By contrast, there is no translation potential between the verb *xashel* and *backen* as the co-occurrences of the *Substance* parameter value {water} with xashel are overrepresented by a factor of 5.4 times the standard deviation, whereas *backen* implies the absence of a cooking substance.

3. Den Kuchen jetzt weitere 15 bis 20 Minuten *backen*, bis die Eiersahne gebräunt ist. *(REZ_DE)*
   [Now bake the cake* for another 15-20 minutes until the egg cream turns brown].

The Armenian *thxel*, on the contrary, does not implicate {oven} but can rather be defined as "to cook something from *dough*"[155] with no presupposition concerning any heat source. In the second definition of *thxel*, however, *tonir*[156] (tandoor) is mentioned as a heat source, which is too culture-specific and cannot be generalized as {oven}. Therefore, in the Armenian annotated

---

[154] "Backen"- 1.a) aus verschiedenen Zutaten einen Teig bereiten und diesen unter Hitzeeinwirkung im *Backofen* gar und genießbar machen; b) durch Backen (1a) herstellen; Kuchen, Plätzchen backen (DUDEN Universal Wörterbuch online retrieved on 16.06.2022 <https://www.duden.de/rechtschreibung/backen_herstellen_garen>.
"Backen"- Eine unfertige Speise der Ofenhitze aussetzen, sodass sich eine Kruste bildet und sie essbar wird a) Teig im Ofen (in einer Form) hochgehen und locker, gar werden lassen b)eine Speise im zerlassenen Fett in der Pfanne rösten, braten ("Backen", in: *Wörterbuch der deutschen Gegenwartssprache (1964–1977)* retrieved on 16.06.2022 via *DWDS*, <https://www.dwds.de/wb/backen>*)*.
[155] "thxel" (partial definition)- 1. Խմորից որևէ բան (հաց, գաթա) պատրաստել: [1. To prepare something (bread, gata) from dough] *("թխել"*, in Արդի հայրենի բացատրական բառարան, Է. Աղայան (1976) retrieved on 16.06.2022 via *nayiri.com*, <http://nayiri.com/imagedDictionaryBrowser.jsp?dictionaryId=24&dt=HY_HY&query=%D5%A9%D5%AD%D5%A5%D5%AC>).
"thxel"- 2. Բացած խմորը թոնրի կողը խփելով լավաշ թխել, կողել: [2. To bake lavash by rolling the dough open and hitting on the side of tonir] *("թխել"*, in *Հայկական խոհանոցի բառարան* (2015) retrieved on 16.06.2022 via *nayiri.com* <http://nayiri.com/imagedDictionaryBrowser.jsp?dictionaryId=124&query=%D5%A9%D5%AD%D5%A5%D5%AC>).
[156] "թոնիր"- tonir (a big jar dug in the earth in which fire is made used for baking bread). ("*թոնիր*", in *Հայերեն-անգլերեն բառարան (Armenian-English Dictionary)* (2009), p. 207. Retrieved on 15.07.2022 via library.anau.am, *<https://library.anau.am/images/stories/grqer/bararanner/Baratyan.pdf>).*

examples {djeroc} (oven) is verbalized to specify that something is to be baked (*thxel*) in *oven* and not, for instance, in *tonir* (tandoor) (example 4).

4. *Թխել* թխվածքաբլիթները, 25 րոպե 180 աստիճան տաքացրած *ջեռոցում*: (ARM)
   [*Bake* the cookies in a preheated *oven* at 180ºC for 25 minutes].

However, *thxel* also differs from the German *backen* in the parameter *Utensil* by correlating with {thava} (pan) (example 5), which once again restates the definition of *thxel*, viz. differentiating it from *backen* with the presence of *dough* as an ingredient but not necessarily {oven} as a heat source.

5. Սովորական բլիթների խմորի մեջ 2-3 ճաշի գդալ շաքարավազ ավելացնել եւ *թխել թավայի* վրա, երկու կողմից կարմրեցնել: *(ARM)*
   [Add 2-3 tbl. spoons of sugar to the regular pancake dough and bake in a pan, until both sides are red*/brown].

In comparing the verb pair *backen-bovel*, the Armenian culinary verb *bovel* differs from the German *backen* in the parameters *Substance*, *Utensil*, *Heat intensity, Resultative* (adjectives) and *Manner*. For instance, {fat} is overrepresented with *bovel* by 6.1 times the standard deviation, and {pan} exceeds the expected value by 7.4 times. There is statistically no significant difference between *backen* and *bovel* only in the parameter *Heat source*. Thus, the German *backen* and the Armenian *bovel* have little potential, if any, to be considered as possible translation pairs.

The German verb *backen* differs drastically from the remaining Armenian culinary verbs, viz. *shogexashel*, *ephel*, and *tapakel*. The difference is observed in all six annotation parameters; however, the parameter *Substance* has the highest impact on the distribution of the annotated occurrences between the verb pairs *backen-ephel* and *backen-shogexashel*, since its three separate values, viz. {liquid}, {stock}, and {water} differentiate them from *backen*. The latter, on the contrary, correlates strongly with {N_A_SUB} (no substance), for instance, in the verb pairs *backen-ephel* and *backen-tapakel*. In comparing the *backen* and *tapakel*, the co-occurrence of the *Substance* parameter value {oil} with *tapakel* is overrepresented by 8.8 times the standard deviation. The parameter *Utensil* also strongly differentiates *backen* from *tapakel*, *ephel* and *shogexashel*. For instance, in the case of the *tapakel-backen* pair, the number of occurrences of the *Utensil* parameter value {pan} with *tapakel* exceeds the expected value by 8.2 times the standard deviation. Therefore, the aforementioned Armenian verbs have no potential to be possible translation equivalents for the German verb *backen*.

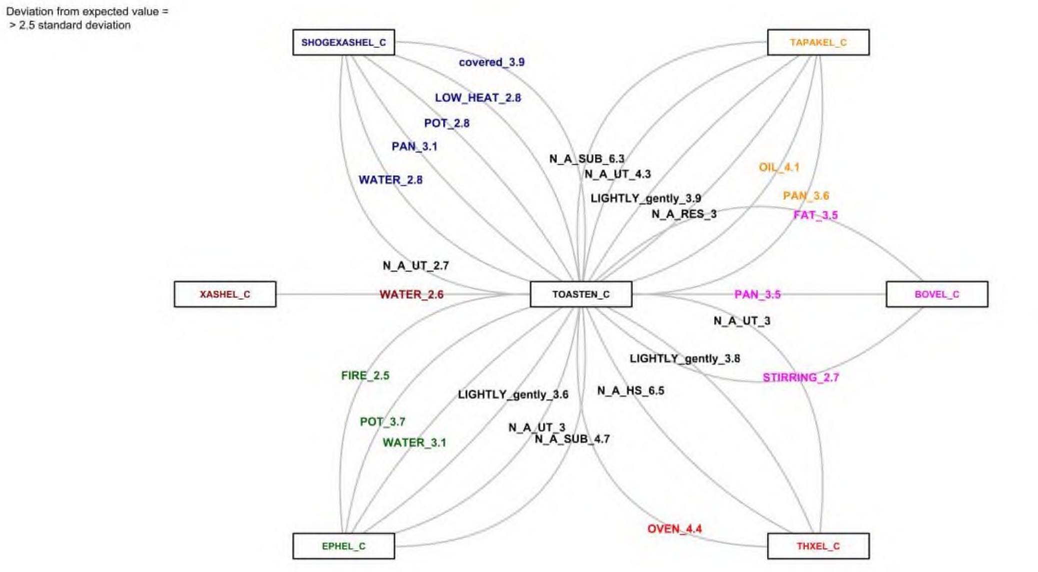


Fig. 2. Potential translation equivalents based on the *German-Armenian* language pair

The first noticeable outcome of the comparison of the German verb *toasten* to the Armenian counterparts is that only one edge connects *toasten* to *xashel* (arm.) (to boil*, to cook in water*) (Fig. 2). However, *toasten* and *xashel* might not be considered here as semantically closer verbs and potential translation equivalents due to their statistically significant semantic differences in the parameter *Substance*. In the rest of the annotation parameters for these two verbs, no value was verbalized, resulting in an insufficiency of data to compare them. The Armenian culinary verb *shogexashel* (to braise*/to sauté/to cook with little water*) differs from *toasten* in the parameters *Substance*, *Heat intensity*, *Utensil*, and *Manner*. In the pair *toasten-bovel* (to roast), the Armenian verb *bovel* differs from the German *toasten* in three parameters. In the distribution of the *Utensil* and *Substance* parameter values, the co-occurrences of {pan} and {fat} with *bovel* both exceed the expected value by 3.5 times the standard deviation (example 6). *Bovel* also correlates with the *Manner* parameter value {stirring}.

6. Հացի շերտերը թեթևակի *բովել* չոր *թավայի* մեջ: *(ARM)*
   [Lighly *toast* the bread slices in a dry pan].

However, since the values of at least the parameter *Resultative* (adjectives) are verbalized, it could be stated that there are no statistically significant differences between *bovel* and *toasten* here. Thus, these two verbs could potentially be translation equivalents, with some consideration regarding the semantic differences in cooking substance, utensil, and manner, which are also partially anchored in the definition of the Armenian verb *bovel*.[157] *Toasten,*[158] on the contrary, is defined by the parameter *Ingredient*, which was excluded from the analysis in this work due to objective reasons elaborated on in detail in Chapters 2.1 and 2.2 on German corpus analysis (examples 7–8).

7. Das *Brot toasten* und mit der Knoblauchzehe einreiben. *(detenten13)*
   [Toast the bread and rub with a garlic clove].

8. *Toasten* Sie die *Toastscheiben* nach Belieben und bestreichen Sie diese mit der Remoulade. *(detenten13)*
   [Toast the slices of toast bread* to your taste and coat them with remoulade sauce].

In considering the translation potential between *toasten* and *thxel* (to bake), the difference is observed in the parameters *Heat source* and *Manner*. The semantic difference in the *Heat source* parameter contrasts *toasten*, with no verbalized heat source, with *thxel* strongly correlating with the value {oven}. Besides, in comparing *toasten* and *thxel*, the co-occurrences of the *Manner* parameter value {lightly_gently}[159] with the verb *toasten* surpass the expected value by 3.8 times of the standard deviation. However, since most of the parameters of the verbs *toasten* and *thxel* have no verbalized values due to insufficiency of data, it is difficult to state whether there is any potential for equivalency between these verbs. In the pairs *toasten-tapakel* (to fry) as well as *toasten-ephel* (to boil*, to cook), the parameters *Substance* and *Utensil* differentiate the Armenian verbs *tapakel* and *ephel* from *toasten*, while the latter lacks

[157] "բովել"-bovel- 1. Տապակի մեջ **առանց ջրի կամ յուղի** աղանձել: 2. Յուղով խարկել, բոհրել: *[1. To roast (parch) in a **pan** without any water or fat. 2. To roast with **fat**, toast]*. *("բովել"*, in *Արդի հայերենի բացատրական բառարան* (1976) retrieved on 20.07.2022 via *nayiri.com*

[158] "toasten" 1. Brotschnitten, Weißbrotschnitten rösten ("toasten", in: *Wörterbuch der deutschen Gegenwartssprache (1964–1977)* retrieved on 20.07.2022 via *DWDS*, <https://www.dwds.de/wb/toasten>).
"toasten": (besonders von Weißbrot) in Scheiben rösten 1(a). (DUDEN Universal Wörterbuch online) <https://www.duden.de/rechtschreibung/kochen> retrieved on 16.11.2022

[159] The *Manner* parameter value {lightly_gently} was coined as a generalization of the values {lightly, gently} (en.) {leicht, kurz} (dt.), {թեթև, թեթևակի, մի փոքր, մի քիչ} (arm.) [light, lightly, a little, a bit] to cover them on a conceptual level as well as to allow the comparison of the verbs in the three languages without having to excessively reduce the annotated examples and/or risk incorrect abstractions. For a full list of generalizations, see Appendix 9.

verbalized data in the respective parameters. *Toasten*, in turn, is differentiated from *tapakel* and *ephel* again in the parameter *Manner* by correlating strongly with the value {lightly_gently} (examples 9–10).

The parameter *Manner* is probably the most distinctive feature when contrasting *toasten* to the Armenian verbs.

9. *Toasten* Sie den *Toast ganz leicht* an und bestreichen ihn mit etwas Margarine. *(detenten13)*
   [Toast the toast bread very lightly and rub it with some margarine].

10. Beginnen wir nun damit, das Brot *kurz* zu *toasten*, so dass es nur ganz leicht braun wird.
    [Now we start by shortly toasting the bread so that it becomes only light brown].

5.2.2 CCCG Visualization of Language Pairs: *Armenian-German*

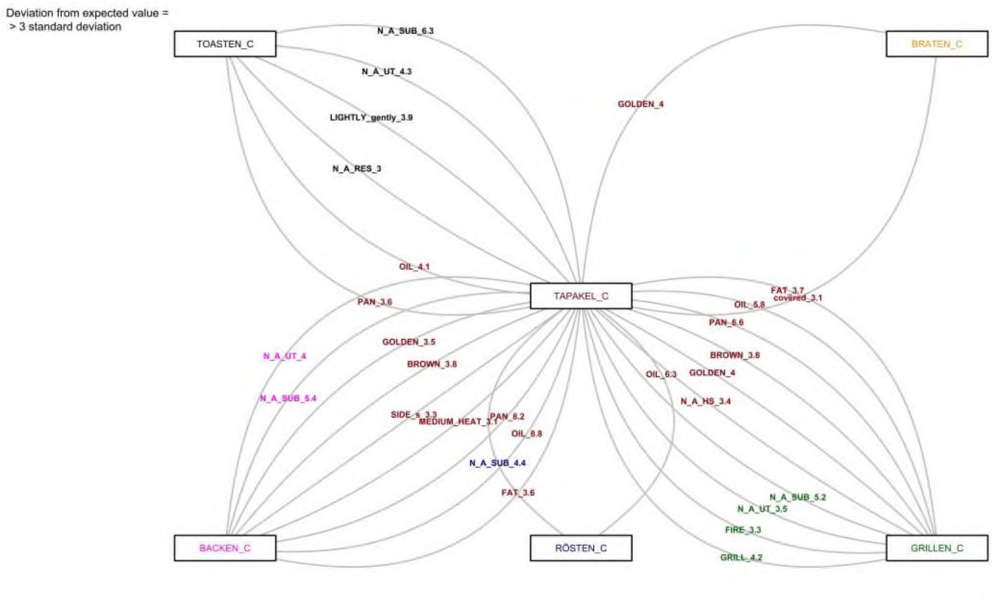


Fig. 3a. Potential translation equivalents based on the language pair *Armenian-German*

This *CCCG* (Fig. 3a) presents pairwise comparisons of one the Armenian hyperonym cooking verbs, *tapakel* (to fry), to the five German verbs on the conceptual level. Obviously, *tapakel* is semantically closer to the German verbs *braten* and *rösten*, judging from the fewer edges connecting them, and differs drastically from the other three verbs. In the pair *tapakel-braten*, statistically significant differences between these verbs are observed only in parameters *Manner* and *Resultative* (adjectives), viz. *tapakel* strongly correlates with {covered} (examples 11–12) and {golden} (examples 13–14).

11. Սմբուկը լցնել մսի վրա, *տապակել փակ կափարիչով*: *(ARM)*
    [Add the egg plants to the meat and *fry closed/with the lid on*].

12. Դդմիկը փոխանցել խորը թավայի մեջ և *տապակել* մոտ կես ժամ *փակված կափարիչի տակ*: *(ARM)*
    [Transfer the courgettes into a deep pan and fry them around half an hour closed/with the lid on].

13. Սոխը *տապակել* մինչև *ոսկեգույն* դառնալը: *(ARM)*
    [*Fry*/sauté* the onion until *golden*].

14. Հավի միսը կտրատել փոքր կտորներով, *տապակել* քիչ քանակությամբ բուսական յուղով, մինչև առաջանա թեթև *ոսկեգույն* կեղև: *(ARM)*
[Cut the chicken into small pieces, *fry* with little vegetable oil until it gets light *golden* crust].

The parameter *Substance* has a strong impact on the distribution of occurrences in the verb pair *tapakel-rösten.*[160] While the Armenian verb *tapakel* shows a strong correlation with the *Substance* parameter value {oil}, the German *rösten* strongly correlates with {N_A_SUB}.
In the pairs *tapekel-toasten* (to fry-to toast) and *tapakel-backen* (to fry-to bake), the differences in the parameters *Substance*, *Resultative* (adjective), and *Utensil* are partially generated by non-verbalized values. However, e.g. when comparing *tapakel* to *toasten*, the latter strongly correlates with the *Manner* parameter value {lightly_gently}. The co-occurrences of the *Utensil* and *Substance* parameter values {pan} and {oil} with *tapakel* are overrepresented respectively by 3.6 times and 4.1 the standard deviation. Statistically significant differences between the verbs *tapakel* and *backen* are observed in five of our six annotation parameters, except for *Heat source*.
In comparing the verb pair *tapakel-grillen*, the statistically relevant semantic differences are observed in the parameters *Utensil*, *Substance*, *Resultative* (adjectives), and *Heat source*. The latter is mostly observed to contain less verbalized values. However, in this pair, the German *grillen* differs from the Armenian *tapakel* because it strongly correlates with the *Heat source* parameter values, viz. {grill} and {fire} (examples 15–16).

15. Am Abend versammelten sich alle Kinder am *Lagerfeuer*, um dort Würstchen zu *grillen*. *(detenten13)*
[In the evening the children gathered around the *campfire* to *grill* sausages].

16. Auf dem heißen *Grill* die Koteletts auf jeder Seite rund drei Minuten *grillen*. *(detenten13)*
[*Grill* the cutlets on both sides roughly 3 minutes on hot *grill*].

[160] See *CCCG* excel-tables for translation purposes for the verb pairs in Appendix 11.

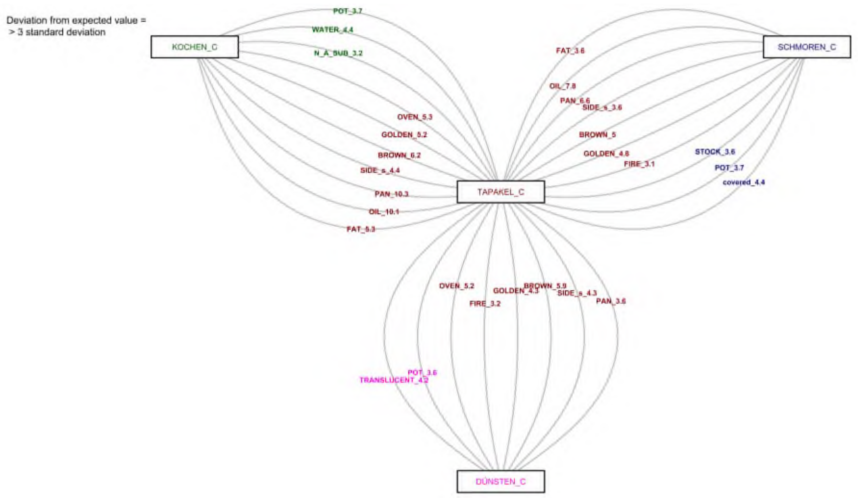


Fig. 3b. Potential translation equivalents based on the language pair *Armenian-German*

In the pairs *tapakel-dünsten* (to fry-to sauté), *tapakel-schmoren* (to fry-to braise) and *tapakel-kochen* (to fry-to cook*/to boil*), the difference between the verbs being compared is observed in almost all annotation parameters, except for *Heat intensity* (Fig. 3b). However, the parameter *Substance*, with its values denoting some type of *liquid*-like cooking substance, has the most decisive impact on differentiating *tapakel* from *dünsten*, *schmoren*, and *kochen*, hence excluding its potential to be considered as a translation equivalent for its counterparts, as *tapakel* exclusively correlates with *oil*-like cooking substances.[161] For instance, when comparing the verbs *tapakel* and *kochen*, the number of co-occurrences of the *Substance* parameter value {oil} with *tapakel* exceeds the expected value by 10.1 times the standard deviation under the hypothesis of a uniform independent distribution.

