Bachelorarbeit

im Studiengang Computerlinguistik

an der Ludwig- Maximilians- Universität München

Fakultät für Sprach- und Literaturwissenschaften

Creating a Multilingual Gold Standard for Case Marker Extraction

vorgelegt von
Xaver Maria Krückl

Betreuer: Leonie Weißweiler, M.Sc.
Prüfer: Prof. Dr. Hinrich Schütze
Bearbeitungszeitraum: 04. April - 13. Juni 2022

Selbstständigkeitserklärung

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, den 13. Juni 2022

. . . .
Xaver Maria Krückl

Abstract

Case depicts the basic dependencies that hold between constituents of sentences. Whilst this definition mainly describes its syntactic function, case similarly carries a certain semantic meaning as expressed in dependent nominals. In order to grasp such deep cases, the case marking strategies of world languages need to be analyzed. Using a morphological approach to this, case markers are defined as inflectional affixes or whole inflected word forms. This presents a both linguistically and computationally sensible definition that also partly applies to syntactically diverse languages. Based on this, the following thesis creates a multilingual gold standard of case markers that aims to help solving the task of automatically detecting cases in low-resource languages. It is compiled to further evaluate *CaMEL* (Weissweiler et al., 2022), an algorithm for the automated marking of surface cases in parallel corpora. From this, a step towards labeling deep cases, fine-grained semantic categories of nouns in sentences, can be taken. Similarly, differences and similarities between case systems of the analyzed languages emerge as a by-product of the analysis. Following the morphological definition of case markers, multilinguality is restricted in therms of language diversity to a set of inflectional languages. Moreover, the gold standard mainly contains flexion markers following the regular declension patterns of these languages. Yet, it is as detailed as possible to ensure reliability, validity and the possibility to test automated methods with it.

Kasus zeigt die grundlegenden Abhängigkeiten an, die zwischen den Konstituenten eines Satzes herrschen. Während diese reduzierte Definition vor allem die syntaktische Funktion beschreibt, werden durch Kasus gleichzeitig auch semantische Bedeutungen durch abhängige Nominale ausgedrückt. Um solche Tiefenkasus zu erfassen, müssen die Strategien der Kasusmarkierung in den Weltsprachen analysiert werden. Basierend auf einem morphologischen Ansatz sind Kasusmarker flektierende Affixe oder komplette flektierte Wortformen. Dies stellt eine sowohl linguistisch als auch rechnerisch sinnvoll zu verarbeitende Definition dar, die zum Teil auch für syntaktisch unterschiedliche Sprachen gilt. Darauf aufbauend wird in der folgenden Arbeit ein mehrsprachiger Goldstandard von Kasusmarkierern erstellt. Dieser soll helfen, die automatisierte Erkennung von Kasus in ressourcenarmen Sprachen zu verbessern. Genauer soll damit *CaMEL* (Weissweiler et al., 2022), ein Algorithmus zur automatischen Markierung von funktionalem Kasus in parallelen Korpora weiter evaluiert werden. Auf dieser Grundlage können folglich auch Tiefenkasus und damit feine semantische Kategorien von Substantiven in Sätzen markiert werden. Ebenso werden Unterschiede und Gemeinsamkeiten zwischen den Kasussystemen der analysierten Sprachen als Nebenprodukt der Analyse sichtbar. In Anlehnung an die morphologische Definition von Kasusmarkern wird Mehrsprachigkeit in Bezug auf die Sprachenvielfalt auf flektierenden Sprachen beschränkt. Außerdem enthält der Goldstandard hauptsächlich Flexionsmarker die den regelmäßigen Deklinationsmustern dieser Sprachen folgen. Dennoch werden, um Zuverlässigkeit und Validität zu gewährleisten und automatisierte Methoden zu testen, auch alle dafür notwendigen Details miteinbezogen.

Contents

1 Introduction

Case. As one of the most fundamental grammatical categories of natural languages, it depicts basic dependencies between the constituents of sentences. In more detail, *case* marks the type of relationship that dependent nouns or noun phrases bear towards their syntactic head, i.e. the verb of a sentence (Blake, 1994). In order to fulfill this, case is realized in world languages by three main strategies. Apart from adpositions or the word order, *case* is most prominently displayed by inflectional affixes attached to word stems. This morphological approach leads towards the central definition of case markers that is to be used in this thesis. Besides this grammatical function of *case*, Blake (1994) also mentions its power to indicate meaning by representing certain semantic roles that syntactic dependents assume. Regarding terminology, it is referred to deep cases in this context, a notion which emerged from the underlying case grammar approach of Charles Fillmore (see Fillmore (1987) and Cook (1989)). Fillmore proposes a set of relational syntactic-semantic cases as opposed to traditional and functional surface cases. The semantics of deep cases has been specified in several following works as in Nilsen (1972), who used case features for identifying them. In addition, Cook (1989) indicates the usage of case grammar theory and its models in early times of computational linguistics for the purpose of sentence analysis and sentence semantics. Based on this theory, Butt (2006) suggests that by the surface type of case marking, the deeper semantic relationships between dependents and head are similarly indicated. Recently, this finding has been taken up by Grimm (2011) and in more detail by Cysouw (2014), who explicitly draw a connection between the usage of surface case markers and semantic content.

Thus, the main motivation for this work is to make a contribution to an automated and more detailed analysis of surface and thereby also deep cases in multiple languages. In concrete, this thesis should help to further evaluate *CaMEL: Case Marker Extraction without Labels* by Weissweiler et al. (2022), a recent contribution to the topic of case marker extraction in low-resource languages by means of computational morphology. It is aimed to refine the machine-generated silver standard on which *CaMEL* is evaluated with the final product of this work, a manually-generated multilingual gold standard of morphological case markers. A detailed and automatic extraction of these would reduce the step towards grasping semantic meaning via the connected deep cases. Accompanying this, the following analysis will provide a deeper (computational) linguistic insight into the case systems of various languages. Due to the byproduct of summing up the case systems of various languages, possible similarities and differences can be visualized.

As far as recognized, there is no similar work on a multilingual gold standard of case markers until this point. The previously mentioned automatically constructed silver standard builds on humanly-annotated morphological data but was mainly constructed by determining possible case marking suffixes according to frequency calculations (Weissweiler et al., 2022). It is therefore necessary to develop a basic methodology of how case markers are to be extracted manually. For each language, chosen for the gold standard two or more scientific sources, i.e. linguistic grammars, are selected from which the regular pattern of inflectional case marking can be extracted. Then, for each declinable part of speech, these are merged into one comprehensive paradigm displayed in the following by several tables. It is aimed to create a gold standard that is as reliable and valid as possible without drifting into an overly extensive analysis. Therefore, finer grained differences and important special forms that may contribute to the detailed finding of surface and deep cases are added whereas foreign-language words and their specific declensions will be left out.

The following thesis is divided into three main parts. At first, multilingual case marking will be approached theoretically. Beginning with a basic linguistic definition of case concerning its function and a short overview on its abstract semantic meaning, the basics of case systems and cases of world languages are introduced. The findings of these subsections also suggest a closer look into the connection of language typology and case. Based on the overall insights, a both linguistically and computationally sensitive definition of actual case markers is created. This definition merges into the next chapters in which the methodology of extracting case markers used in this thesis is described. The selection of the consulted languages which proceeds according to language families and genera as well as number and kind of case marking behaviour is also justified. Subsequently, the most extensive chapter of the thesis follows. In this, the case markers of the selected languages are practically extracted and collected in a separate paradigms for each inflected part of speech. Finally, after the results have been summarized and classified in the conclusion, an outlook to possible future work on the topic is given.