Thus, the parameters *Utensil*, *Substance*, as well as partially *Resultative* (adjectives), have the highest impact on comparing *tapakel* (arm.) to the German verbs. Their semantic closeness in these parameters with *braten* and *rösten* therefore allows them to be considered as potential translation equivalents.

[161] See also the *CA* of the German and Armenian verbs in Chapters 2.3 and 4.2 of this work.

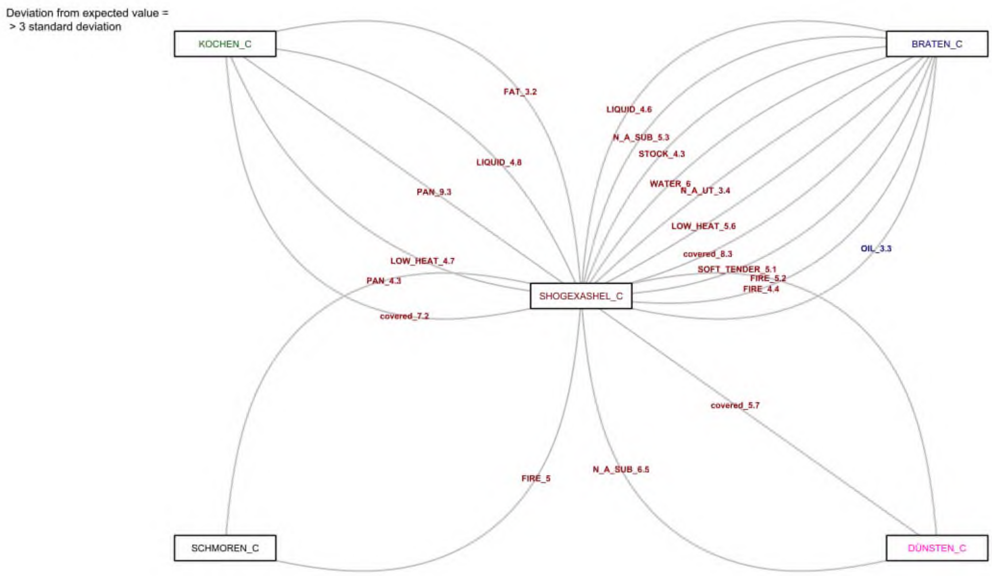


Fig. 4a. Potential translation equivalents based on the language pair *Armenian-German*

The *CCCG* (Fig. 4a) compares one of the Armenian specific (hyponym) verbs, *shogexashel* (to braise*/to steam*, to cook with little water*), to the verbs *dünsten*, *braten*, *schmoren*, and *kochen*. Since fewer connecting edges indicate possible semantic closeness between the verb pairs in comparison, a higher translation equivalence potential (on a conceptual level) is observed between the verbs *shogexashel-schmoren* and *shogexashel-dünsten*.

The Armenian cooking verb *shogexashel* differs from the German *schmoren* only in parameters *Heat source* and *Utensil*. For instance, the number of co-occurrences of the *Heat source* parameter value {fire} with *shogexashel* surpasses the expected value by the 5.0 times the standard deviation (example 17). Therefore, with all the semantic similarities in the rest of the parameters, *shogexashel* and *schmoren* have a high potential for translation equivalence.

17. Բանջարեղենը մանր կտրատել, շաքարավազ, աղ, քացախ և բուսայուղ ցանել վրան, խառնել: Դնել *կրակին* և շոգեխաշել 15 րոպե: *(ARM)*
    [Cut the vegetables in small pieces, sprinkle some sugar, salt, vinegar, and vegetable oil and mix. Put on *fire*/heat** and braise for 15 minutes].

The semantic difference between the verbs *shogexashel* and *dünsten* is observed in the parameters *Manner*, *Heat source*, and *Substance*. The issue with the *Heat source* parameter was already elaborated upon in comparing the verb pair *shogexashel-schmoren*. Moreover, the observed strong correlation between *shogexashel* and the *Substance* parameter value {N_A_SUB} is explained by the definition of the verb itself.[162] No substance is verbalized and therefore not annotated, as *shogexashel* presupposes շոգի *[shogi]* (steam, vapor) thus "little water" to generate the steam. S*hogexashel* strongly correlates with the *Manner* parameter value {covered}, with the number of co-occurrences between the two exceeding the expected value by 5.7 times the standard deviation. Considering the aforementioned difference, there is still some potential for them to be translation equivalents (examples 18).

18. Դդմիկի հյութը մզել, շերտիկները տապակել 2 րոպե, ապա *ծածկել կափարիչով*, *շոգեխաշել* 3-4 րոպե: *(ARM)*
    [Rinse the courgettes off and fry the slices for 2 minutes, then cover with the lid and braise for 3-4 minutes].

There are statistically no significant differences between the verb *shogexashel* and *kochen* in parameters *Resultative* (adjectives) and *Heat source*. In the rest of the annotation parameters, *kochen* differs semantically from *shogexashel*. However, in the sense of "to cook/to prepare," *kochen*[163] and *shogexashel* might have some translation potential.[164]

---

[162] "շոգեխաշել"-shogexashel- շոգիով խաշել, շոգու ջերմությամբ խաշել: *[to cook* with steam, to cook* with the help of the heat from steam]*. (Translation is mine). *("Շոգեխաշել"*, in *Արդի հայերենի բացատրական բառարան* (1976) retrieved on 27.07.2022 via *nayiri.com*

[163] In the *"Wörterbuch zur Verbvalenz"* *kochen* is defined as *„etwas befindet sich zusammen mit einer Flüssigkeit irgendwielange bei Siedetemperatur in einem Topf o.Ä, um gar zu werden*". This means that, if *kochen* is used in the sense of cooking in a substance, then it is a liquid and not fat. As soon as *fat* or some type of *oil*-like substance – with the possibility of being mixed with *liquid* – co-occurs, then the potential translation equivalent for the verb *shogexashel* would rather be *dünsten* and not *kochen*.
<*https://grammis.ids-mannheim.de/verbs/view/400712/3*> retrieved on 28.07.2022.
"kochen"- 1 a) (ein festes Nahrungsmittel) auf dem Herd, auf einer Feuerstelle o. Ä. durch Hitze in einer oder unter Zusatz einer Flüssigkeit gar werden lassen. (DUDEN Universal Wörterbuch online)
<*https://www.duden.de/rechtschreibung/kochen*> retrieved on 28.07.2022
"kochen"- etw. in reichlich Flüssigkeit auf Siedetemperatur erhitzen und dadurch ("*Kochen*", in: *Wörterbuch der deutschen Gegenwartssprache (1964–1977)* retrieved on 28.07.2022 via *DWDS*, <*https://www.dwds.de/wb/backen*>*).*

[164] In example 19, the *Substance* parameter value {Wasser} (water) is verbalized, while in the next examples, the verb *kochen* already presupposes {water} and would never be mistaken for other verbs, e.g. *braten*, as the latter strongly correlates with some type of *oil* or *fat*, thus excluding *water* (see Chapter 4.2 on CA vs. MCA of German verbs).

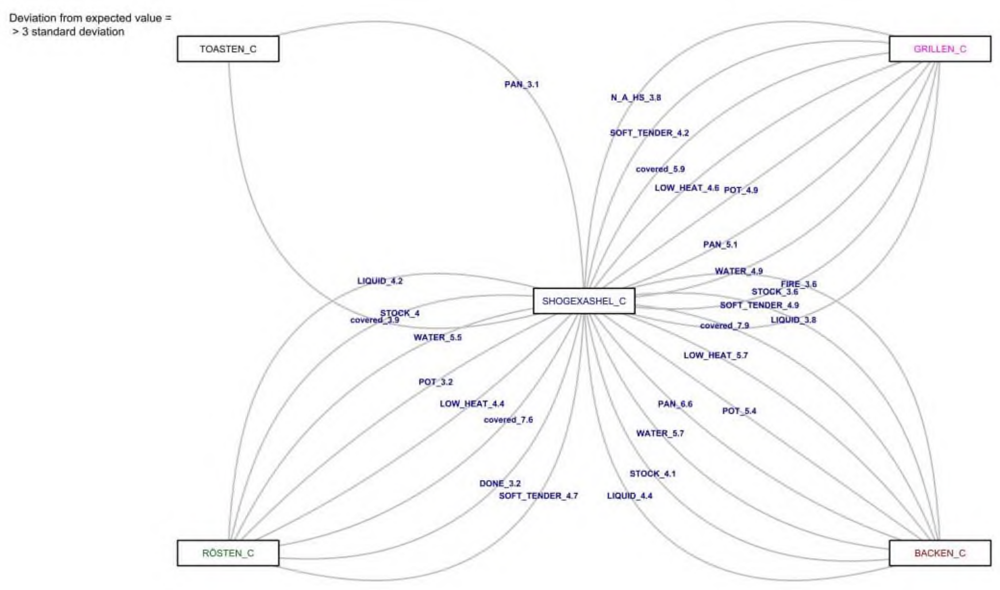


Fig. 4b. Potential translation equivalents based on the language pair *Armenian-German*

There is no potential for the verb pairs *shogexashel-backen*, *shogexashel-grillen*, and *shogexashel-rösten*, as they differ in all annotation parameters (Fig. 4b, cf. examples 19–20). Even though the following graph illustrates statistically significant differences between the verb *shogexashel* and *toasten* only with respect to the parameters *Utensil* and *Manner*, they cannot be considered as potential translation equivalents for two reasons. First, there are also statistically significant differences between *shogexashel* and *toasten* in the parameters *Heat intensity*, *Utensil*, and *Substance* when the *CCCG* is generated for deviations above 2.5 (see Chapter 5.2.1, Fig. 2). Second, in the parameters *Resultative* (adjectives) and *Heat source*, almost no value is verbalized for both *shogexashel* and *toasten*, making it difficult to draw conclusions about the semantic closeness of these verbs.

19. *Wasser* in einem Topf zum Kochen bringen und die Eier darin ca. 6 Minuten *kochen* und abschrecken. *(detenten13)*
    [In a pot bring the water to boiling and hard-boil the eggs in it for 6 minutes and hold under cold water].

20. Nebenbei einen großen Topf aufsetzen, um die Eier hart zu *kochen*. *(detenten13)*
    [Put a large pot near it to hard-boil the eggs].

5.2.3 CCCG visualization of language pairs: *Armenian-English*

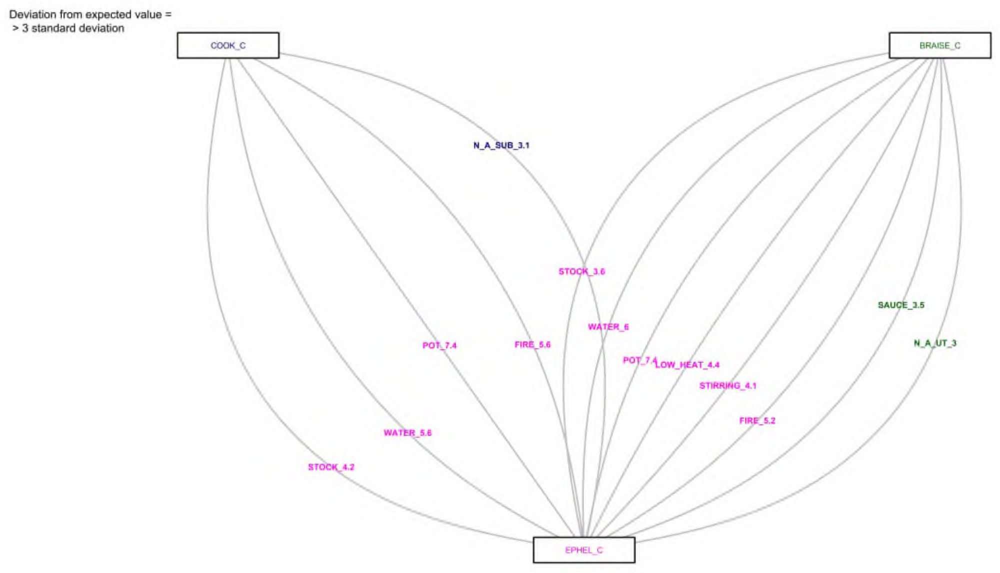


Fig. 5a. Potential translation equivalents based on the language pair *Armenian-English*

Fig. 5a compares the Armenian general verb (hyperonym) *ephel* (to boil, to cook) to the English verbs *cook* and *braise*. The comparison of *ephel* to the English verbs *sauté, fry, roast,* and *bake* is discussed separately below (Fig. 5b) to ensure the legibility of the *CCCG*. Since the number of edges connecting the verb pairs in comparison denotes the semantic differences in the identified annotation parameters, almost all English verbs differ remarkably from the Armenian *ephel*, except for *cook*.

In comparing the Armenian-English verb pair *ephel-cook*, *ephel* differs from *cook* in the parameters *Substance, Utensil,* and *Heat source*. The strong correlation of the Armenian verb *ephel* with the *Heat source* parameter value {fire} (see also Fig. 5b)[165] is due to the semantic density of {կրակ} (fire), which indicates any heat source that food can be cooked on. The semantic difference between *ephel* and *cook* in the parameter *Substance* and *Utensil* is explained by the definitions and classifications of the verbs themselves.[166] *Cook*[167] is an umbrella verb for all other hyponyms in English, from *fry* to *bake,* with the possibility of all

[165] See also the comparison of the Armenian-German verbs based on the example of *shogexashel* in Chapter 5.2.2 of this work (Fig.4a and Fig. 4b).
[166] cf. Lehrer on the classification of the English verbs (1972).
[167] "cook"- 1. to combine (a number of ingredients) to make a dish out of them in a way that involves heating a) (of food) to complete the process of being heated, so as to become fit for human consumption, b) to heat (food) to the point of it becoming edible. ("Cook" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*.

respective parameters being verbalized in the co-text, viz. *to cook in the oven* and not necessarily *to bake in the oven* (example 21).

21. <u>Boil the water</u> to *cook* the rice in a large pan or pot. (*enTenTen15*)

This is not the case with *ephel*. Even though *ephel* is a more general verb, a hypernym for *shogexashel* (to braise*/to steam*/to cook with littel water*) and *xashel* (to boil, to cook in water*), it cannot be substituted for all other Armenian verbs. While *ephel*[168] has the meaning of *cook* in its definition, denoting *"to prepare"*, it primarily presupposes some type of *liquid* substance and/or the action of boiling (examples 22–23).

22. Կարտոֆիլը կաթսայի մեջ աղ արած *ջրով* մի քանի րոպե *եփել*, որպեսզի կտորները դառնան փափուկ, բայց չփլվեն: *(ARM)*
[*Cook* the potatoes for a couple of minutes in boiling salted water so that the pieces become soft but do not fall apart].

23. *Եփել* բարձր ջերմության վրա, մինչեւ *ջուրը եռա*: *(ARM)*
[*Cook* over high heat until the *water* starts to *boil*].

This is also illustrated by the distribution of the *Substance* parameter values in comparing *ephel* and *cook*, where the number of annotated examples of {water} and {stock} with *ephel* exceeds the expected value by respectively 5.6 and 4.2 times the standard deviation. Thus, *ephel* and *cook* can only be semantically closer verbs in the meaning of "to make, to prepare" a dish and not an ingredient (examples 24–27).

24. *Եփել* <u>սոուսը</u> ցածր կրակի վրա: *(ARM)*
[*Cook* the <u>sauce</u> of low heat].

25. <u>Բրնձով շիլա</u> *եփել*: *(ARM)*
[*Cook* rice porridge].

26. I asked him if he'd teach me how to *cook* a <u>dish</u> for Immigrant Kitchens. (*enTenTen15*)

[168] "եփել"-ephel- 1. Ուտելիքը կամ խմելիքը եռացնելով պատրաստել, տաքացնելով անհրաժեշտ վիճակին հասցնել: 2. Կրակով՝ ջերմությամբ եփ տալ, խաշել, 3. Եռացնել:
[1. To prepare food or drink by boiling, to bring to necessary state by warming. 2. Boil with the help of fire (heat), xashel. 3. Bring to boil.]. (Translation is mine). *("Եփել"*, in *Արդի հայերենի բացատրական բառարան* (1976) retrieved on 31.07.2022 via *nayiri.com*

27. When <u>the soup</u> is *cooked*, puree with a stick blender. (*enTenTen15*)

The semantic similarity (closeness) of the verbs *ephel* and *cook* in the remaining parameters, viz. *Heat intensity*, *Manner*, and *Resultative* (adjectives), is primarily an outcome of data insufficiency, with the overwhelming majority of the examples for both verbs were annotated as N_As.[169] There is little translation potential in the verb pair *ephel-braise* as they differ in five of the six annotation parameters, except for *Resultative* (adjectives). Besides, *braise* strongly correlates with the *Substance* parameter value {sauce}, which could be classified as *liquid*, thus opening up space for semantic closeness between these two verbs. In fact, *braise*[170] is a two-stage process and comprises in itself the meaning of first *frying* lightly and then *cooking* (ephel).*

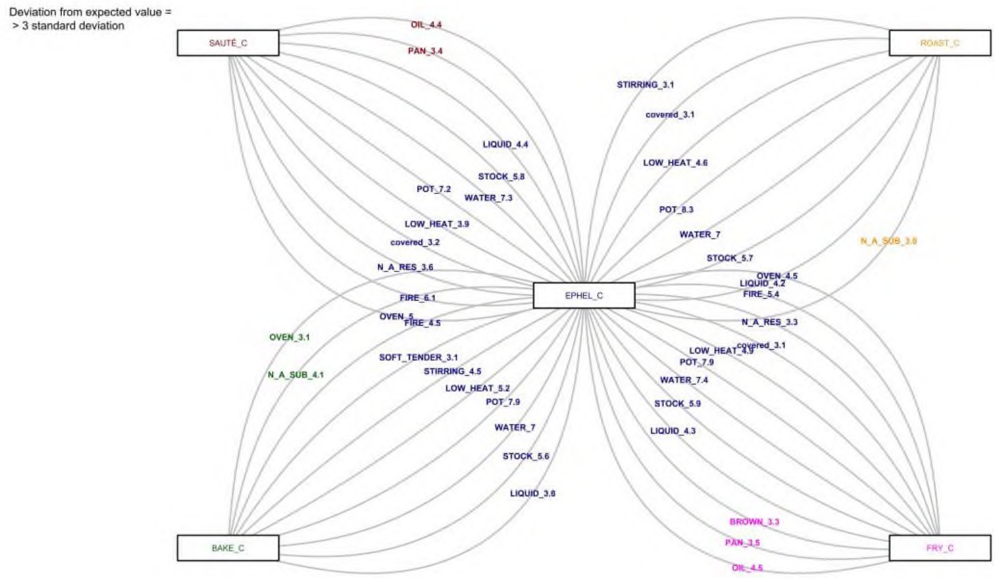


Fig. 5b. Potential translation equivalents based on the language pair *Armenian-English*

There is hardly any translation potential between the components of the verb pairs *ephel-fry*, *ephel-sauté*, *ephel-roast*, and *ephel-bake* as they differ in all six annotation parameters. However, their translation possibility might already be excluded due to the differences in the *Substance* parameter alone, as some type of *liquid*-like substance that correlate with *ephel*

---

[169] For a complete list of all German, English, and Armenian annotated verbs see Appendices 1, 5 and 12.