2 Theoretical Approaches to Multilingual Case Marking

This chapter introduces the theoretical foundations and approaches to multilingual case and its marking. Starting from a general definition which is followed by a short overview on the semantic relations of case, concrete case systems and cases of world languages are set out. Additionally, the connection of case to classical language typology will be outlined. Resulting from this, a linguistically and computationally sensible definition of case markers will be developed. This sets the basis for the practical analysis of creating the multilingual gold standard of case markers.

2.1 A Linguistic Definition of Case

In order to approach the term *case*, it is helpful to initially analyze its denotation. Following the typical western tradition of a term's etymological analysis, *case* derives from Latin *casus* 'fall(ing)' which is on its part a loan translation from ancient Greek *ptõsis* that similarly denotes 'fall(ing)' (Haspelmath, 2012). According to Blake (1994), the term was therefore used to describe the form of words that fall away from their standard imprint. Based on this, ancient Greek philosophers developed a first concept of *case* as a grammatical category. They describe it as a system of word forms that signifies the relationship amongst words in a sentence (Blake, 2012). Until today, this finding has not really changed but further refinements have been made. Thus, *case* is used at present to describe two phenomena. At first, it similarly defines the system in which way dependent nominals are marked for the type of relationship they bear towards their syntactic head (Blake, 1994). It is this basic definition on which not only contemporary literature but also the following understanding of *case* will build. In a second sense, the term also generally qualifies a language to have a case system or rather to be a case language (Haspelmath, 2012). This basic distinction shows that a comprehensive definition of *case* is difficult to create. The concept is better described by several core notions common amongst linguists of various fields (Butt, 2006). Therefore, such key concepts that are necessary for the further elaboration on multilingual case marking are being presented in the following to extend Blake's basic definition.

Most works on *case* initialize a more detailed analysis by going ahead with the classical philological approach (see Blake (1994), Butt (2006) and Malčukov and Spencer (2012)). At first, this serves to refine the basic definition as it indicates case concordance. Case is not only marked on nouns but also on noun's dependents, i.e. in complete noun phrases. In addition, ancient Greek and Latin provide a first answer on the question of what specifically may count as *case*. In these languages, *case* is realized via a system of inflection that is also possibly matched with a set of prepositions (Blake, 2012). By this, concrete grammatical functions are created. Traditionally, case is therefore grasped as a system of inflectional markings; one would talk about *inflectional case*. This approach leads to a morphologically oriented definition as inflection is accomplished by appending morphological affixes (Blake, 1994). However, in several languages like Japanese, only a system of adpositions is used to indicate case functions. Instead of inflection, *case* is concretely marked in Japanese by a larger system of prepositions without a morphological extension of nominals (Blake, 2012). The introductory definition becomes even more imprecise for languages like present day English that mainly use the word order as a third alternative of marking *case*. Admittedly, this only allows a differentiate between syntactic functions like subject as well as direct and

indirect objects from which the specific cases need to be concluded in a further step. At this point, it therefore seems rather insufficient to define the display of *case* by its marking via grammatical morphemes only. However, as Blake (1994) indicates, it has been found that the concrete cases of nominals which were given through the word order are similar to the ones that could have been extracted if they showed morphological inflection. On top of that, morphological case marking is supported by adpositions in several languages. Hence, presenting morphological case marking as the prevailing method of realizing case is justifiable. Overall, without adopting this assumption, this realization of *case* provides an attractive mean to compare languages by distinguishing these three different types of its realization.

2.2 A Brief Overview of Semantic Case Relations

When moving towards the grammatical relations expressed by *case*, this firstly refers to basic syntactic categories like subject or direct and indirect object. Although Blake (1994) indicates the presence of a few dependent nominals in certain languages without a grammatical reference, he classifies these syntactic categories as *case functions*. Consecutively, this serves as a starting point to the phenomenon of *case meanings*. With this, he refers to semantic relations like location or instrument that are similarly expressed by *case*. The first analysis of these abstract semantic roles was presented in the case grammar approach by Charles Fillmore (see for example Fillmore (1987) and (Cook, 1989) for an overview). Initially, Fillmore carved out a set of thematic relations which he labeled alike to terms of traditional cases. However, he then switched to the semantically more transparent terms *agent*, *instrument*, *experiencer*, *object*, *source*, *goal*, *place* and *time* in order to describe syntactic-semantic relational cases which also fit a cross-linguistic analysis (Blake, 1994). Later, the term *deep cases* as being delimited from traditional syntactic *surface cases* has become common. Although Haspelmath (2012) prefers the less confusing term semantic roles instead of deep cases, this is not done in the following in order to maintain a clear differentiation to surface cases. According to a practical analysis of Fillmore's findings, Blake (1994) indicates that for some cases like the instrumental or the locative, their grammatical function and semantic meaning coincide. By contrast, however, other syntactically combined roles like agent and patient need to be isolated by using semantic tests.

Fillmore himself located his approach on a syntactic level, but already suggested the possibility of converting deep structure cases into surface representations by analyzing the means of affixation, suppletion and adpositions (Fillmore, 1987). For the first and partly the third of these vehicles, this implies an examination of morphological case markers, a procedure which is similarly indicated by Butt (2006) and Grimm (2011). A concrete induction of deep cases, respectively semantic roles, has been conducted by Cysouw (2014) though a cross-linguistic analysis of case markers. By using parallel texts, semantic roles can be derived particularly well for languages with prominent case marking. Still, some semiotic issues like *case polysemy* remain. Based on a comparison of the dative in Latin and Turkish, it becomes apparent that a range of meanings may be expressed by case (Haspelmath, 2012). While Grimm (2011) aims to solve this problem via relating case to the more abstract semantic level, the difficulty of *case syncretism* needs to be solved. This denotes the possibility of several cases being indicated by only one marker (Baerman, 2012).

2.3 Case Systems and Cases of World Languages

In order to introduce the different systems and labels that were used and are still common to categorize the comprehensive topic of case, it is useful to start with a classical approach again. Ancient Greek and Latin theorists developed the metaphorical notion *oblique* for such cases that differ from the basic nominative form (Haspelmath, 2012).

This is especially useful for instances in which different formal properties are shared by the nominative and the other cases. Exemplary, this refers to a discrepancy of the nominative stem in Latin or a missing nominative case marker of German nouns. In two-case systems, the term *oblique* case is also used as the opposite to the basic *direct* case. Blake (2012) additionally indicates that oblique cases are often complemented with prepositions, for example to express local meanings in ancient Greek. Even today, this systematic is still used in some grammars to describe phenomena of case.

In his survey on case marking, Blake (1994) introduces global case systems by differentiating the organization of core and peripheral relations. Whilst noun phrases in core relations are often unmarked, those in peripheral relations are classically marked by inflectional case, adpositions or both. The core grammar of a majority of the world's languages is the *accusative system* in which the subject is encoded in the nominative and the indirect object in the accusative. If both cases are indicated, they often show the same marker. Otherwise, the nominative is usually unmarked. This holds true for most of the European languages as well as for Latin. A subordinate role which makes up for around twenty percent of the world's languages is taken over by the *ergative system*. It is mainly used by languages in the Caucasian and Austronesian area of the world. Whilst the subject in the nominative is unmarked, the agent is a marked as the absolutive or ergative. Even though, both preceding systems are distinguished, a mixing is possible and frequent. This is what Haspelmath (2012) refers to when he indicates that both terms are only defined for idealized systems.

When it comes to interpreting core relations, it becomes apparent that they mainly exist due to syntactic reasons with a semantic basis that is only implied. Generally, semantic roles are not exclusively marked on complements but primarily given by the meaning of the predicate. Although Blake (1994) describes the core relation of accusative system as rather discourse pragmatic than semantically motivated, the nominative as the subject clearly describes the agent. Similarly, the accusative as the direct object bears semantic properties like the role of a patient or the effects on itself in case it is non-patient. A semantic basis of the ergative system is rather difficult to detect. Roughly, it is described as an entity that acts on or towards another entity which is not itself, complemented by the absolutive. Besides, a range of patients and a few agents may be encoded.