[170] "braise"- to lightly fry (food) and then simmer it in a sauce in a closed container. ("Braise" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*

explicitly excludes *oil* or *fat*-like cooking substances, which are strongly related to *fry*. *Roast* and *bake* create strong correlations with *no cooking substance*, annotated here as {N_A_SUB}.[171] The definitions of *roast*[172] and *bake*[173] also allow us to state that the most decisive parameter in differentiating *ephel* from the English verbs is *Substance*, as *ephel* even correlates with the *Heat source* parameter {ջեռոց} (oven), which is not verbalized in the annotated examples with *roast* and *bake*.

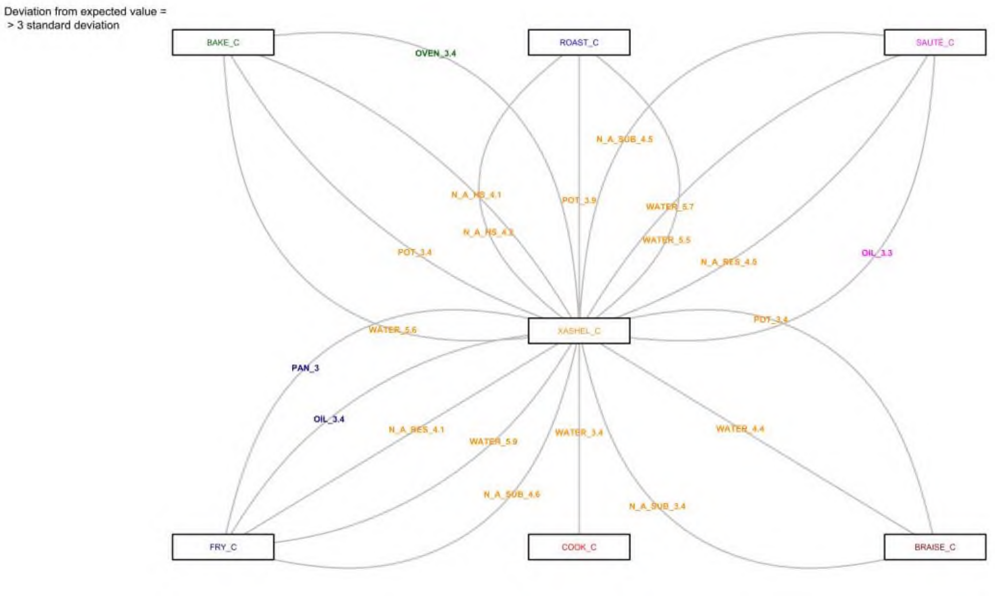


Fig. 6a. Potential translation equivalents based on the language pair *Armenian-English*

When comparing the Armenian specific (hyponym) verb *xashel* (to boil*, to cook in water) to the English culinary verbs, the first noticeable outcome visualized by the *CCCG* (Fig. 6a and

---

[171] At the beginning of this chapter, it has already been mentioned that, due to our initial annotation criteria, both *non-verbalized cooking substances* and *the absence of them* has been annotated as {N_A_SUB}, which led to slight misinterpretations of the *CCCGs* generated by *R*. However, other statistically evaluative visualizations, e.g. *Correspondence Analysis*, as well as specific examples from the corpora, help differentiate them.

[172] "roast"- culinary- a) to cook (meat, fish, chicken, etc.) in an oven or over a fire, b) to heat (coffee beans, nuts, etc.) in order to dry them ("Roast" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*

"roast"-v. **a.** *transitive*. To cook (food, esp. meat) by prolonged exposure to heat at or before a fire or similar source of radiant heat. Also *intransitive*. **b.** *intransitive*. Of food: to be cooked by roasting. **c.** *transitive*. To expose (coffee beans) to heat in order to prepare for grinding. ("roast" in *OED Online* This Edition (2020), retrieved on 23.09.2022 via < https://www.oed.com/view/Entry/166571?rskey=QMqje7&result=3#eid >)

[173] *"bake"- transitive*. To cook (food) by dry heat, without direct exposure to flame, typically in an oven, or sometimes on a heated surface such as a griddle, baking stone, etc. ("bake" in *OED Online* This Edition (2020), retrieved on 01.08.2020 via < https://www.oed.com/view/Entry/14759?rskey=6PgYyw&result=3#eid >)

"bake"- to cook something by heating it in a closed space ("Bake" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*

Fig. 6b) is that *xashel* significantly differs from *fry* and is semantically closest to *cook*. They differ only in the *Substance* parameter, with *xashel* strongly correlating with {water}. However, the actual differences can be revealed when the deviation of distribution is lowered to 2.5 times the standard deviation, which is still statistically significant (Fig. 6b).

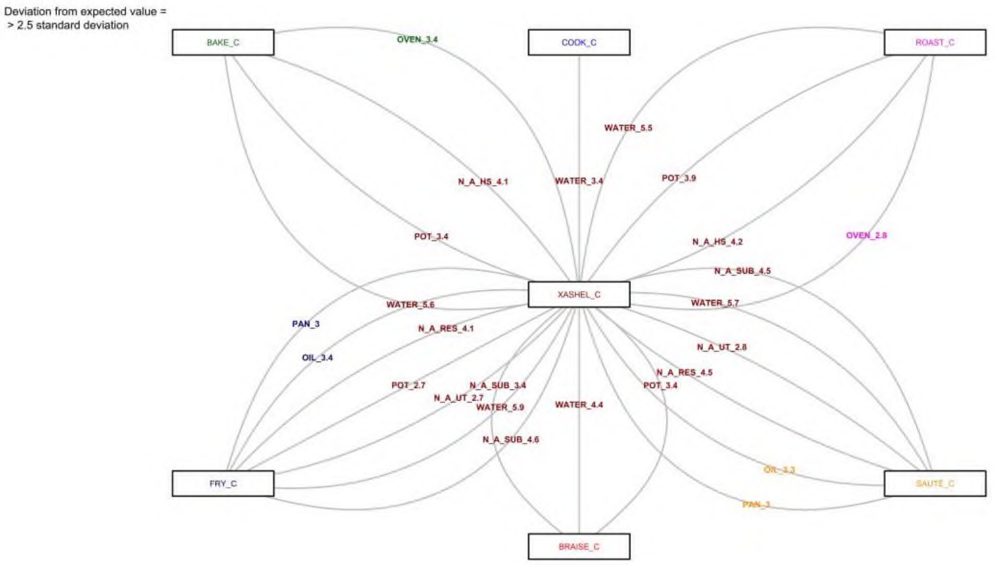


Fig. 6b. Potential translation equivalents based on the language pair *Armenian-English*

Even though there are no considerable differences between the first and the second *CCCG* visualizations in terms of reducing the non-verbalized parameter values – since the deviation for most of the parameters remained unaltered (except for the parameter *Utensil*) – some additional differences can be observed in a few parameters and/or their separate values. For instance, in the pair *xashel-roast*, the difference in the distribution of occurrences is only observed in the parameters *Substance*, *Utensil*, and *Heat source*. Thus, the English verb *roast* strongly correlates with the *Heat source* parameter value {oven}. In the pairs *xashel-braise*, *xashel-bake*, and *xashel-cook*, the distribution remains unaltered. In the first pair, the difference is observed in the parameters *Utensil* and *Substance*. In the second pair, *xashel* differs from *bake* in the parameters *Substance*, *Utensil*, and *Heat source*. The co-occurrences of the *Heat source* value {oven} with *bake* are overrepresented by the 3.4 times the standard deviation. The number of the co-occurrences of the *Substance* parameter value {water} and *xashel* surpasses the expected value by 5.6 times the standard deviation. Besides, the verb *xashel* strongly correlates with the *Utensil* parameter value {pot}. As for the last remaining pair, with no differences in the distribution of relevant parameters values at both above 3.0 and 2.5 times the standard deviation, *xashel* is in both cases semantically closest to *cook*.

In the next two pairs, additional differences between the verbs under comparison are illustrated, although these include partially non-verbalized values. For instance, if in the first *Conditional Correlation Graph* (Fig. 6a), the distributional difference between the verbs *xashel* and *fry* is influenced by the parameters *Substance*, *Utensil*, and *Resultative* (adjectives). In the second graph (Fig. 6b), two additional values of the *Utensil* parameter, viz. {pot} and {N_A_UT}, also differentiate these two verbs by showing strong correlation with the verb *xashel*. However, it can be stated that in both pairs, *xashel*-*fry* and *xashel*-*sauté*, the most decisive differentiating parameters based on their verbalized values are *Substance* and *Utensil*. While *fry* and *sauté* strongly correlate with {oil} and {pan}, *xashel* prefers {water}.[174] Thus, *Substance* and *Utensil* remain the key parameters distinguishing *xashel* from the English culinary verbs.

[174] As mentioned above, *xashel* also shows a strong correlation with the *Utensil* parameter value {pot} (Fig. 2a) when the *CCCG* is calculated using 2.5 times the square root of the expected value.

5.2.4 CCCG Visualization of Language Pairs: *English-Armenian*

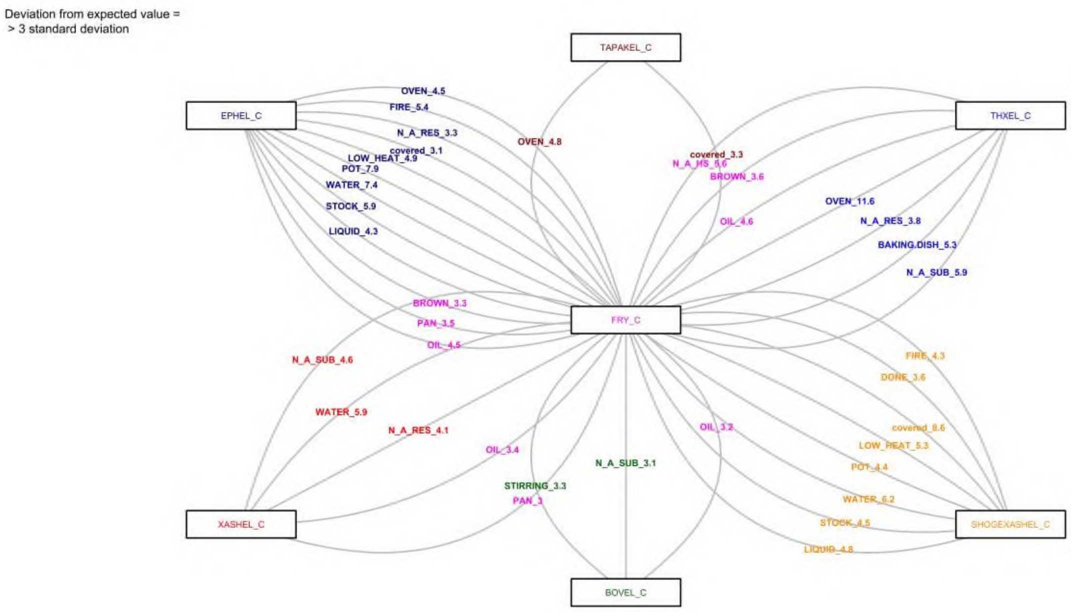


Fig. 7. Potential translation equivalents based on the language pair *English-Armenian*

This *CCCG* (Fig. 7) compares the English generic (hypernym) verb *fry* to six Armenian counterparts to find potential translation equivalents on the conceptual level. A quick look shows that the Armenian verbs *tapakel* (to fry) and *bovel* (to roast) are the closest semantic matches to *fry*. The latter differs from *tapkel* only in the parameters *Heat source* and *Manner* (examples 28–29).

28. Նախապես 180 աստիճանով տաքացված *ջեռոցում* դնել ձևամանը եւ *տապակել* կարտոֆիլը մինչև բաց շագանակագույն երանգ ստանալը: *(ARM)*
[Place the baking from in the preheated oven and fry*/roast the potatoes until they become light brown].

29. Սմբուկը լցնել մսի վրա, *տապակել փակ կափարիչով*: *(ARM)* (Fat or any type of oil is assumed by the definition of *tapakel*).
[Add the eggplants to the meat and *fry closed/with the lid on*].

Thus, *tapakel*[175] and *fry*[176] have high potential to be considered translation equivalents since they do not differ – most importantly – in the parameter *Substance*, which is also anchored in their encyclopedic definitions (examples 30–31). Besides that, they are semantically close in the parameters *Utensil*, *Resultative* (adjective), and *Heat intensity*.

30. *Ձիթապտղի յուղի* մեջ սմբուկը *տապակել* երկու կողմից մինչեւ ոսկե դարչնագույնը: *(ARM)*
    [Fry the eggplants on both sides in olive oil until they become golden brown].

31. Meanwhile *fry* the bacon in *olive oil* until crispy. (*enTenTen15*)

In the verb pair *fry-bovel*, statistically significant differences are observed in the parameters *Substance* and *Manner*. While *fry* strongly correlates with {oil}, *bovel*, on the contrary, with {N_A_SUB}, which – a thorough look in the examples and the verb's definition[177] reveals – indicates *no cooking substance* (examples 32-33).

32. Քունջութը *բովել առանց յուղի*, սառեցնել: *(ARM)*
    [*Roast* the sesame seeds *without any fat*, let it cool down].

33. 5-7 րոպե չոր տապակի վրա բովել ընկույզը, որ մի փոքր շագանակագույն դառնան (միջին ջերմության վրա, *առանց յուղի*): *(ARM)*
    [*Roast* the walnuts in *a dry* pan* until they become a bit (light) brown (on medium heat *without fat*)].

This is a very distinctive feature: when any type of *oil*-like substance is used, *tapakel* (to fry) is semantically the most appropriate translation equivalent of the English *fry*. In the *absence of a cooking substance, bovel* (to roast) is the only Armenian appropriate verb, even though *bovel* also allows for the presence of *fat* according to its secondary meaning, particularly with certain ingredients such as *ալյուր* (flour) (example 34). Moreover, *bovel* strongly correlates with the

[175] "տապակել"- tapakel- 1. Յուղի մեջ <u>առանց ջրի</u> եփել: 2. Բովել, աղանձել: 3. Տե՛ս խորովել: [1. To cook (food) in fat without water. 2. To roast, to parch. 3. See *grill*.] *("Տապակել"*, in *Արդի հայերենի բացատրական բառարան* (1976) retrieved on 09.08.2022 via *nayiri.com*

[176] "fry"- to cook (something) in (a container) without water over (a heat source), usually while partially or completely covered in oil or fat. ("Fry" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*

[177] "բովել"-bovel- 1. Տապակի մեջ **<u>առանց ջրի կամ յուղի</u>** աղանձել: 2. Յուղով խարկել, բոհրել: *[1. To roast (parch) in a **<u>pan</u>** without any water or fat. 2. To roast with **<u>fat</u>**, toast.]* (Translation is mine). *("բովել"*, in *Արդի հայերենի բացատրական բառարան* (1976) retrieved on 20.07.2022 via *nayiri.com*,

Manner parameter value {stirring}, which conceptually differentiates it from *fry*. Presumably, *bovel* without oil needs *constant stirring* (example 35-36).

35. Թավայի մեջ *կարագով բովել ալյուրը*: *(ARM)*
    [In a pan *roast* the *flour* with *butter*].

36. Սուրճի հատիկները չուգունե թավայի մեջ *բովել* մարմանդ կրակի վրա, *անընդհատ խառնել* փայտե գդալով: *(ARM)*
    [In a cast iron pan *roast* the coffee beans over low heat *by constantly stirring* it].

There is practically no translation potential in the pairs *fry-ephel*, *fry-shogexashel*, *fry-xashel*, and *fry-thxel* since they differ in almost all six annotation parameters. Most importantly, the difference in the *Substance* parameter alone is sufficient to exclude their translation potential. *Fry* strongly correlates with *oil*-like substances in the first three pairs, while *shogexashel* and *xashel*, both hyponyms of *ephel*, correlate with *water*-like ones. *Thxel*, on the other hand, does not presuppose *any substance* at all. In the verb pair *fry-thxel*, however, the most differentiating parameter is *Heat source*. The co-occurrences of {oven} with *thxel* are overrepresented by 11.6 times the standard deviation, under the hypothesis of a uniform independent distribution.