When it comes to peripheral relations, these are mostly denoted by the dative, the genitive and other local cases. The dative as the case of the indirect object is rather unproblematic (Haspelmath, 2012). As Blake (1994) shows, its denotation as the 'giving case' results from ancient Greek *didonai* and Latin *dare*, verbs that translate to 'giving'. Since cases are not isomorphic across languages, the range that the dative covers in different languages varies. Similarly, it has a larger set of semantic functions that range from roles of purpose to possessor or destination. However, the main case to denote possession is the genitive which is widely spread amongst the world's language families. Its usage as a relative case is only marginal. Regarding local cases like the locative, ablative or allative, the relationship between case marking and semantic function is quite straightforward. Via these, notions concerning location, destination, source or path may be expressed. However, destination may also be taken over by the accusative, as in Latin, or the dative, as in Turkish. Such specific usages need to be analyzed based on concrete languages. Larger case systems as in the Uralic family are usually based on a more detailed elaboration of local cases to which the aspect of relative orientation is added. Eventually, some further marginal cases should be mentioned. The instrumental, which is found in Indo-European and further smaller language families clearly indicates the instrument by which an action is carried out. The comitative case of Finnish, which expresses accompaniment, is also known as the sociative in Dravidian languages. Similarly in Finnish, the abessive, or also privative, denotes the 'lacking' or 'not having' of an object.

Finally, Blake (1994) rises the question of whether there is a concrete hierarchy amongst the presented inflectional cases. From a cross-linguistic perspective on non-isomorphic case

systems, this is practically not possible as no case of one system completely corresponds to the same in the other. However, considering the mere functions of case, a hierarchy emerges for the majority of systems that is quite clear. As core functions, the nominative and the accusative or ergative are followed by the peripheral cases, the genitive and the dative. After these local cases, the ablative or instrumental and finally other cases are situated. If a language has a case to the end of that list, it has at least one of the previous cases as well. Additionally, the lowest case is usually used to fulfil a range of functions.

2.4 Language Typology and Case

Based on the final attempt of the previous section, this passage similarly aims to detect universals amongst languages concerning case. In more detail, the following considers the correspondences between case marking and language typology according to the landmark work of Comrie (1989). Besides several holistic typologies, it is both the word order and morphological typology that contribute the most to this approach. Based on these two, a classical differentiation between *analytic* and *synthetic* languages is drawn. In concrete, it is the word order that dictates the relations between the dependents of a sentence in analytic languages. In their most extreme specification, these languages are referred to as being *isolating*. Ideally, they involve no diverse morphology so that a one-to-one correspondence holds between morphemes and words. Examples of such languages include Vietnamese and present day English. In contrast, the relations between dependents of a sentences are marked via morphological agglutination or inflection in synthetic languages. As being indicated in this defining sentence, this category may be further divided into two sub categories referred to as *agglutinative* and *fusional* languages. Before this further differentiation of synthetic languages is outlined in the following subsections, this passage is concluded by a critical view on the presented classical approach.

From a contemporary perspective, it becomes apparent that the classification is predominantly built on a selection of prototypical members that were be used to create the categories. However, as indicated throughout the theoretical expositions, many languages combine several of criteria when it comes to case. Additionally, Bickel and Nichols (2013) indicate that several variables are conflicted by the traditional classification. Therefore, they propose new criteria by which languages are to be typologized. These include the phonological fuison between case markers and their host words, formative exponence as the number of different categories that may be expressed by a case marker and flexivity which describes allomorphy and further inflectional categories (Bickel and Nichols, 2013). Whilst these new criteria allow for a more precise distinction of languages, it is the traditional approach which is more suitable for this thesis by depicting concrete morphological case marking as the main criterion of linguistic similarities.