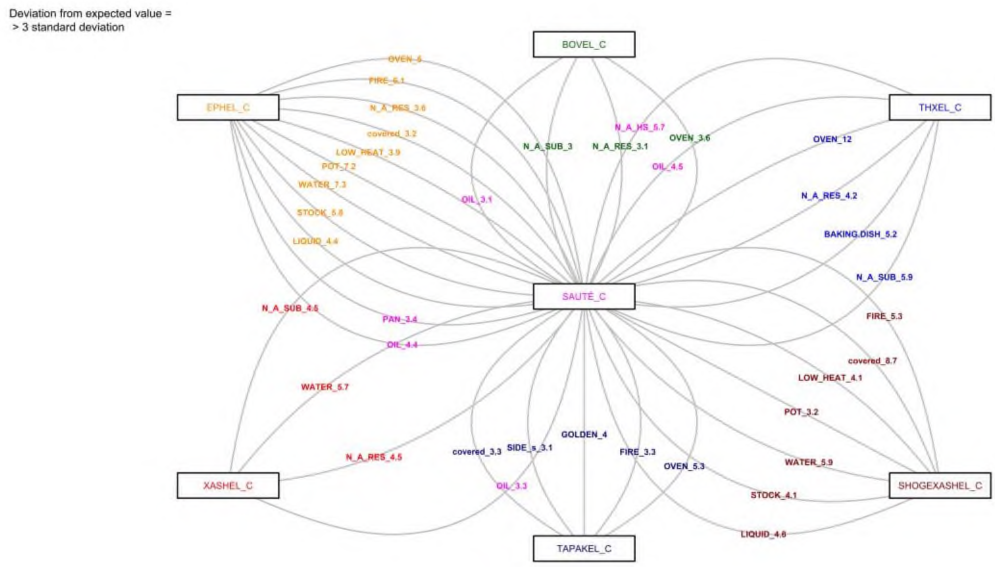


Fig. 8. Potential translation equivalents based on the language pair *English-Armenian*

Based solely on the *CCCG* (Fig. 8), which compares the English specific (hyponym) verb *sauté* to six Armenian verbs, the pairs *sauté-bovel*, *sauté-xashel* might initially appear to have considerable translation potential. However, in the pair *sauté-bovel*, the potential of considering *sauté* and *xashel* (to boil*, to cook in water*) as semantically closer verbs represented in a few connecting edges is mostly generated due to data insufficiency in parameters *Manner*, *Resultative* (adjectives), and *Heat source* for both verbs. As discussed earlier in this chapter, such cases may lead to misinterpretation of these verbs as semantically closer than they truly are. Crucially, they differ in the most distinguishing parameter: *Substance*. Even though *xashel* strongly correlates with non-verbalized substance values too, the number of annotated examples of {water} co-occurring with *xashel*,[178] exceeds those with *sauté* by 5.7 times the standard deviation, under the hypothesis of a uniform independent distribution. The verb *sauté*, on the contrary, shows a strong correlation with {oil}. Therefore, there is no potential for *sauté* and *xashel* to be considered as possible translation equivalents. In the verb pair *sauté-xashel*, *sauté* differs from *xashel* in the parameters *Substance*, *Heat source,* and *Resultative* (adjectives). However, the difference in the latter parameter is difficult to evaluate since a thorough look at the annotated examples shows that, in this case, the insufficiency of data is due to the non-verbalized values. The difference in the *Substance* parameter, by contrast, is rather clear-cut; *bovel* (to roast) shows a strong correlation with *no substance* at all, whereas *sauté* with {oil} (examples 37–39).[179]

37. *Sauté* the shallot in *canola oil* over medium high heat until translucent. (*enTenTen15*)

38. Heat the *oil* in a small skillet and *sauté* the shallot or onion for about one minute. (*enTenTen15*)

39. *Sauté* the shallots in 1 tablespoon butter in a small saucepan until wilted. *(EN_REZ)*

Last but not least, in the distribution of *Heat source* parameter values for the verbs *sauté* and *bovel*, the number of annotated examples of {oven} with *bovel* exceeds those with *sauté* by 3.6 times the standard deviation. However, the deviation from the expected value is primarily due

[178] *"խաշել"*-xashel- 1. Եռացրած ջրի մեջ կարճ ժամանակ եփել, կիսաեփ անել: 2. Ջրի մեջ եռացնելով եփել: 3. Եռացրած ջրի ազդեցությանը ենթարկել՝ կակղացնելու համար: *[1. To cook briefly in boiling water, to parboil. 2. To cook in water by boiling. 3. To expose to boiling water in order to soften.] ("Խաշել"*, in *Արդի հայերենի բացատրական բառարան* (1976) retrieved on 11.08.2022 via *nayiri.com*

[179] See also Fig. 1 of this subchapter comparing the English verb *fry* to the six Armenian verbs for further clarification on the annotation of the occurrences with *bovel*.

to the non-verbalized values in the *Heat source* parameter for the verb *sauté*, rather than a particularly high number of co-occurrences of *bovel* and {oven}. In fact, all 286 annotated examples of *sauté* contain exclusively *non-verbalized values* in the aforementioned parameter. By contrast, only six out of 89 annotated examples with *bovel*, six co-occur with {oven}. Nevertheless, it is once again the parameter *Ingredient* (with values like {nuts} or {coffee beans}) that differentiates *bovel* as a distinct type of cooking process from its semantically closer counterparts, e.g. *tapakel* (40-41).

40. *Գետնանուշը բովել ջեռոցում*: Կեղեւը հեռացնել եւ թողնել, որ հովանա: *(ARM)*
[*Roast* the *almonds* in *oven*. Peel and let it cool down].

41. Մաքրած *նուշը* լցնել ջեռոցի սապակի մեջ, որի վրա յուղաթուղթ է փռված: *Բովել*, մինչեւ ոսկեգույն դառնան: *(ARM)*
[Place the peeled *almonds* on the baking sheet (on the parchment paper) in *oven*. Roast until golden].

There is some translation potential for the pair *sauté-tapakel* as they are semantically closer in the parameter *Substance*;[180] however, they differ notably in the parameters *Manner* and *Heat source*, as well as *Resultative* (adjectives). The etymology of the verb *sauté*,[181] which is in fact the past participle of the French verb ‘sauter’, meaning ‘to leap’, ‘to jump’, suggests that the verb describes the movement of the ingredients in a pan by tossing and turning quickly over high heat. In this respect, *Manner* is the parameter that most clearly differentiates it from *tapakel* (to fry). The latter, on the contrary, strongly correlates with the *Manner* parameter values {covered} and {side_s} (examples 42–43).[182]

[180] Also in the parameter *Heat intensity*, however, it is not the most distinguishing parameter for the verb pair *sauté-tapakel*.

[181] “sauté”- to fry (something) for a small time. (“*Sauté*” in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*
“sauté”- *cookery* A. adj (Sometimes as past participle.) Of meat, vegetables, etc. Fried in a pan with a little butter over a high heat, while being tossed from time to time. B. n dish cooked in the above manner ("sauté" in *OED Online* This Edition (2020), retrieved on 11.08.2020 via <https://www.oed.com/view/Entry/14759?rskey=6PgYyw&result=3#eid >)

[181] Even though *sauté* is presented by Lehrer as well as by the antidote 9 dictionary as a verb, the OED has entries for *sauté* as an adjective (described as the technique of ‘sautéing’) and a noun (the dish itself). The OED also cites in its examples ‘sauter’ as a verb and quotes Gouffé (1869) Royal Cookery Bk. i. 5 “*To sauter* is to fry with little butter over a brisk fire” ("sauter, v." *OED Online*. This Edition (2020), retrieved on 11.08.2020 (<https://www.oed.com/view/Entry/171422?redirectedFrom=sauter#eid >)

[182] The *Manner* parameter value {side_s} was generalized and includes values such as {on both sides}, {on one side}, {on one side}, {on the other side} (EN), {երկու կողմից} (ARM), {auf jeder Seite}, {von beiden Seiten}, {beideseitig}, {rundherum} (DE). For a full list of annotation parameter neutralizations (generalizations) for translation purposes, see Appendix 9.

42. Սմբուկը լցնել մսի վրա, *տապակել փակ կափարիչով*: *(ARM)*
[Add the egg plants to the meat and *fry closed/with the lid on*].
43. Թավան տաքացնել, բուսայուղ լցնել, պանրաբլիթները *տապակել երկու կողմից* 3-5 րոպե մինչեւ *ոսկեգույն* դառնան: *(ARM)*
[Heat the oil in pan, fry the cheese pancakes *on both sides* for 3-5 minutes until *golden*].

In the distribution of the *Heat source* parameter values for the verbs *sauté* and *tapakel*, the number of occurrences of {oven} and {fire} with *tapakel* exceeds those with *sauté* by respectively 5.3 and 3.3 times the standard deviation (examples 43-44).

44. Այս բաղադրատոմսով առաջարկում ենք *ջեռոցում* կարտոֆիլը հավի հետ *տապակել*: *(ARM)*
[This recipe suggests *frying*/roasting* the potatoes with chicken in *oven*].

45. *Կրակին* դնել թավան եւ *տապակել* բանջարեղենը արեւածաղկի յուղի մեջ մինչեւ կիսով չափ պատրաստ լինի: *(ARM)*
[Put the pan on *fire*(on any heat source)* and *fry* the vegetables in sunflower oil until half ready/al dente].

The difference in the *Resultative* (adjectives) parameter between the verbs *sauté* and *tapakel* is manifested by a strong correlation between the value {golden} and *tapakel* (see example 42 above).

There is no translation potential in the verb pairs *sauté-ephel, sauté-shogexashel*, *sauté-thxel*, due to the difference in almost all parameters, and, most importantly, in the parameters *Manner*, *Substance*, and *Heat source.*

5.2.5 CCCG Visualization of Language Pairs: *German-English*

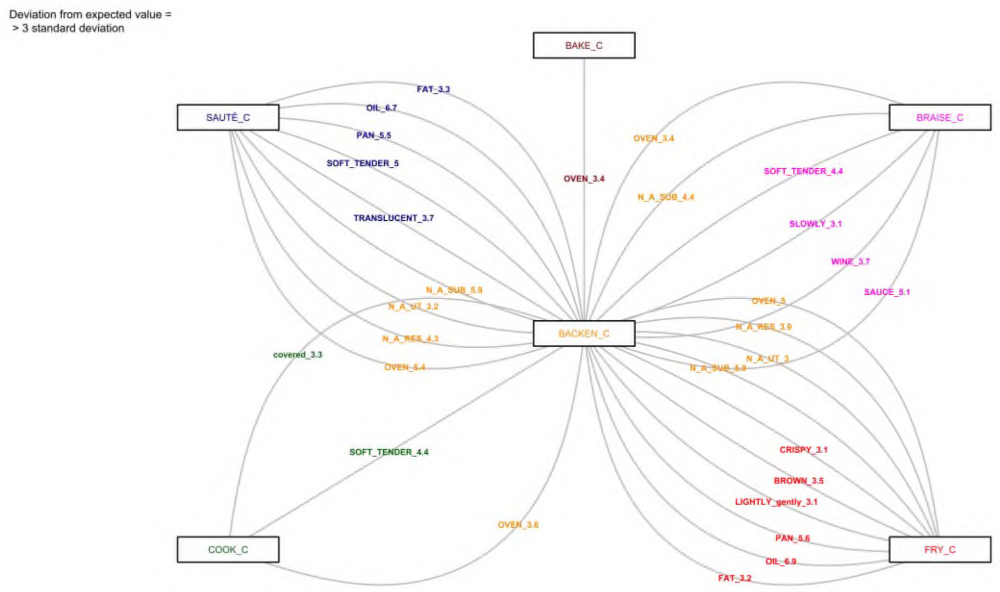


Fig. 9. Potential translation equivalents based on the language pair *German-English*

In comparing the German general (hypernym) verb *backen* to the six English verbs (Fig. 9), the most noticeable outcome is the absence of the verb *roast*, which indicates that there are no statistically significant differences between these two verbs in all six annotation parameters. However, their "semantic closeness" is largely due to insufficient data, i.e. *non-verbalized values* in the parameters *Substance*, *Heat intensity*, *Manner*, and *Utensil*. *Backen* and *roast*[183] might nevertheless be semantically close in the parameters *Resultative* (adjectives) and *Heat source*, as both have a considerably high number of annotated examples with verbalized values. The lack of data due to non-verbalized values is also the reason why it is difficult to draw any firm conclusions about the translation potential between *backen* and *roast*. It is difficult to state about the translation potential between *backen* and *roast*.[184] There is a huge translation potential between the verbs *backen* and *bake*, as they differ only in the parameter *Heat source*, where the

[183] "roast"- culinary- a) to cook (meat, fish, chicken, etc.) in an oven or over a fire, b) to heat (coffee beans, nuts, etc.) in order to dry them ("Roast" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*
"roast"-v. **a.** *transitive*. To cook (food, esp. meat) by prolonged exposure to heat at or before a fire or similar source of radiant heat. Also *intransitive*. **b.** *intransitive*. Of food: to be cooked by roasting. **c.** *transitive*. To expose (coffee beans) to heat in order to prepare for grinding. ("roast" in *OED Online* This Edition (2020), retrieved on 23.09.2022 via < https://www.oed.com/view/Entry/166571?rskey=QMqje7&result=3#eid >)
[184] Lehrer classified *roast* as a hyponym verb, but did not provide a precise hypernym for it. *Roast* is mostly described as a hyponym of *bake*, but having ties with *broil* and with such co-hyponyms as *barbecue*, *charcoal*, and *grill* (cf. 1972:157–58).

value {oven} strongly correlates with the English verb *bake*. However, as elaborated upon in detail at the beginning of Chapter 5.2.1 (Fig.1), the German *backen* implicates some type of {oven} (*Ofen*, *Steinofen*, etc.). Despite the fact that the English *bake* also mentions the presence of some type of an *oven* in its definition, a number of annotated examples contain an explicitly verbalized value {oven}, which leads to a considerably higher deviation value in the distribution (examples 46–47).

46. The six-pound loaves *are baked* in *wood-fired ovens* for fresh bread every day. (*enTenTen15*)

47. *Bake* in a pre-heated 350degree *oven* 40 minutes, then reduce heat to 300 degrees and *bake* 35 to 40 minutes more, or until cake tests done. *(EN_REZ)*

There might be some translation potential in the pair *backen-cook* (to bake-to cook), as they differ in three of the six annotation parameters, viz. *Heat source*, *Resultative* (adjectives), and *Manner*. Their apparent closeness[185] in the parameters *Heat intensity*, *Utensil*, and *Substance* is nevertheless due to data insufficiency – an issue mentioned repeatedly throughout this chapter – as well as the broad semantic scope of the verb *cook*. Therefore, for each pair of a German verb with *cook*, different parameters (and even different values of one and the same parameter) may be significant in order to find any translation potential between them. In comparison with the German *backen*, for instance, in our annotations, *cook* strongly correlates with the *Manner* parameter value {covered}, as well as the {soft_tender} value from the *Resultative* (adjectives). *Backen* itself is no hyponym of any other German verb and is distinguished by the *Heat source* parameter value {oven} (cf. *CCCG*, Fig. 1 in Chapter 5.2.1 of this book). Therefore, due to the broadness of the meaning of *cook*, there might be some translation potential between these two verbs on the conceptual level, but only with respect to the parameters that specifically characterize *backen* (examples 48–49). In this case, *backen* becomes a relatively more specific verb than *cook*. While *cook* is so generical that it might also include *backen* in its meaning, the reverse would often result in a severely wrong lexical choice, e.g. *cook* in an *oven* is acceptable, but not {soft_tender} *backen*. It's difficult to draw conclusions based solely on this *CCCG*. In general, verbs with a broader semantic scope are harder to compare.

[185] *Cook* is a very generic verb in English and stands as a hypernym for all of the English culinary verbs and not just a specific group of verbs distinguished by any parameter, e.g. substance. Thus, cook is not a hypernym for *boil*, *steam*, *simmer*, etc., annotated as [+water], but a generic verb for both *boil* and *fry*, as well as *bake* (cf. Lehrer 1972). Therefore, *Substance* is not the most distinguishing parameter for the verb *cook* in general.

48. We fire the *oven* with wood only and *cook* pizzas at approximately 800°F. (*enTenTen15*)

49. You basically *cook* the chicken in the *oven* for around 15-20 min. from this point. *(EN_REZ)*

There is no translation potential between the pairs *backen-braise*, *backen-fry*, and *backen-sauté*, as they differ in five of the overall six annotation parameters.[186] However, even the *Substance* parameter alone is sufficient to exclude the aforementioned verb pairs as potential translation equivalents. While the English *fry* and *sauté* strongly correlate with some type of *oil-* or *fat*-like substance, *braise* correlates with some *liquid*-like ones (cf. CA of the German, English, and Armenian verbs, Chapter 5.1, Fig. 1–3). The German *backen*, on the contrary, does not correlate with any of the these substances but rather with *no substance* at all, annotated as {N_A_SUB}. For instance, when comparing the verbs *backen* and *fry*, the co-occurrences of the *Substance* parameter value {oil} and *fry* is overrepresented by 6.9 times the standard deviation, under the hypothesis of a uniform independent distribution.

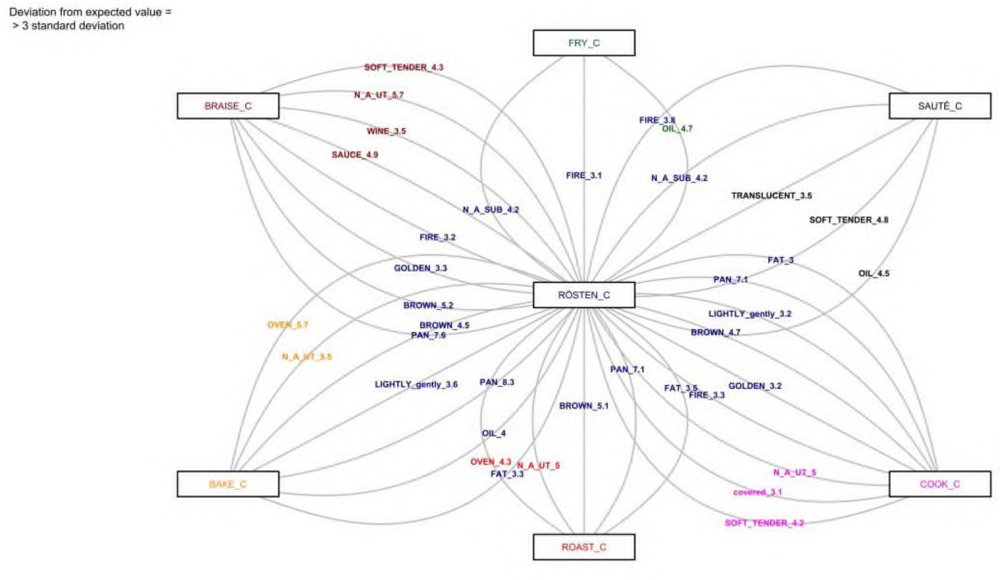


Fig. 10. Potential translation equivalents based on the language pair *German-English*

This *CCCG* (Fig. 10) compares the German specific (hyponym) verb *rösten* (to roast) to the six English verbs *bake*, *braise*, *cook*, *fry*, *roast*, and *sauté*, in pairs. The German culinary verb

[186] The parameter *Heat intensity* is not illustrated in the comparison of the German *backen* to the English culinary verbs due to insufficiency of data as an outcome of annotating non-verbalized values as N_As.

rösten [187] differs from the aforementioned English verbs in a number of parameters, leaving little space for translation potential. However, for instance, in the verb pair *rösten-fry*, the difference is observed merely in the parameters *Substance*, *Resultative* (adjectives) and *Heat source*, with at least one of them differentiating the verb *rösten* in general from all other English culinary verbs. When comparing the verbs *rösten* and *fry*, the co-occurrences of the *Substance* parameter value {oil} with *fry* exceed the expected value by 4.7 times the standard deviation, while *rösten* strongly correlates with {N_A_SUB}. It is worth noting that {N_A_SUB} here denotes *no substance*, fully corresponding to the first meaning of the definition of *rösten*, and should not be considered as annotation insufficiency.[188] Thus, the difference between the German *rösten* and the English *fry* in the *Substance* parameter is manifested solely in the first meaning of the verb rösten, while in the second meaning denoting '*braten*' there is even more potential for translation (cf. examples 50–51).

50. Das gibt uns die Gelegenheit unsere Kürbiskerne in einer Pfanne *ohne Öl* zu *rösten* und dabei immer mal zu wenden. *(detenten13)*
    [It gives us the opportunity to roast our pumpkin seeds in a pan *without oil* by stirring them occasionally].

51. *Fry* the onions in coconut *oil* till it gets translucent or even slightly golden. *(ententen15)*

On the contrary, when comparing the verb pair *rösten-roast*, the difference in the parameter *Substance* appears in the first meaning of *rösten*. Even though the English *roast* seems to cover both meanings, it lacks verbalized values denoting any type of *fat*. Thus, in the distribution of *Substance* parameter values for the verbs *rösten* and *roast*, the number of occurrences of {fat} with *rösten* exceeds those with *roast* by 3.5 times the standard deviation (examples 52–53).