2.4.1 Agglutinative Languages

As results from the denotation of 'agglutinating', these languages are characterized by the fact that various affixes are glued on a nominal stem in order to create inflected word forms (Comrie, 1989). As Blake (2012) indicates, it is possible to talk of clear stems, bases and case markers in such circumstances as these forms hardly change. Comrie (1989) furthermore points out to clear cut boundaries between morphemes of which one but mostly several form a word. A classic example of agglutinative languages is Turkish which appends case, number and possession marking suffixes to a nominal stem. Therefore, in such languages, case can be grasped and extracted in its most tangible form. Possible phonologically motivated changes due to different surroundings are still easily recognizable. Eventually, Blake (2012) also refers to the issue of very large case marker inventories that may result from further orientation markers attached to the case suffix in agglutinative languages. In this context, the case affix is to be seen as an infix.

2.4.2 Fusional Languages

Similarly indicated by its denotation, fusional languages are characterized by a fusion of different morphological categories into one word (Comrie, 1989). No clear-cut boundary may be drawn between morphemes and similarly, the definition of stems as well as case markers is more problematic. As for example in Russian, the stem of a noun may coincide with its unmarked form of the nominative singular. However, in another declension class, the nominative singular may also be indicated by a case marker which was previously added to the unmarked form to represent the genitive. Moreover, the genitive marker of this declension class again represents the nominative plural of both classes. As this marker only consists of one letter, it is impossible to segment the morpheme into the two meanings that it expresses. The category case and number are fused into only one inflectional morpheme, a portmanteau morpheme. The example similarly indicates that it is fusional languages in which the problem of case syncretism needs to be dealt with.

2.5 A Linguistically and Computationally Sensible Definition of Case Markers

In the course of the previous sections, a central topic of *case* was repeatedly mentioned but not yet considered in more detail. This last theoretical passage deals with the concrete marking of *case* and aims to develop a both linguistically and computationally sensible definition for this operation. From a most basic point of view, Moravcsik (2012) refers to *case markers* as formal devices associated with nominals or noun phrases that indicate their grammatical role according to the initial definition of *case*. In a further step, she introduces the notion of *segmented case morphemes* for such devices. These case morphemes may be realized as affixes, clitics, via stem modification or suppletion - concepts that are being presented throughout the following passage. At first, this agrees with the broad proposal of Blake (1994) who defines case markers as affixes which can be separated from the stem of inflectable words. Again, he based this morphological approach on the classical ancient Greek and Latin case systems. In these, the case of core functions like subject and object is marked via suffixes. For peripheral relations like location or instrument, additional adpositions may be used as a further option. This overlap between adpositions and inflectional case markers is also generally indicated for further languages by Haspelmath (2012). However, as adpositions are not part of each inflectional system on which this morphological definition builds, they are left out in the following. Additionally, they are not part of noun phrases that are under observation. In this context, the term *case particles* is also rejected by Haspelmath (2012) as being contradictory. Particles are generally defined as words and not as inflectional elements. When it comes to clitics, these elements are closer connected to the nominals whose case they indicate. Although he refers to a certain indeterminacy in this context, clitics are also seen as suffix-like case markers. Overall, Haspelmath (2012) therefore talks of *inflectional case exponents*. In a further step, Blake (1994) highlights that it is not possible to isolate an affix from the basic morpheme of a word, usually the stem, for certain parts of speech of several languages. In these instances, the whole word form marks the case and is therefore seen as a case marker. However, in the following, this will not apply to instances of suppletion as it would be almost impossible to find all root variants of a word.