52. *Margarine* in einer beschichteten *Pfanne* erhitzen und Vollkorntoast goldgelb *rösten*. *(REZ_DE)*

---

[187] "rösten"- 1. ⟨jmd., etw. röstet etw.⟩ ohne Zusatz von Fett oder Wasser durch Erhitzen (trocknen und) bräunen; 2. [landschaftlich] ⟨jmd. röstet etw.⟩ (in der Pfanne, dem Backofen oder auf dem Rost) braten oder grillen. ("Rösten", in: *Wörterbuch der deutschen Gegenwartssprache* (1964–1977) retrieved on 18.09.2022 via *DWDS*, < *https://www.dwds.de/wb/r%C3%B6sten#1* >*)*. The underlying of key words illustrating the differentiating parameter of *rösten* is mine.
"rösten"- 1. a) etwas längere Zeit (über einem Feuer, im Backofen o. Ä.) ohne Zusatz von Fett oder Wasser großer Hitze aussetzen, sodass es gar wird, eine braune Kruste bekommt, knusprig wird: Brot, Kastanien, Fleisch, Haferflocken, Nüsse rösten; einen Fisch auf dem Grill rösten; frisch gerösteter Kaffee; 2. (landschaftlich) a) braten (a): Kartoffeln rösten; b) (selten) braten (b): die Steaks in der Pfanne rösten. (DUDEN Universal Wörterbuch online) retrieved on 18.09.2022 via <*https://www.duden.de/suchen/dudenonline/r%C3%B6sten* >)).
[188] For more details, see Chapter 2.4.1 of this work.

[In a coated pan heat the *margarine* and *roast* the whole wheat toast bread].

53. You could *roast* your red bell peppers in the *oven* but I prefer mine right on the *flame* of my gas range! *(EN_REZ)*

The parameter *Utensil*,[189] however, has the greatest impact on the distribution of the annotated occurrences for the verb pair *rösten-roast*, viz. the co-occurrences {pan}[190] and the German *rösten* exceeding the expected value by 7.1 times the standard deviation (examples 54–55). The English *roast*, on the contrary, strongly correlates with *no utensil* at all, annotated here as {N_A_UT}, as it presupposes a direct (in a pan) or indirect (in an oven/over fire) heat exposure, usually with no verbalized cooking appliance.

54. Dazu *röstet* die hübsche Gastgeberin Kaffeebohnen in der *Pfanne*. *(detenten13)*
[For this *roasts* the hostess coffee beans in a *pan*].

55. Pinienkern in eine *Pfanne goldbraun* rösten. *(REZ_DE)*
[*Roast* the pine nuts in a *pan* until *golden brown*].

Thus, the *Heat source* parameter with its value {oven} (examples 56–57) is the most distinguishing of the four parameter differentiating the English verb *roast* from the German *rösten*, whereas the latter also differs from the former in the parameter *Resultative* (adjectives), in addition to s *Utensil*, and *Substance* as mentioned above. For instance, when comparing the verbs *rösten* and *roast*, the co-occurrences of the *Resultative* (adjectives) parameter {brown}[191] with *rösten* is overrepresented by 5.1 times the standard deviation, under the hypothesis of a uniform independent distribution (examples 56, 58).

[189] The parameter *Ingredient* also differentiates *rösten* from the other English verbs, but it indicates high translation potential with *roast*, which strongly correlates either with *beans (Kaffee- Kakaobohnen rösten* as *to roast coffee beans)*, *seeds (Samen)* (e.g. *Sesam*, *Kurbiskerne* rösten) or nuts *(Haselnuss, Mandeln rösten* as *to roast almond, cashew, chestnut)* (detenten13). *Bread* or meronyms as well show a strong correlation with *rösten*, for instance *Brotscheibe rösten, Brotwürfel rösten*; however, in this case, the English verb *'to toast'* is more appropriate, e.g. *toast the bread crumbs* (ententen15). Although the parameter *Ingredient* has been left out from our data analysis due to the large number of different values – making generalization difficult and yielding very small clusters, e.g. in CITs, and rather complicated plots, e.g. CA and MCA – the full list of annotations for all verbs in Appendix 1 and 5 allow us to draw conclusions regarding the high translation potential on the conceptual level between the German *rösten* and the English *roast* with respect to the *Ingredient* parameter..

[190] {Pan} is a conceptual generalization of the German and English *Utensil* values denoting some type of *pan* comprising such values as {Pfanne}, {Grillpfanne} (dt.), {roasting pan}, {pan}, {frying pan}, and {sauté pan} (en.).

[191] The *Resultative* (adjectives) parameter value {brown} is a generalization on the conceptual level (conceptual generalization) of the German {goldbraun, hellbraun, dunkelbraun, braun}, the English {brown, light brown, dark brown}, and the Armenian {darchnaguyn, shaganakaguyn, bac shaganakaguyn, mug shaganakaguyn} values .

56. *Roast* the red pepper and garlic in the *oven*. *(detenten13)*

57. *Roast* the beef in the middle of the *oven* for 20 minutes. *(EN_REZ)*

58. Butter in einer *Pfanne* aufschäumen und Pinienkerne darin bei mittlerer Hitze *hellbraun* rösten. *(detenten13)*
    [In a *pan* froth up the butter and roast the pine nuts over mediam heat until *light brown*].

There is nearly no translation potential when comparing the German verbs *rösten* to the remaining English verbs. In the verb pairs *rösten-bake* and *rösten-cook*, differences are observed in all six parameters except for *Heat intensity*.No translation potential is observed in the verb pairs *rösten-cook* and *rösten-braise* either, due to mutually exclusive values in *Substance* parameter, i.e. {oil} vs. {liquid} (examples 59–60). While *rösten* correlates with substances denoting some type of *oil* or *fat*, the verbs *cook* and *braise*, on the contrary, show strong correlations with *liquid*-like cooking substances, e.g. {wine} and {sauce} (example 61).

59. Pour in 1 cup of *broth*, cover the pan, and *braise* the squash for 10 minutes until partially cooked. (*detenten13)*

60. The recipes are evaluated by staff members and are printed in a Deliciously rich oxtail stew recipe, with oxtails braised in red wine and stock. *(ententen15)*

61. Dann in der Pfanne *Öl* erhitzen und die klein geschnittene Zwiebel hellbraun *rösten*. (*detenten13)*

When comparing the verb pair *rösten-sauté*, the differences are illustrated in the parameters *Substance*, *Resultative* (adjectives), and *Heat source*. The German verb *rösten* strongly correlates with *no substance*, while the English *sauté* with the *Substance* parameter value {oil}, indicating no translation potential between these two verbs in the first meaning of *rösten*. *Sauté* also differs from *rösten* in its strong correlation with the *Resultative* (adjectives) values {soft_tender} as well as {translucent}, excluding their translation potential in this parameter as well. *Heat source* is also a decisive parameter in considering the translation potential for *rösten* and *sauté*. Out of the total 517 annotated examples for these two verbs (231 for *rösten* and 286 for *sauté*), all 21 co-occurrences with {fire} are with *rösten*, resulting in a deviation of 3.8 times the standard deviation.

5.2.6 *CCCG* Visualization of Language Pairs: *English-German*

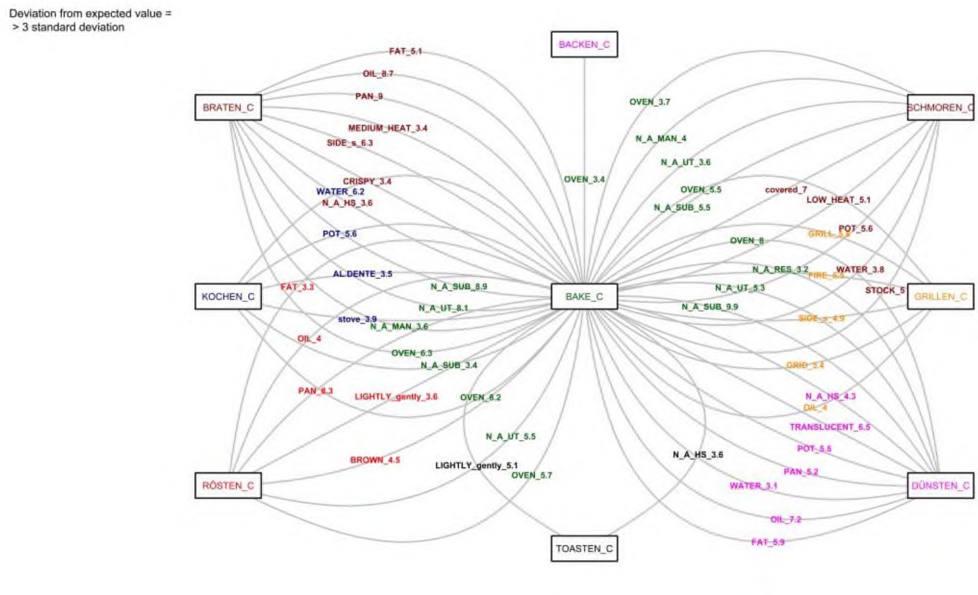


Fig. 11a. Potential translation equivalents based on the language pair *English-German*

This *CCCG* (Fig. 11a) represents the pairwise comparison of the English generic (hypernym) verb *bake* with the German verbs *backen, braten, dünsten, grillen, kochen, rösten, schmoren*, and *toasten*, with the objective of identifying, on the conceptual level, any translation potential between them. The graph visualizes drastic differences between the English verb *bake* and most of the German culinary verbs, which in most cases differ in all six annotation parameters. Unsurprisingly, the English *bake* is semantically closest to the German *backen*, differing only in the parameter *Heat source*. The difference in the distribution of the values of this parameter between the verbs *bake* and *backen*, particularly {oven}, is due to the specificity of English culinary texts, as {oven} is anchored in the definitions of both of the verbs.[192] Nevertheless,

[192] "Backen"- 1.a) aus verschiedenen Zutaten einen Teig bereiten und diesen unter Hitzeeinwirkung im *Backofen* gar und genießbar machen; b) durch Backen (1a) herstellen; Kuchen, Plätzchen backen ("backen", in: DUDEN Universal Wörterbuch online retrieved on 16.06.2022 <*https://www.duden.de/rechtschreibung/backen_herstellen_garen*>).
"Backen"- Eine unfertige Speise der Ofenhitze aussetzen, sodass sich eine Kruste bildet und sie essbar wird a) Teig im Ofen (in einer Form) hochgehen und locker, gar werden lassen b)eine Speise im zerlassenen Fett in der Pfanne rösten, braten ("backen", in: *Wörterbuch der deutschen Gegenwartssprache (1964–1977)* retrieved on 16.06.2022 via *DWDS*, <*https://www.dwds.de/wb/backen*>*).*
"bake"- transitive. To cook (food) by dry heat, without direct exposure to flame, typically in an oven, or sometimes on a heated surface such as a griddle, baking stone, etc. ("bake", in: *OED Online* This Edition (2020), retrieved on 01.08.2020 via < https://www.oed.com/view/Entry/14759?rskey=6PgYyw&result=3#eid >)
"bake"- to cook something by heating it in a closed space ("Bake" in *Definitions* Dictionary, *Antidote 9 (software, version 3). Druide informatique, Montreal, 2016*

more than half of the annotated examples for *backen* and *bake* in the parameter *Heat source* have no verbalized values, as they both presuppose some type of “heating in closed space,” e.g. *oven, brick oven, Backofen, Steinbackofen,* etc. (examples 62–63).

62. *Backen* Sie die Torte bei 180 Grad 40 bis 50 Minuten lang auf der untersten Rille. *(detenten13)*
[Bake the cake at 180°C for 40-50 minutes on the lowest shelf].

63. *Bake* the bread for 60 minutes. (*enTenTen15*)

There might be translation potential between the English verb *bake* and the German *toasten*, as they only differ in the *Heat source* and *Manner* parameters. Throughout this work, the importance of the parameter *Heat source* in differentiating *bake* from the rest of the verbs in comparison was emphasized multiple times. The absence of further edges denoting differences between *bake* and *toasten* is due to the insufficiency of data, as the overwhelming majority of values in *Substance*, *Utensil*, and *Heat Intensity* parameters are non-verbalized. The difference between *bake* and *toasten* in the *Manner* parameter has the strongest impact on excluding possible translations within this particular verb pair. When comparing *bake* and *toasten*, the co-occurrences of the *Manner* parameter value {lightly_gently} with *toasten* are overrepresented by 5.1 times the standard deviation. Even though the number of initially annotated examples with *bake* is three times greater than those with *toasten*, all 12 occurrences of {lightly_gently} are exclusively with *toasten*. Moreover, the *CCCG* (Fig. 11a) fails to single out the correlation between the English verb *bake* and the rather significant and specific *Heat source* parameter value {oven}, which would leave little space for translation potential between *bake* and *toasten*. Considering this a significant loss of information, a small *CCCG* (Fig. 11b), based on deviations above 2.0 times the square root of the expected value under the hypothesis of a discrete uniform independent distribution, was generated to compare the English *bake* with the German verbs *toasten* and *backen*, in order to better assess their translation potential.

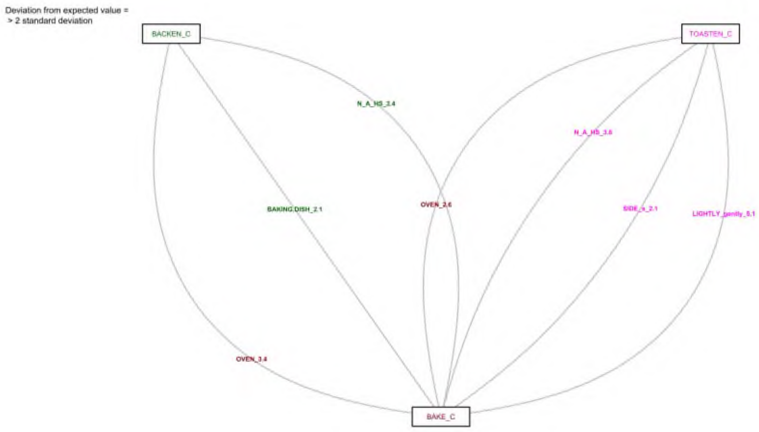


Fig. 11b. Comparison of the English verb *bake* to the German verbs *backen* and *toasten*

Since one of the primary objectives of this work is to identify translation potential based solely on intralinguistic analysis – rather than making use of bilingual dictionaries – it is of great importance to gain as much significant information as possible out of statistical evaluation (also including the *CCCG*). The partial graph (Fig. 11b) identifies strong correlations between the English verb *bake* and the *Heat source* parameter {oven}, in contrast to the German *toasten*, which strongly correlates with the non-verbalized value of the aforementioned parameter {N_A_HS}. In addition, the graph reveals another strong correlation between *toasten* and the *Manner* parameter value {sides_s}, which further excludes the translation potential between these *bake* and *toasten*, at least with respect to the aforementioned two parameters.[193] Fig.11b) thus confirms the strong correlation between {oven} and the English verb *bake*, further supporting the statement at the beginning of the graph evaluation that the semantical component of *some type of oven* inherent in the verb *backen* is not verbalized in the majority of cases. The correlation between the German verb *backen* and the *Utensil* parameter value {baking_dish} in this graph is of less significance compared to the *Heat Source* parameter. Nevertheless, it

[193] It might be claimed that simply relying on verb definitions in the respective languages could have yielded the same results as with the statistical evaluation of our linguistic data. This is because, as part of specialized texts, culinary verbs tend to be semantically dense. However, since one of our objective is to identify possible translation equivalents through the use of *CCCGs*, it is essential to first visualize the results of statistic evaluations with the suggested methods in order to fully allow readers to weigh their potential. The semantically dense definitions of the culinary verbs in this work serve nevertheless to support and double check the results of statistical evaluation of qualitative linguistic analysis. This might not hold true if the same method is used to analyze lexemes with less semantic density and/or a great spectrum of full or partial synonymic lexical relations. Thus, the semantic analysis of culinary verbs, in fact, comes to substantiate the effective implementation of multi-dimensional visualization tools based on statistical evaluation for contrastive linguistics as well as translation studies.

provides useful insight, as 11 out of 12 annotated examples with {baking_dish} occur with *backen*, yielding specific information about this particular value-verb relationship.

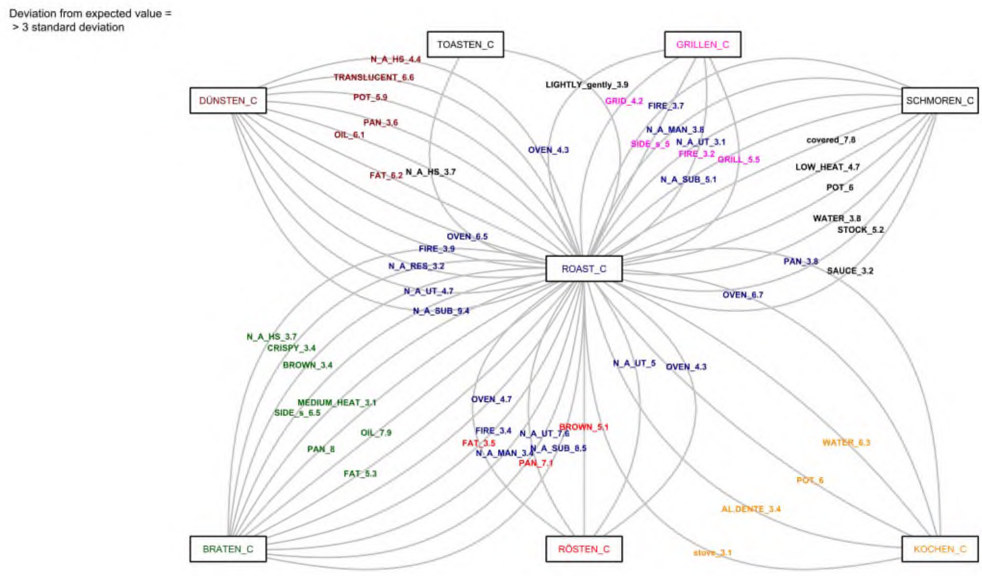


Fig. 12a. Potential translation equivalents based on the language pair *English-German*

Counting the edges which connect the verb *roast* to the German counterparts, thereby revealing semantical differences between the verb pairs in almost all six annotation parameters, practically excludes any translation potential in the verb pairs *roast-dünsten*, *roast-schmoren*, and *roast-braten* (Fig. 12a). Even though the verb pair *roast-kochen* is connected by much fewer edges, the differences in the parameter *Substance* and *Heat source* alone do not allow for any translation potential. The German verb *kochen* strongly correlates with the *Substance* parameter value {water} as well as the *Heat source* parameter {stove}, while *roast* shows a strong correlation with {oven}. Due to non-verbalized values in the parameters *Heat intensity* and *Manner* for both *kochen* and *roast*, no difference is displayed for these parameters on the graph. Of great interest are those pairs with *roast* that have less connecting edges, indicating a certain degree of semantic closeness between the verbs and, consequently, some translation potential. However, generating the preceding *CCCG* with a deviation value above 3.0 times the standard deviation caused a significant loss of information in the identification of verb pairs with a higher translation potential: *backen*, for instance, is excluded from the graph. Therefore, another graph (Fig. 12b), based on a deviation value above 2.5 times the standard deviation, has been generated to provide a more detailed visualization of the semantic differences in parameters that might still allow for some translation potential between the verbs.