Overall, the use of morphological affixes emerged as the main type of case marking. This also follows the language typology approach by Comrie (1989) from which it becomes apparent that it is more precisely suffixes that mark case. This is based on the characteristic of agglutinative as well as fusional languages. For the former, however, suffixes may also turn into infixes when they are being followed by another inflectional category (Blake, 2012). A possible fusion of these into a much more comprehensive set of case markers due to further appended orientation suffixes is rejected here due to the characteristics

of agglutination. Additionally, a cross-linguistically survey on the type of case marking in the World Atlas of Language Structures (WALS) shows a great majority of case suffixes as compared to other affixes based on the analysis of nouns (Dryer, 2013). Yet, the somewhat problematic aspect of case syncretism remains when case markers are predominantly seen as suffixes (Baerman (2012) and Blake (1994)). According to Moravcsik (2012), the problem of similar orthographic spelling could be solved by using the category of suprasegmental case markers in which stress is used to indicate case. However, leading over to the last point of defining case markers, stress as a phonological criterion is usually not indicated in text, neither handwritten nor digital. Therefore, when it comes to the computationally sensible aspect that needs to be fulfilled, is is finally affixes, especially suffixes as well as clitics, and complete word forms that are defined as case markers in the following. Additionally, extracting such parts of words or complete words requires only rather basic computational operations.

3 A Methodology for Case Marker Extraction

This chapter shortly describes the concrete methodology by which case markers are extracted in the following to create a gold standard of multiple languages. Additionally, the selection of the analyzed languages is explained and justified.

3.1 Aspects of Manual Case Marker Extraction

Whilst comprehensive language traits concerning case marking have been presented theoretically in the previous chapter, this section explains the handling of language specifics for the following practical extraction of case markers. Most importantly, this describes both the inclusions and simplifications that will be made in the process. Essentially, this multilingual gold standard should be used to test the extracted case markers of various languages by the CaMEL algorithm of Weissweiler et al. (2022). Therefore, it is aimed to treat all languages as similarly as possible to receive a structurally identical product. This should also meet the requirements of a gold standard in terms of reliability, validity and sensibility. Given the following conditions, it is aimed to provide an accurate solution for performing the task.

For each language to be analyzed, at least two grammar books are used of which one clearly shows the morphological case marking. If the second work is not sufficient for checking all forms or if any differences emerged, further literature on the respective language is consulted. In order to ensure traceability of the collected case markers, it is mentioned in detail which grammar provides the basic knowledge and case morphemes as well as which is used to revise them afterwards. For each language, language specifics and the respective parts of speech are outlined in an initial overview. Then, the morphological case markers as defined previously are extracted from regular declension tables. A differentiation in terms of number is essential for fusional languages and therefore similarly conducted for agglutinative ones. Special forms that mostly originate from foreign words are left out whilst it is aimed to capture such that originate from phonetic or register based differences. Furthermore, all indicated more extensive and finer grained markers that result from fusion or joint elements between stem and marker are similarly extracted. This is to allow a more precise distinction of similarly marked cases. In the following tables in the text, affixed case markers are distinguished by a hyphen into pre-, in- and suffixes.

3.2 A Selection of Inflectional Case Marking Languages

In order to describe the dimensions of multilinguality in which the gold standard is to be created, the selection of languages is initially limited to such that are available in the Parallel Bible Corpus (PBC) by Mayer and Cysouw (2014). This parallel corpus is necessary for the *CaMEL* algorithm that builds on the parallel extraction of noun phrases (Weissweiler et al., 2022). Whilst some languages are already excluded in the paper, the following presents further analyses that limits the suitable languages and leads to the final selection. At first, all possible languages have been classified according to their language families and language genera by using their exact ISO encoding in the PBC and the WALS (Dryer and Haspelmath, 2013). Then, if specified, a more detailed analysis