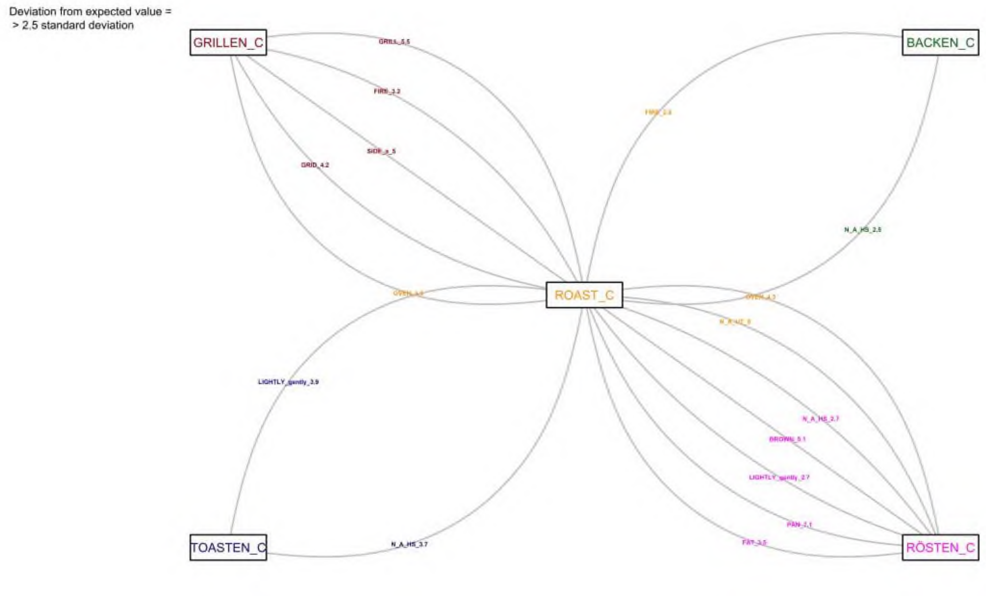


Fig. 12b. Comparison of the *English* verb *roast* to the *German* verbs *toasten*, *rösten, grillen*, and *backen*

For the sake of clarity and legibility, semantically drastically different verb pairs were left out here. Nevertheless, they were elaborated upon above (Fig. 12a). Two pairs, viz. *roast-backen* and *roast-toasten*, appear to be semantically closer verbs, both differing only in the parameter *Heat source*. Besides, *toasten* also differs from *roast* in the parameter *Manner*. When comparing these two verbs, the number of co-occurrences of the *Manner* parameter value {lightly_gently} and *toasten* exceeds the expected value by 3.9 times the standard deviation.[194] However, this effect is primarily due to the insufficiency of data in almost all parameters. Moreover, the strong correlation between *backen* and the *Heat source* parameter value {N_A_HS} is the outcome of non-verbalized values, generally *Backofen, Ofen* (oven), which are anchored in the definition of the verb itself. Thus, according to this graph alone, there is considerably large potential for this verb pair to be considered translation equivalents; however, the parameters *Ingredient* and *Dish* would shed more light on the specific group of ingredients and/or dishes which strongly correlate with *roast* and *backen*, respectively, and are clearly semantically different, thereby excluding any translation potential. For instance, the English

[194] Despite the considerable difference between the number of overall annotated examples for the verbs *roast* and *toasten*, 270 and 87 respectively, out of 17 co-occurrences of both verbs with {lightly_gently}, 12 is with *toasten*, which gives rise to such a high deviation in the distribution.

verb *roast* in the meaning of *heating in order to dry* correlates only with *nuts*, *beans* and *seeds* and would not be semantically close to the German *backen* (examples 64–65).

64. *Coffee beans are roasted* at temperatures between 370 and 540 degrees Fahrenheit. *(ententen15)*

65. Now that you've had your *chestnuts* roasted by an open fire, it's time to savor a various kind of chestnut. *(ententen15)*

In the meaning of "cooking in the oven", the verb pair *roast-backen* could serve as equivalents (examples 66–67).[195]

66. We cooked up a simple dinner – sautéed green beans, oven *roasted* small *potatoes* with rosemary and olive oil, a mushroom side dish, and a wonderful salad. *(ententen15)*

67. Wir haben im Kaminofen *Kartoffeln gebacken* - wie am Lagerfeuer. *(detenten13)*
    [We baked potatoes in the wood-burning stove - just like around a campfire.]

The same is true for the verb pair *roast-toasten*. The parameter *Ingredient* clearly distinguishes the two verbs, differentiating them semantically in at least one additional parameter.
In the verb pair *roast-grillen*, the semantic difference is observed in parameters *Heat source*, *Manner* and *Utensil*. Moreover, the German verb *grillen*[196] strongly correlates with two different values of the *Heat source* parameter, viz. {grill} and {fire}, while the English *roast* shows a strong correlation with {oven} (examples 68–69).

68. In der Glut des *Lagerfeuers grillen*, bis das Fruchtfleisch des Apfel weich gegart ist. *(detenten13)*
    [Grill in the embers of the campfire until the pulp of the apple is softly cooked].

[195] The same is true for the verb pair *roast-toasten*. The parameter *Ingredient* clearly distinguishes the two verbs differentiating them semantically in at least one additional parameter. This parameter is even anchored in the definition of these two verbs. If the process of cooking* is supposed to be *heating without oil*, then *roast* strongly correlates with different types of nuts, beans, and seeds, while *toasten* correlates with bread, including *bread crumbs, a slice of bread,* etc. For instance, …*roast* the *pumpkin seeds* (ententen2015) and …die *Brotscheibe* für die Suppe *toasten* (detenten2013).
[196] grillen- ⟨jmd. grillt (etw.)⟩ *von Fleischerzeugnissen, Fisch, Gemüse* unter Zufuhr von starker Hitze (ohne direkten Kontakt zur Wärmequelle) rösten. ("grillen", in: *Wörterbuch der deutschen Gegenwartssprache* (1964–1977) retrieved on 02.12.2022 via *DWDS*, <*https://www.dwds.de/wb/grillen* >).
grillen- schwaches v. auf dem Grill rösten. ("grillen", in: DUDEN Universal Wörterbuch online) retrieved on 02.12.2022. <*https://www.duden.de/rechtschreibung/grillen_roesten_braten_zubereiten*>))

69. Lammlachse auf dem heißen *Grill* rundum ca 8 Minuten *grillen*. *(detenten13)*
    [Grill the lamb fillets on the hot grill on both sides for ca. 8 minutes].

The parameter *Heat source* is decisive in semantically differentiating the verbs *roast* and *grillen* and should be taken into consideration regarding their translation potential. The strong correlation of *grillen* with the *Utensil* parameter value {grid}[197] also reduces the translation potential for this verb pair, since a {grid} is typically used over an open fire and not in an oven (example 72). In addition, the co-occurrences of the *Manner* parameter value {side_s} with the German verb *grillen* are overrepresented by 5.0 times the standard deviation, implying that the process of *grillen* might also include *turning* the food to grill on one or both sides (examples 70-72).

70. Forellen auf dem heißen *Rost* ca. 12 – 15 Minuten *grillen*. *(REZ_DE)*
    [Grill the trouts on hot grids for ca. 12-15 minutes].

71. Nun das Fleischstück *auf beiden Seiten* über heißer Glut eineinhalb Minuten *grillen*. *(REZ_DE)*
    [Now grill the piece of meat over hot embers for 1,5 minutes on both sides].

72. Den Fisch *auf jeder Seite* 3-4 min. *grillen* (nicht braten) bis er an den Rändern gut gebräunt ist. *(detenten13)*
    [Grill the fish on both sides for 3-4 minutes (do not fry) until the edges are good brown].

The semantic difference between the verbs *roast* and *rösten* is manifested in all six parameters except for *Heat intensity*. Especially the strong correlation of the German *rösten* with the *Utensil* parameter value {pan}, which is overrepresented by 7.1 times the standard deviation considerably lowers their translation potential.

[197] The *Utensil* parameter value {grid} is a generalization of the values {grid, rack} (en.) and {Rost} (dt.) to ensure the comparison of the given verbs on the conceptual level. For a full list of annotation parameter neutralization (generalization), see Appendix 9.

Conclusions and Future Research Desiderata

This bottom-up, corpus-based work describes the Armenian, German, and English culinary lexical fields, and in particular culinary verbs, with the objective, among other things, to identify possible translation potential based on *intra-* and *inter*linguistic analysis. The TenTen Corpus Family available in *Sketch Engine* (https://www.sketchengine.eu/documentation/tenten-corpora/), viz. the German Web 2013 (deTenTen13) and English Web 2015 (enTenTen15), as well as manually compiled English *(EN_REZ)*, German *(REZ_DE)* and Armenian *(ARM)* much smaller corpora served as the basis for developing the methodology of this work and carrying out the analysis. By combining qualitative (manual semantic annotation) and quantitative (computational analysis based on statistical significance) methods, eight German (*backen, braten, dünsten, grillen, kochen, rösten, schmoren*, and *toasten), six English (bake, braise, cook, fry, roast*, and *sauté*) as well as six Armenian culinary verbs (*բովել [bovel], եփել [ephel], թխել [thxel], խաշել [xashel], շոգեխաշել [shogexashel]*, and *տապակել [tapakel]*) were investigated with regard to their collocational meanings. Among the different association scores measuring the strength of two words co-occurring in the same contextual environment, *logDice* (Killgariff et al. 2010) served as the basis for extracting the collocations of the aforementioned culinary verbs in the respective languages. A special annotation quantity determination model was built to identify an appropriate distribution of the examples for the semantic annotation (see Chapter 2.2, 3.1 and 4.1). Accordingly, 200 occurrences were systematically annotated for the more general (hypernym) culinary verbs, e.g. the German verb *braten*, while for the more specific (hyponym) verbs, e.g. the German verb *toasten*, the annotation covered 100 occurrences. The following parameters were annotated: *Substance*, *Utensil*, *Manner, Resultative* (adjectives), *Heat intensity*, and *Heat source*.[198] Fig. 1 below illustrates, in numbers and in the respective languages, the initially annotated and subsequently reduced occurrences based on their relevance for the *intra*linguistic analysis and, on a conceptual level, for translation potential (Fig.1).

[198] Initially the parameters were eight; the *Ingredient* and *Dish* proved to be less relevant for the comparison, however, small qualitative analysis were carried out with respect to the parameter *Ingredient*, e.g. on the German verb *rösten* in Chapter 2.4 of this work. For the introduction of the annotation parameters see Chapter 2.1.

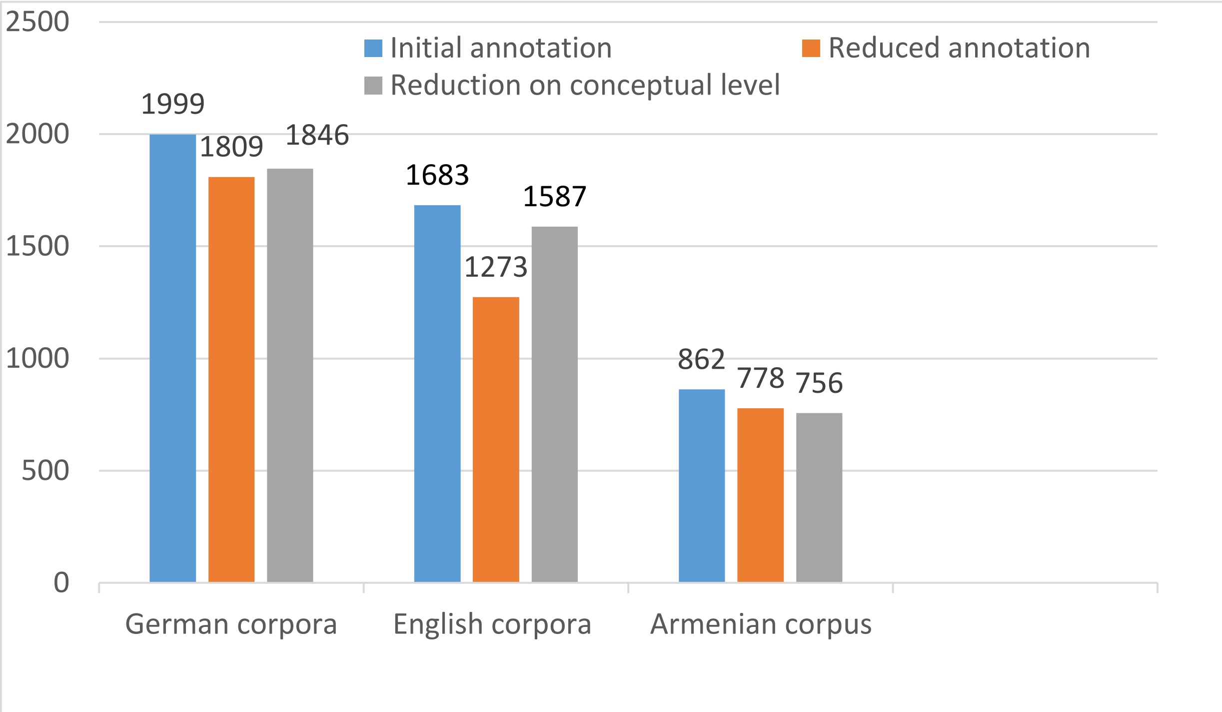


Fig. 1. Initially annotated and the final reduced occurrences

This work is predominantly methodological since it pursues its objectives by adopting the collocation analysis method and trying to bridge the gap between the conventional logic-based lexical semantics, i.e. the componential analysis of describing fields and frames as well as the distributional approach to the analysis of linguistic data. The statistically-significant co-occurrences of the investigated culinary verbs and the manual semantic annotation of authentic corpus context examples across the aforementioned parameters were then analyzed using different visualization techniques, such as *Correspondence Analysis* (CA) and *Multiple Correspondence Analysis* (MCA) (Chapters 2.3, 3.2. and 4.2.), *Mosaic-Plots* and *Context-Conditional Correlation Graphs* (CCCG) (Chapters 2.4 and 2.4.1, 3.3 and 4.3, 5.2.1-5.2.6) as well as *Conditional Inference Trees* (CIT) (Chapters 2.5, 3.4 and 4.4).

The combination of quantitative and qualitative analysis methods, viz. the collocation extraction based on the *LogDice* association score and the manual semantic annotation of the occurrences, accompanied by the visualizations of the correlations and their distribution in the respective corpora, has resulted in interpretable word profiles. This methodology differs insofar from the vector space model (VSM) (*word embedding*) and similar methods considered as "computational implementation of distributional hypothesis" (Desagulier 2017, 2018, Mikolov et al. 2013). VSM lacks interpretability due to its considerable reduction of dimensions by applying mathematical techniques, such as the *Principal Component Analysis* (PCA), making it impossible to explain the underlying contextual and semantic similarity between the correlated words and to interpret them directly as collocational dimensions. While such methods are useful for Machine Translation (MT), they are less suitable for linguistic analysis to some extent. The methodology suggested in this work retains the interpretability of these word profiles, since the models developed in this work analyze semantic categories and not merely word similarity relationship.

The initial theoretical background and the starting point of the manual semantic annotation in the scope of this work are based on A. Lehrer's (1972, 1974) classification of German and English culinary verbs with respect to the culinary triangle of Lévi-Strauss (1969). Based on authentic corpus evidence, Lehrer's classification was revisited in the framework of this research by adding more distinctive features for distinguishing the culinary verbs derived from the manual semantic annotation parameters. Moreover, Lehrer introduced seven binary as well as three non-binary components (1974:157-160), for instance, the English verb *fry* being [+Fat] while *boil* [+Water]. However, she did not group these components in semantic categories, such as *Substance*. Besides, culinary verbs allowing both the presence or the absence of a cooking substance in their first and second senses respectively were also left out in Lehrer's analysis. This work comprehensively covers such [±Fat] verbs, e.g. German *rösten*, Armenian *bovel* English *roast*, in particular through the *CCCGs* (Chapters 2.4, 3.3 and 4.3) analysis, where they are compared in pairs, and in *CIT* analyses, which single out the parameter determined use context clusters (Chapters 2.5, 3.4 and 4.4).

Since there is no classification of the Armenian culinary verbs in previous researches, an attempt is made to classify at least those considered in the linguistic analyses of this work, based on the manual semantic annotation and in analogy to the German and English culinary verbs suggested by Lehrer's componential analysis (Table 1).

եփել [ephel] (to boil/to cook in/with water/to prepare)			**տապակել** **[tapakel]** (to fry)	**խորովել** **[xorovel]** (a very culture-specific verb, to grill without any cooking substance)		**թխել** **[thxel]** (to bake)
եռացնել [eratsnel] (to boil) transitive [+Water]	**եռալ [eral]** (to boil) intransitive [+Water] [+Vigorous] [±Kettle] e.g. tea, water, milk	**եփել [ephel]** (full boil) [+Water] [+Vigorous] **շոգեխաշել [shogexashel]** (to braise*/to cook in little water*/ to steam*) [+Water] [-Vigorous] [+Covered] **խաշել** [xashel] (to cook in water) [±Vigorous]	**կարմրացնել [karmratsnel]** (to make red*/brown) [-Water] [+Fat] [±Direct heat]	**խորովել [xorovel],** (to barbecue, to grill) [-Fat] [±Direct heat] e.g. meat or veggies (with the skin on)	**բովել [bovel] (roast)** [±Fat] [+Browning] [+Direct heat] e.g. coffee beans (without fat) e.g. flour (both with and without fat)	[+Oven] [+Tonir] [±Direct heat]

Table 1. Classification of the Armenian culinary verbs based on corpus analysis

The corpus-based authentic evidence and analysis within the scope of this work might restate Lévi-Strauss' (1969) idea of universal differentiating features, e.g. the presence or absence of *air*, *water*, or *fat* during the preparation process. This hold true, however, only for certain verbs, such as *bake* (en.), *backen* (dt.), *thxel* (arm.), being [+Air], *fry* (en.), *braten* (dt.), *tapakel* (arm.) [+Fat, -Water], *boil* [+Water] (not considered in this work), *xashel* (arm.) and *ephel* (arm.), being [+Water]. By contrast, for verbs such as *roast* (en.), *bovel* (arm.), *rösten* (dt.) with [±Fat], *cook* (en.) [±Fat], [±Water], *schmoren* (dt.) [±Water], [±Fat] no universal features could be identified. Thus, more specific hyponym verbs demand further differentiating semantic features and could benefit, for instance, from pairwise comparison, namely the *CCCGs* for both intralinguistic and contrastive analysis.

Verbs with more verbalized values in more parameters show greater differences than those with non-verbalized parameters. The absence of verbalized values in any parameter could denote two different things: either the value is already strongly implicated in the usage of the verb, e.g.

the *Substance* parameter value {water} for the verbs *ephel and xashel* in Armenian, as well as *kochen* in German, or the actual absence of the value, e.g. {Fat} in the description of the German verb *rösten*. More general verbs (hypernyms) may have, at least theoretically, greater translation potential then specific ones (hyponyms) due to their semantic density.

Even though statistically significant differences between verbs may be observed in many parameters, the *Context-Conditional Correlation Graph* analysis in this work suggests that, for some verbs, certain parameters are more decisive than others, so that only the difference in that specific parameter is sufficient to semantically differentiate one verb from another. Table 2 below groups the verbs according to the most significantly differentiating parameter.

Substance	**Ingredient**	**Utensil**	**Resultative (adjectives)**	**Manner**	**Heat intensity**	**Heat source**
braten (dt.), e.g. oil, fat	rösten (dt.), e.g. Pinienkerne ohne Fett rösten	bake (en.), e.g. baking dish, baking form	dünsten (dt.), e.g. glasig	schmoren (dt.), e.g. *zugedeckt*	schmoren (dt.), e.g. *auf kleiner Flamme* [over low heat]	backen (dt.), e.g in Backofen, Ofen
braise (en.), e.g. wine, beer	roast (en.), e.g. roast the sesame seeds in a dry pan	thxel (arm.), [bake], e.g. ձևաման (dzevaman) [baking dish, form]	braten (dt.), e.g. knusprig, braun, kross	shogexashel (arm.), [braise, cook* with little water], e.g. կափարիչը փակ (kaparichy pak) [with the lid on/covered]		grillen (dt.), e.g. auf dem Grill, Glut
tapakel (arm.) [fry], e.g. oil, fat	xashel (arm.), [cook* in water], e.g. ձու, կարտոֆիլ, միս (dzu, kartofil, mis)		sauté (en.), e.g. translucent			bake (en.), e.g. in oven

	e.g.[eggs, potatoes, meat]						
bovel (arm.) [roast], e.g. without any substance used to roast nuts, however, with oil as in *to roast* flour*							thxel (arm.), [bake], e.g. թոնիր (tonir) [tandoori]
rösten (dt.), e.g. without oil, viz. ohne Öl, die Kerne or Samen rösten							
roast (en.), e.g. dry, without oil							
shogexashel (arm.), e.g. with little water							

Table 2. Most significant verb-specific annotation parameters

In general, *Substance* can undoubtedly be considered the most important among our annotation parameters. Given the specialized nature of culinary texts, these annotation parameters could effectively be extended as a tagging framework for additional languages.

Beyond contributing to language acquisition and language teaching, the methodology and models developed in this work could also be applied in computation linguistics, computational lexicography, frame semantics (e.g. FrameNet), as well as manually compiled dictionaries (both monolingual and bilingual). Especially for lexicographic purposes, this work could also contribute to developing specific features in the computational lexicography to formulate more fine-grained distinctions between the usages of culinary verbs in the dictionaries. For instance, the entries for the German verb *rösten* in Duden und DWDS (see below) could benefit from the addition of at least two of our annotation parameters, viz. *Manner* and *Utensil*, with their respective values {kurz} and {sanft}, as well as {Pfanne}, correlating with *'rösten ohne Fett'* (see sub Chapter 2.4.1 for the *Mosaic-Plot* visualization of *'rösten'* and *'ohne Fett rösten'*).

- rösten"- 1. ⟨jmd., etw. röstet etw. ⟩ ohne Zusatz von Fett oder Wasser durch Erhitzen (kurz, sanft) (in der Pfanne) (trocknen und) bräunen; 2. [landschaftlich] ⟨jmd. röstet etw.⟩ (in der Pfanne, dem Backofen oder auf dem Rost) braten oder grillen; 3. [Hüttenwesen] ⟨jmd., etw. röstet etw⟩ von Erzen: bei der Aufbereitung stark erhitzen, um Schwefel oder Arsen zu entfernen (DWDS).
- "rösten"- 1. a) etwas längere Zeit (über einem Feuer, im Backofen o. Ä.) ohne Zusatz von Fett oder Wasser großer Hitze aussetzen, sodass es gar wird, eine braune Kruste bekommt, knusprig wird: Brot, Kastanien, Fleisch, Haferflocken, Nüsse rösten; einen Fisch auf dem Grill rösten; frisch gerösteter Kaffee; 2. (landschaftlich) a) braten (a): Kartoffeln rösten; b) (selten) braten (b): die Steaks in der Pfanne rösten (DUDEN).

Future Work

Despite the tangible results and outcomes of this work, there are still future research desiderata in this field.

- Semantically, the model could be tested for ambiguous words, treated theoretically as cases of polysemy and homonymy. The model has been successfully applied to a small group of mostly unambiguous words, viz. the culinary verbs. Due to the specificity of annotation parameters, homonymies, e.g. *to smoke* (a cigarette vs. meat) would not overlap as the types of context parameters would be different. For ambiguous words overlapping in their central meaning, e.g *dämpfen* (dt.) ('Fisch dämpfen' vs. 'Lärm dämpfen') and diverging in their sub-meaning, viz. secondary central meaning, the model could be further developed to systematize lexicographical dictionaries and suggest a better and more objective structure for these hierarchies.
- Develop a specific annotation tag set including particular features for values not explicitly verbalized in the text, such as implications or semantic density, for instance in FrameNet, for encompassing cases like the word *սոխառած* (sokharadz) in Armenian, which denotes *already peeled and diced onions sautéed in oil*.
- Develop further annotation tag sets, for example, by combining the existing *Heat source* with *Cooking time* for analyzing verbs such as the German *ziehen lassen* where the *Heat source* has stopped but the cooking process continues.

- Test out culinary verbs outside the classifications used in this work, especially the culture-specific ones which are also used for creating metaphors, e.g. the German *abschrecken* (literally: to frighten) (‘die Eier kochen, mit kaltem Wasser abschreicken’) vs. the Armenian *վախեցնել* [vakhetsnel] (literally: to frighten, but denoting “to frighten in hot water” rather than in cold, as in German). Here the target domain is *cooking* while the source domain is *fear*.

References

Alberti, Gianmarco (2013). "An R Script to Facilitate Correspondence Analysis. A Guide to the Use and the Interpretation of Results from an Archaeological Perspective". Archeologia e Calcolatori 24:23–53.

Alian, Marwah, and Arafat Awajan (2023). "Arabic Word Sense Disambiguation Using Sense Inventories". International Journal of Information Technology 15. doi: 10.1007/s41870-022-01147-w.

Allan, Keith. (2001). Natural Language Semantic. UK: Blackwell Publishers

Allan, Keith et al. (2010). The English Language and Linguistics Companion. UK: Palgrave Macmillan.

Apresjan, J.D. (1966). Analyse distributionnelle des significations et champs sémantiques structurés. Langages, 1, 44-74.

Apresjan, J.D. (1974). Regular polysemy. Linguistics, 12(142), 5-32. https://doi.org/10.1515/ling.1974.12.142.5

Apresjan, J.D. (1992). Systemic Lexicography as a Basis of Dictionary-making. Dictionaries: Journal of the Dictionary Society of North America 14, 79-87.

Apresjan, Juri D. (2000). Systematic Lexicography. Translated by Kevin Windle. Oxford University Press.

Apresjan, Juri D. (2002). "Principles of systematic lexicography". In Marie-Hélène Corréard (ed.), *Lexicography and Natural Language Processing:* A Festschrift in Honour of B. T. S.

Atayan, Vahram, Volker Gast, Bettina Fetzer, Julien Biege, Sophie Hettrich, and Anne Weber (2019). "Unmittelbare Nachzeitigkeit Im Deutschen Und Französischen: Eine Studie Auf Grundlage Des OpenSubtitles-Korpus". *Akten Der VIII. Internationalen Arbeitstagung Zum Romanisch-Deutschen Und Innerromanischen Sprachvergleich*, Innsbruck, 29.8.-1.9.2016. 223–49.

Atayan, Vahram, and Gast, Volker (2024). " Data visualization in corpus-based translation studies and contrastive linguistics". (In preparation), in Tomáš Svoboda, Olga Nádvorníková, Věra (eds.) *Advancing interdisciplinarity in empirical translation research*.

Altenberg, Bengt, and Sylviane Granger (2002). "Recent Trends in Cross-Linguistic Lexical Studies". In *Lexis in contrast:corpus-based approaches* 7, 3–48. Amsterdam [a.o]: Benjamins.

Atkins, S., Clear, J. and N. Osler. (1992). "Corpus Design Criteria". *Literary and Linguistic Computing* 7.1, 1-16. doi: 10.1093/llc/7.1.1.

Babych, Bogdan, Anthony Hartley, Serge Sharoff, and Olga Mudraya (2007). "Assisting Translators in Indirect Lexical Transfer". *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics.* June 23-30, 2007. 136–43.

Baker, C. F., Fillmore, C. J., & Lowe, J. B. (1998). The Berkeley FrameNet project. In *COLING 1998 Volume 1: The 17th International Conference on Computational Linguistics*.

Baker, C. F., Fillmore, C. J., & Cronin, B. (2003). The structure of the FrameNet database. *International Journal of Lexicography*, *16*(3), 281-296.

Baker, M., Francis, G. (Ed.), & Tognini-Bonelli, E. (Ed.) (1993). "Corpus Linguistics and Translation Studies: Implications and Applications". In Text and Technology: In Honour of John Sinclair John Benjamins Publishing Company

Baker, M. (2011). In Other Words. A coursebook on translation (2nd ed.). London and New York: Routledge.

Barsalou, Lawrence W. (1992). "Frames, Concepts, and Conceptual Fields". In A. Lehrer & E. F. Kittay (eds.). *Frames, Fields, and Contrast. New Essays in Semantic and Lexical Organization.* 21-74. Hillsdale, New Jersey, Hove and London: Lawrence Erlbaum Associates, Publishers.

Bartsch, Sabine. (2004). Structural and functional properties of collocations in English: a corpus study of lexical and pragmatic constraints on lexical co-occurrence. Tübingen: Narr.

Bergenholtz, Henning/Sven Tarp (1994). "Mehrworttermini und Kollokationen in Fachwörterbüchern", In Schaeder, Burkhard/Henning Bergenholtz (eds.): Fachlexikographie: Fachwissen und seine Representation in Wörterbüchern. Tübingen: Gunter Narr, 385-419 [=Forum für Fachsprachen-Forschung; Bd. 23]

Berty, Katrin (2017). Kollokationen als Problemgrößen der Her-Übersetzung: empirische Untersuchungen syntagmatischer Übersetzungseinheiten zur Herleitung allgemeiner wie sprachspezifischer Übersetzungsstrategien. Wissenschaftlicher Verlag Trier, Trier.

Biber, D. (1994). "Representativeness in Corpus Design". In Current Issues in Computational Linguistics: in Honour of Don Walker, Zampolli A., Calzolari N. and M. Palmer (eds.). Linguistica Computazionale IX.X. Giardini Editori e Stampatori in Pisa and Kluwer Academic Publishers.

Blank, A. (2001). Einführung in die lexikalische Semantik für Romanisten (Bd. 45). Tübingen: Niemeyer.

Bolinger, D. and Donald A. Sears (1968). Aspects of language. New York [u.a.]: Harcourt, Brace & World.

Camacho-Collados, José, and Mohammad Taher Pilevar (2018). "From Word To Sense Embeddings: A Survey on Vector Representations of Meaning". Journal of Artificial Intelligence Research 63:743–88. doi: 10.1613/jair.1.11259.

Caro Cedillo, A. (2004). Fachsprachliche Kollokationen: ein übersetzungsorientiertes Datenbankmodell Deutsch-Spanisch (Bd. 63). Tübingen: Narr.

Carnap, Rudolf (1956): Meaning and Necessity. Chicago: University of Chicago Press.

Church, Kenneth Ward, and Patrick Hanks (1990). "Word Association Norms, Mutual Information, and Lexicography". *Computational Linguistics* 16(1): 22–29.

Coseriu, E. (1964). "Pour une sémantique diachronique structurale". *Travaux de linguistique et de littérature 2 (1)*. University of Strasbourg.

Coseriu, E. (1967). Lexikalische Solidaritäten. *Poetica*, *1*(3), 293-303. https://doi.org/10.30965/25890530-00103002

Coste, Octavian (2013). "Lexical Gaps in Romanian From the Perspective of English: The Case of Culinary Verbs". *Romanian Journal of English Studies 9(1):* 22–29. doi: doi:10.2478/v10319-012-0003-1.

Couturat, Louis (1901). La logique de Leibniz [The Logic of Leibniz]. Paris: Felix Alcan.

Couturat, Louis (1903). Opuscules et fragments inédits de Leibniz [Minor Works and Unpublished Fragments of Leibniz]. Felix Alcan. Republished 1966, Georg Olms.

Croft, William A. (2000). Explaining Language Change: An Evolutionary Approach. Longman.

Croft William and D. Alan Cruse (2004). Cognitive Linguistics. Cambridge University Press.

Cruse, David A. (1986). Lexical semantics. Cambridge [u.a.]: Cambridge University Press.

Demaecker, Christine (2017). "Wine-Tasting Metaphors and Their Translation: A Cognitive Approach". *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication 23(1):* 113–31.

Desagulier, Guillaume (2017). Corpus Linguistics and Statistics with R: Introduction to Quantitative Methods in Linguistics. Springer Cham.

Desagulier, Guillaume (2018). Word embeddings: the (very) basics. Around the word. Retrieved May 29, 2024, from https://doi.org/10.58079/n4uj

Di Franco, G. (2016). "Multiple correspondence analysis: one only or several techniques?", *Quality & Quantity, 50(3):* 1299-1315. https://doi.org/10.1007/s11135-015-0206-0

Dournes, Jacques (1969). "Modèle Structural et Réalité Ethnographique (A Propos Du «Triangle Culinaire»)". *Homme 9(1):* 42–48. doi: 10.3406/hom.1969.367017.

Ducháček, Otto (1959). Champ conceptuel de la beauté en français moderne. Vox Romanica 18(1): 297-323.

Eberle, Thomas S. (1984). 'Sinnkonstitution in Alltag und Wissenschaft. Der Beitrag der Phänomenologie an die Methodologie der Sozialwissenschaften'.

Evert, Stefan (2005). "The Statistics of Word Cooccur-Rences: Word Pairs and Collocations".

Filmore, Charles J. (1982). Frame semantics. In *Linguistics in the Morning Calm*. Seoul: Hanshin Publishing Co., 111-137.

Firth, J. R. 1950. "Personality and language in society". *Sociological Review 62.* Reprinted 1957 in *Papers in Linguistics 1934–1951.* Oxford University Press.

Firth, J. R. [1951] (1957a). "Modes of meaning". In *Papers in Linguistics 1934–1951.* Oxford University Press.

Firth, J. R. (1957b). "A synopsis of linguistic theory 1930–1955". In *Studies in Linguistic Analysis.* Philological Society, Oxford. Reprinted 1968 in F. R. Palmer (ed.), *Selected Papers of J. R. Firth.* Longman.

Frege, G. (1892). Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, *100*, 25-50.)

Friedrich, Dieter (2020). Verstehen, lernen und übersetzen: Kollokationen als Einheiten des korpusbasierten Lernens in der Praxis der Fachübersetzung: das deutsch-amerikanische Doppelbesteuerungsabkommen als Fall. WVT Wissenschaftlicher Verlag: Trier.

Friendly, Michael (1994). "Mosaic Displays for Multi-Way Contingency Tables". Journal of The American Statistical Association 89(425): 190–200. doi:10.1080/01621459.1994.10476460.

Frost, Jim (2019). Introduction to Statistics. An Intuitive Guide for Analyzing Data and Unlocking Discoveries. ebook. 248 pages.

Geeraerts, Dirk (2006). "Methodology in Cognitive Linguistics". In G. Kristiansen, M. Achard, R. Dirven & F. Ruiz de Mendoza Ibáñez (ed.), *Cognitive Linguistics*. 21-50. Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110197761.1.21

Geeraerts, Dirk (2010). Theories of lexical semantics (1. Aufl.). Oxford [u.a.]: Oxford Univ. Press.

Gennaro, Giovanni, Amedeo Buonanno, and Francesco Palmieri (2021). "Considerations about Learning Word2Vec". *The Journal of Supercomputing 77.* doi: 10.1007/s11227-021-03743-2.

Giacomini, Laura (2011). An Onomasiological Dictionary of Collocations: Mediostructural Properties and Search Procedures. 241–67.

Giora, Rachel (1997). "Understanding figurative and literal language: The graded salience hypothesis". *Cognitive Linguistics 8 (3).* Reprinted in Hanks and Giora (eds., 2012), vol. 4.

Giora, Rachel (2003). On Our Mind: Salience, Context, and Figurative Language. Oxford University Press.

Gipper, Helmut (1959). "Sessel oder Stuhl? Ein Beitrag zur Bestimmung von Wortinhalten im Bereich der Sachkultur". In Helmut Gipper (ed.), *Sprache, Schlüssel zur Welt: Festschrift für Leo Weisgerber.* Pädagogischer Verlag Schwann.

Greenacre Michael and Jörg Blasius (2006). Multiple correspondence analysis and related methods. CRC Press.

Greimas, Algirdas (1966). Sémantique structurale. Recherche de méthode. Paris: Larousse.

Gries, Stefan Thomas (2013). Statistics for linguistics with R: a practical introduction (2nd ed.). Berlin [u.a.]: Mouton de Gruyter.

Gries, Stefan Thomas (2021). Statistics for Linguistics Using R: A Practical Introduction (3rd ed. Berlin & Boston: Mouton de Gruyter.

Halliday, Michael (1985). "Dimensions of discourse analysis: Grammar". In Teun A. van Dijk (ed.), *Handbook of Discourse Analysis. Vol. 2, Dimensions of Discourse.* Academic Press.

Hanks, Patrick (2013). Lexical Analysis Norms and Exploitations. Cambridge, Massachusetts, London, England: The MIT Press.

Hareide, Lidun (2019). In PaCor <1., 2016, Santiago de Compostela>. *Comparable Parallel Corpora: A Critical Review of Current Practices in Corpus-Based Translation Studies*. (I. Doval & M. T. Sánchez Nieto, Eds.) (Bd. volume 90) (19-38). Amsterdam; Philadelphia: John Benjamins Publishing Company. http://doi.org/10.1075/scl.90

Harris, Zellig S. (1954). "Distributional Structure", *WORD, 10:* 2-3, 146-162, doi: 10.1080/00437956.1954.11659520

Hartigan, John A., and Beat Kleiner (1981). "Mosaics for Contingency Table". In (Ed.) W. F. Eddy, *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*. New York, NY: Springer US. 268–73.

Hausmann, Franz Josef (1984): "Wortschatzlernen ist Kollokationslernen. Zum Lehren und Lernen französischer Wortverbindungen". In *Praxis des neusprachlichen Unterrichts 31*, 395-406.

Hausmann, Franz Josef (1988). "Grundprobleme des zweisprachigen Wörterbuchs". In K. Hyldgaard-Jensen (Ed.), Proceedings of the Third International Symposium on Lexicography: May 14–16, 1986, at the University of Copenhagen (pp. 137-154). Berlin, Boston: Max Niemeyer Verlag. https://doi.org/10.1515/9783111347349-004

Hausmann, Franz Josef. (1989). "Le dictionnaire de collocations". In F. J. Hausmann, O. Reichmann, H. E. Wiegand, L. Zgusta (ed.), *Dictionnaires. Encyclopédie internationale de lexicographie, tome I (Manuel de linguistique et des sciences de communication, 5/1),* Berlin/New York: De Gruyter, 1010- 1019

Hausmann, Franz-Josef (2003). "Was sind eigentlich Kollokationen?" In: Steyer, Kathrin: *Wortverbindungen- mehr oder weniger fest. Institut für Deutsche Sprache Jahrbuch.* 309-334

Hausmann, Franz Josef (2007). "Die Kollokationen im Rahmen der Phraseologie – Systematische und Historische Darstellung". *Zeitschrift für Anglistik und Amerikanistik 55(3).* doi: 10.1515/zaa.2007.55.3.217.

Heřman, Ondřej (2021). "Precomputed Word Embeddings for 15+ Languages". In *RASLAN.*

Hjelmslev, Louis (1958). "Dans quelle mesure les significations des mots peuvent-elles être considérées comme formant une structure?" In Eva Siversten (ed.), *Proceedings of the Eighth International Congress of Linguists.* 636-654. Oslo: Oslo University Press.

Hjelmslev, Louis (1953). "Prolegomena to a theory of language". Baltimore: Indiana University Publications in *Anthropology and Linguistics (IJAL Memoir, 7) (2. ed.* (slightly rev.): Madison: Univ. of Wisconsin Press, 1961)

Hornby, A. S. (1954). A Guide to Patterns and Usage in English. Oxford University Press.

Hödl, Nicola (2004). "Fein Schneiden – Hacher Finement – Picar Finamente Das Deutsche Verbkompositum Mit Resultativem Adjektiv Und Seine Französischen Und Spanischen Entsprechungen Im Kochrezept". Vox Romanica 63(1):169–89.

Hunston, Susan (2022). Corpora in applied linguistics (2. Aufl.). Cambridge, United Kingdom ; New York, NY, USA ; Port Melbourne, VIC, Australia ; New Delhi, India ; Singapore: Cambridge University Press. http://doi.org/10.1017/9781108616218

Hunston, Susan, and Gill Francis (2000). Pattern grammar: a corpus-driven approach to the lexical grammar of English (Bd. 4). Amsterdam [u.a.]: Benjamins.

Ipsen, Günther (1924). Stand und Aufgaben der Sprachwissenschaft. Heidelberg

Jackendoff, Ray (1983). Semantics and Cognition. MIT Press.

Jackendoff, Ray (1990). Semantic Structures. MIT Press.

Jackendoff, Ray (1991). "Parts and boundaries". *Cognition 41:* (1–3).

Jackendoff, Ray (2002). Foundations of Language. Oxford University Press.

Jakobson, Roman and Morris Halle. (1971). Fundamentals of language (2. Aufl., Bd. 1). The Hague [u.a.]: Mouton

Jakobson, Roman (1959/1971). "On Linguistic Aspects of Translation". In Roman Jakobson *Selected Writings II: Word and Language*. 260–266. Hague and Paris: Mouton. https://doi.org/10.1515/9783110873269.260

Jenset, G. B. and McGillivray, B. (2017). Quantitative historical linguistics: a corpus framework (1. Aufl., Bd. 26). Oxford: Oxford University Press.

Jolles, Andre (1934). Antike Bedeutungsfelder. *Beiträge zur Geschichte der deutschen Sprache und Literatur* 58. Tübingen.

Kastovsky, Dieter (1982). Wortbildung und Semantik (Bd. 14). Düsseldorf [u.a.]: Schwann-Bagel.

Katz, J. J., & Fodor, J. A. (1963). "The Structure of a Semantic Theory". *Language 39(2),* 170–210. https://doi.org/10.2307/411200

Kilgarriff, Adam, Pavel Rychlý, Pavel Smrž, and David Tugwell (2004). "Itri-04-08 the Sketch Engine". Information Technology.

Killgariff, Adam, Vojtěch Kovář, Simon Krek, Irena Srdanović, and Carole Tiberius (2010). "A Quantitative Evaluation of Word Sketches". *Proceedings of the 14th EURALEX International Congress*. 372–79.

Koch, Peter (2012). "The pervasiveness of contiguity and metonymy in semantic change". In K. Allan & J. A. Robinson (Eds.), *Current methods in historical semantics*. Berlin: De Gruyter Mouton. 259–311.

Lakoff, George (1987). Cognitive models and prototype theory. In U. Neisser (Ed.), *Concepts and conceptual development: Ecological and intellectual factors in categorization*. Cambridge University Press. 63–100.

Langacker, Ronald W. (1991). Concept, Image, and Symbol: The Cognitive Basis of Grammar. Mouton de Gruyter

Leech, Geoffrey. (1992). Corpora and theories of linguistic performance. In J. Svartvik (ed.), Directions in Corpus Linguistics. Proceedings of the Nobel Symposium 82 , Vol. 65 (105–122). Mouton de Gruyter.

Lehrer, Adrienne (1969). "Semantic Cuisine". *Journal of Linguistics 5(1),* 39–55. http://www.jstor.org/stable/4175017

Lehrer, Adrienne (1972). "Cooking Vocabularies and the Culinary Triangle of Lévi-Strauss". *Anthropological Linguistics 14(5):* 155–71.

Lehrer, A. (1974). Semantic fields and lexical structure (Bd. 11). Amsterdam [u.a.]: North-Holland Publ.

Lehrer, Adrienne (1993). "Semantic Fields and Frames: Are They Alternatives?" In Lutzeirer, Peter Rolf (ed.): *Studien zur Wortfeldtheorie* [Studies in Lexical Field Theory]. Tübingen: Max Niemeyer Verlag. 149-162.

Leibniz, Gottfried Wilhelm. c. (1702–4). *Table de définitions.* Published in Couturat (ed., 1903).

Leisi, Ernst (1985). Praxis der englischen Semantik (2. Aufl.). Heidelberg: Winter.

Lévi-Strauss, Claude (1965). "Le Triangle Culinaire". *Food & History 2:* 7–20.

Levshina, Natalia (2020). "Conditional Inference Trees and Random Forests". Pp. 611–43 In A Practical Handbook of Corpus Linguistics: 611-643. Cham: Springer International Publishing.

Lipka, Leonhard (2002). English Lexicology: Lexical structure, word semantics and word-formation. Tübingen: Narr

Lohmann, Arne (2013). "Is Tree Hugging the Way to Go? Classification Trees and Random Forests in Linguistic Study". *VIEWS - Vienna English Working Papers 22:*1–16.

Lyons, John (1977). Semantics (1. Aufl.). Cambridge: Cambridge Univ. Pr.

Lyons, John (1996). Semantics. Cambridge: Cambridge Univ. Pr.

Mandelblit, Nili (1995). "The Cognitive View of Metaphor and its Implications for Translation Theory". In Translation and Meaning Part 3, ed. by Marcel Thelen and Barbara Lewandowska Tomasczyk. Maastricht: Hogeschool Maastricht. 483–495.

Marco, Josep. (2019). In PaCor <1., 2016, Santiago de Compostela>. Parallel corpora for contrastive and translation studies: new resources and applications. (I. Doval & M. T. Sánchez Nieto, Eds.) (Bd. volume 90) (39-56). Amsterdam; Philadelphia: John Benjamins Publishing Company. http://doi.org/10.1075/scl.90

McEnery, Tony, and Andrew Hardie (2012). Corpus linguistics: method, theory and practice (1. Aufl.). Cambridge [u.a.]: Cambridge Univ. Press.

Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg Corrado, and Jeffrey Dean (2013). "Distributed Representations of Words and Phrases and Their Compositionality". In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2, NIPS'13*. Red Hook, NY, USA: Curran Associates Inc. 3111–3119.

Miller, George A., and Walter G. Charles (1991). "Contextual Correlates of Semantic Similarity". *Language and Cognitive Processes 6(1):*1–28. doi: 10.1080/01690969108406936.

Mitchell, T.F. (1975). Principles of Firthian linguistics. London:Longman.

Montague, R. (1970). English as a Formal Language. In: B Visentini et al. (ed.), *Linguaggi nella Società e nella Tecnica. Edizioni di Comunità*, Milan. 189-224.

Newmark, Peter (1988). A Textbook of Translation. London (i.a.): Prentice Hall.

Ogden, C. K., and I. A. Richards (1923). The Meaning of Meaning. London: Routledge & Kegan Paul.

Oster, Ulrike, and Teresa Molés-Cases (2016). "Eating and Drinking Seen through Translation: A Study of Food-Related Translation Difficulties and Techniques in a Parallel Corpus of Literary Texts". *Across Languages and Cultures 17:*53–75. doi: 10.1556/084.2016.17.1.3.

Palmer, Harold E. [1938] Reprint (1968). A Grammar of English Words. One thousand English words and their pronunciation, together with information concerning the several meanings of each word, its inflections and derivatives, and the collocations and phrases into which it enters. London: Longman.

Porzig, Walter (1934). "Wesenhafte Bedeutungsbeziehungen." *Beiträge zur Geschichte der deutschen Sprache und Literatur (PBB)* 58

Pottier, Bernard (1964). "Vers une sémantique moderne". *Travaux de linquistique et de littérature 2:* 102-137.

Pottier, Bernard (1965). "La definition sémantique dans les dictionnaires". *Travaux de linquistique et de littérature 3:* 33-39.

Pustejovsky, James (1995). The generative lexicon. Cambridge, Mass., London: MIT Press.

Pustejovsky, James, Anna Rumshisky, and Patrick Hanks (2004). "Automated induction of sense in context." In *COLING 2004 Proceedings.* Geneva.

Putnam, Hilary (1975a). "The meaning of 'meaning." In Putnam (1975b).

Putnam, Hilary (1975b). "Mind, Language, and Reality". *Philosophical Papers. 2 vols*. Cambridge University Press.

Rosch, Eleanor (1973a). "Natural categories." *Cognitive Psychology 4 (3).*

Rosch, Eleanor (1973b). "On the internal structure of perceptual and semantic categories." In Timothy E. Moore (ed.), *Cognitive Development and the Acquisition of Language.* Academic Press.

Rosch, Eleanor (1975). "Cognitive representations of semantic categories." *Journal of Experimental Psychology 104 (3):* 192-233.

Rubenstein, H., & Goodenough, J. (1965). "Contextual correlates of synonymy." Communications of the ACM, 8, 627–633.

Ruppenhofer, J., Ellsworth, M., Schwarzer-Petruck, M., Johnson, C. R., & Scheffczyk, J. (2016). "FrameNet II: Extended theory and practice." International Computer Science Institute.

Russell, B. (1905). "On denoting". *Mind*, *14*(56), 479-493.

Rychlý, Pavel. (2008). "A Lexicographer-Friendly Association Score." In *Proceedings of Recent Advances in Slavonic Natural Language Processing*, edited by P. Sojka and A. Horák. Masaryk University, Brno: RASLAN.

Trier, Jost (1931). Der deutsche Wortschatz im Sinnbezirk des Verstandes [The vocabulary of German in the sense area of conceptualization]. PhD thesis, University of Bonn.

Trier, Jost (1934). "Das sprachliche Feld: Eine Auseinandersetzung." *Neue Jahrbuch für Wissenschaft und Jugendbildung 10.*

Sánchez Cárdenas, Beatriz. (2016). Corpus analysis and the translation of advebs in specialized texts: Raising student awareness. In Corpas Pastor, G., & Seghiri (Eds.). *Corpus-based approaches to translation and interpreting: from theory to applications (Bd. Band 106).* 195-217. M. Frankfurt am Main; New York: Peter Lang.

Sapir, Edward (1929) [pub. 1960]. Culture, Language and Personality. University of California Press.

Sassenhagen, Jona, and Christian J. Fiebach (2019). "Traces of Meaning Itself: Encoding Distributional Word Vectors in Brain Activity." BioRxiv. doi: 10.1101/603837.

Saussure, Ferdinand de (1916). Cours de linguistique générale. (C. Bally, Hrsg.). Paris: Payot.

Schank, R.C., and Abelson, R.P. (1977). Scripts, goals, plans, and understanding: An inquiry into human knowledge structure. Hillsdale, NJ: Lawrence Erlbaum Associates.

Schweinberger, Martin (2021). Tree-Based Models in R. Brisbane: The University of Queensland. url: https://slcladal.github.io/tree.html (Version 2021.02.09).

Sinclair, J. (1987). Looking up: an account of the COBUILD Project in lexical computing and the development of the Collins COBUILD English Language Dictionary.

Sinclair, John (1991). Corpus, concordance, collocation. Oxford [u.a.]: Oxford Univ. Press.

Sinclair, John (1998). "The lexical item." In Edda Weigand (ed.), *Contrastive Lexical Semantics.* John Benjamins.

Sinclair, John (2004). Trust the Text: Language, Corpus and Discourse. Routledge.

Stefanowitsch, Anatol, and Stefan Gries (2003). "Collostructions: Investigating the Interaction of Words and Constructions. *International Journal of Corpus Linguistics 8.* doi: 10.1075/ijcl.8.2.03ste.

Steyer, Kathrin. (Hrsg.) (2004). Wortverbindungen - mehr oder weniger fest (Bd. 2003). Berlin [u.a.]: de Gruyter.

Steyer, Kathrin. (2008). 'Kollokationen in Deutschen Wörterbüchern Und in der Deutschen Wörterbuchforschung'. Lexicographica 24(2008): 185–208. doi: doi:10.1515/9783484605336.1.185.

Straube, Annika (2022). *Die Lexik des Leidens im Französischen, Italienischen und Deutschen*. Dissertation (Doctoral dissertation). Neuphilologische Fakultät, Heidelberg. DOI: 10.11588/heidok.00033905

Strobl, Carolin, James Malley, and Gerhard Tutz (2009). "An Introduction to Recursive Partitioning: Rationale, application and Characteristics of Classification and Regression Trees, Bagging and Random Forests." *Psychological Methods 14 (4):* 323–48. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2927982/.

Stubbs, Michael (1996). Text and corpus analysis: computer-assisted studies of language and culture (1. Aufl., Bd. 23). Oxford [u.a.]: Blackwell.

Sun, Tongtong, Yongxing Du, Liyong Zhou, Lijing Niu, Baoshan Li, Yanzhang Gong, and Jianmin Dong (2021). "Word Sense Disambiguation Method Based on Graph Model and Word Vector." *Journal of Physics: Conference Series,* 2024. doi: 10.1088/1742-6596/2024/1/012061.

Tognini-Bonelli, E. (2001). Corpus linguistics at work (Bd. volume 6). Amsterdam; Philadelphia: John Benjamins Publishing Company. http://doi.org/10.1075/scl.6

Vilgis, Thomas (2013). "Texture, Taste and Aroma: Multi-Scale Materials and the Gastrophysics of Food." *Flavour 2.* doi: 10.1186/2044-7248-2-12.

Wadud, Md. Anwar, M. Ph. D., and Mohammad Motiur Rahman (2022). "Word Embedding Methods for Word Representation in Deep Learning for Natural Language Processing." *Iraqi Journal of Science,* 1349–61. doi: 10.24996/ijs.2022.63.3.37.

Weinreich, Uriel (1966). "Explorations in semantic theory." In Thomas Sebeok (ed.), *Current Trend in Linguistics 3,* 395-477. The Hague: Mouton.

Weisgerber, Leo (1927). "Die Bedeutungslehre- ein Irrweg der Sprachwissenschaft?" *Germanisch-Romanisch Monatsschrift 15:* 161-83.

Weisser, Martin (2016). Practical corpus linguistics: an introduction to corpus-based language analysis. Chichester: Wiley Blackwell.

Whorf, B.L. (1956). *Language, thought, and reality: Selected writings.* Technology Press of Massachusetts Institute of Technology: Cambridge, Mass.

Wurm, Andrea (2007). Translatorische Wirkung: Ein Beitrag Zum Verständnis von Übersetzungsgeschichte Als Kulturgeschichte Am Beispiel Deutscher Übersetzungen Französischer Kochbücher in Der Frühen Neuzeit. Vol. 17. 1st ed. Frankfurt am Main [u.a.]: Lang.

Corpora

EANC

East Armenian National Corpus (2007-2009). http://eanc.net/ (Last retrieved on 13.06.2024)

SKETCHENGINE WEB CORPORA

TenTen Corpus Family (o.J.). Hrsg.: Lexical Computing CZ s.r.o.
URL: https://www.sketchengine.eu/documentation/tenten-corpora/. (Last retrieved: 13.06.2024).

German Web corpus 2013

deTenTen: Corpus of the German Web (o.J.). Hrsg.: Lexical Computing CZ s.r.o. URL: https://www.sketchengine.eu/detenten-german-corpus/. (Last retrieved: 13.06.2024)

English Web corpus 2015

enTenTen: Corpus of the English Web (o.J.). Hrsg.: Lexical Computing CZ s.r.o. URL: https://www.sketchengine.eu/ententen-english-corpus/. (Last retrieved: 13.06.2024)

NLP and Language Analysis Tools

R Core Team (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Scott, M. (2020). *WordSmith Tools version 8,* Stroud: Lexical Analysis Software.

Dictionaries

Aġayan, Ēdoward Bagrati Աղայան, Էդուարդ Բագրատի (1976). *Արդի հայերենի բացատրական բառարան* [Ardi hayereni bac’atrakan baṙaran] online, Yerevan: “Hayanstan”. URL: http://nayiri.com (Last retrieved: 13.06.2024)

Antidote 9 (2016). (Version 3) [Software]. *Druide informatique*, Montreal.

Baratyan, N.R. (2011). *English-Armenian Dictionary.* Yerevan: YSU Press.

Deutches Universalwörterbuch online. DUDEN. 6. Auflage. Mannheim. Leipzig. Wien. Zürich: Dudenverlag. URL: www.duden.de/woerterbuch . (Last retrieved: 11.06.2024)

Deuter, M. (Hrsg.). (2003). *Oxford collocations dictionary for students of English* (3. ed.). Oxford [u.a.]: Oxford University Press.

Leibniz-Institut für Deutsche Sprache: "Wörterbuch zur Verbvalenz". Grammatisches Informationssystem grammis.
DOI:10.14618/evalbu
Permalink: https://grammis.ids-mannheim.de/verbvalenz

Oxford English dictionary online (2020). Oxford University Press. URL: https://www.oed.com/. (Last retrieved: 01.06.2024)

Wörterbuch der deutschen Gegenwartssprache (1964–1977), retrieved via *Digitale Wörterbuch der deutschen Sprache.* URL: https://www.dwds.de/d/wb-wdg. Last retrieved: 24.06.2024.

Data Management and Appendices

Digital Appendix in HeiDATA

All annotated data frames (Armenian, English, German) as well as all graphics and visualizations *(M(CA), Mosaic-Plot, CCCGs, CIT)* in this work have been published in high resolution in HeiDATA institutional repository for Open Research Data from Heidelberg University. The appendix can be accessed and downloaded through the HeiDATA dataset "Data for the PhD thesis "Modeling Lexical Fields for Translation: a Corpus-Based Study of Armenian, German, and English Culinary Verbs""

DOI : https://doi.org/10.11588/DATA/3MPL7E

Digital Appendix in HeiBOX

Appendices 1-12

Link: https://heibox.uni-heidelberg.de/d/26f363df032a438e9996/

Appendices 1-12, accessible via the aforementioned link, redirect the reader to the respective folders in Heibox cloud storage server of Heidelberg University. Each Appendix is numbered and titled accordingly, with options for 'preview' and 'download' modes.

- **Appendix 1**: Manual Semantic Annotation of the German Corpus Occurrences *(deTenTen13 and REZ_DE)*
- **Appendix 2**: Neutralization (Generalization) of Annotation Parameters (German)
- **Appendix 3**: *Context-Conditional Correlation Graphs* (CCCG) in Excel Tables (German)
- **Appendix 4**: *Conditional Inference Trees* (CIT) (German, Armenian, and English Culinary Verbs)
- **Appendix 4.1**: *Conditional Inference Tree* (CIT) (German Synonymous (Culinary) Verb Groups)
- **Appendix 5**: Manual Semantic Annotation of the English Corpus Occurrences *(enTenTen13 and EN_REZ)*
- **Appendix 6**: Mosaic-Plot Visualization of the Comparison Between the English *enTenTen15 and EN_REZ*
- **Appendix 7**: Neutralization (Generalization) of Annotation Parameters (English)
- **Appendix 8**: Neutralization (Generalization) of Annotation Parameters (Armenian)
- **Appendix 9**: Neutralization (Generalization) of Annotation Parameters for Translation Purposes
- **Appendix 10**: Reduction of Annotated Examples for Translation
- **Appendix 11**: *Context-Conditional Correlation Graphs* (CCCG) Excel Tables Based on Language Pairs for Translation Purposes
- **Appendix 12**: Manual Semantic Annotation of the Armenian Corpus Occurrences *(ARM)*