according to their number of cases as described in Iggesen (2013) and the position of their case affixes as examined by Dryer (2013) was conducted. For both articles, however, it has to be mentioned that the focus was placed on nouns only. Based on these analyses, it becomes clear that most suitable languages belong to the Indo-European family. This holds true for Latin as well which is, although being a dead language, initially analyzed in the following practical part. The main reason for this results from its major influence on the emergence of the case (marking) theory. Due to similar arguments, Modern Greek as a Greek language is chosen. Although it lost one of the Ancient Greek cases, it can still be compared to its predecessor. Hereafter, German and Russian are analyzed as being the main representatives in terms of case marking and distribution of the Germanic and Slavic language genera. Whilst there is a certain similarity between all Slavic languages, the other representatives of the Germanic group are becoming less inflectional. This is already reality in Romance languages which do not use any inflectional case marking at all and are therefore not suited for the following analysis. Since the four presented languages are all fusional, it is aimed to incorporate further agglutinative languages into the standard. Therefore, Turkish and Finnish were selected as the main representatives of the Altaic and Uralic language family. Amongst the Turkic language genus, all members show a similar case marking behaviour. Turkish, however, is the most common of these. The situation of Finnish in the Finno-Ugric languages is alike. As Blake (1994) additionally indicates, these languages distinguish the most cases of world languages with around a dozen. This makes them especially interesting for the possibility of analyzing underlying deep cases. Overall, it was aimed to select prominent case marking languages for which a closer connection between case markers and semantic relations is expected. Further alternative languages with a suitable inflectional behavior will be set out in the conclusion

4 Practical Extraction and Collection of Multilingual Case Markers

In this practical chapter, the presented methodology of case marker extraction is applied to the chosen languages. These are structured according to language families. Language specific details are given at the beginning of the respective subsection. Then, for each inflected part of speech, a table containing the paradigm of case markers is created. The final table appears at the appropriate place in the text after its creation has been outlined.

4.1 Indo-European Languages

In classic Indo-European languages, case markers cannot always be linearly separated from a concrete stem (Blake, 1994). This results from their fusional character which is typical for the language family. Therefore, the outline of case marking morphemes as indicated in the considered grammar works will be followed. If no clear stem can be separated or only full word forms are given, these are taken over into the paradigm as such. Blake (1994) states in this regard that stem and case are ultimately united in Indo-European pronouns. Accordingly, the tables of this part of speech will be quite comprehensive.

4.1.1 Latin

In the process of theoretically analyzing case, Latin serves as a classical example of morphological case marking in several basic works. However, (Blake, 2012) also indicates that due to a partially stronger fusion of stems and inflectional endings, the boundary between these two morphemes is not always as clear as theoretically presented. A similar opinion is stated by Embick and Halle (2005) who argue that considering stems in Latin is actually unmotivated and problematic. This becomes practically evident when different Latin grammars are approached. Panhuis (2015) similarly points out to this problem but still tries to solve it according to theoretical morphology. He argues that Latin case markers as grammatical affixes are attached to the stem of nominals as declension endings. Though, when it comes to separating the stem, his argumentation becomes rather cumbersome. In order to recognize this unit, the genitive plural markers *-um* or *-rum* need to be removed from this form. In concrete, the stem of *rosarum* would be *rosa*. However, problems appear when the form of the dative and ablative plural, *rosis*, is compared to this. It becomes quite obvious that the final sound of the stem, which also determines a words declension class, was either lost or replaced by the case marker *-is*. Similar issues occur in the singular, in which *rosa* would be used for the nominative, the vocative and, with a stressed *-a*, also for the ablative. As previously outlined, such fine grained differences cannot be differentiated. Furthermore, this is just as problematic in other declension classes as well.

Therefore, in order to develop a computationally process-able paradigm of Latin case markers, Rubenbauer (1995) is used to clarify the basic morphological foundations of this dead language. According to this rather classical source, inflexible Latin words consist of a usually monosyllabic root, a hardly changed sequence of sounds common to a word family. Together with a sound or sequence of sounds Rubenbauer calls a 'suffix', a stem is built that contains concrete meaning. Finally, by a further sound or a sequence of sounds he refers to as an 'ending', a grammatically complete word is composed. Usually, with root, 'suffix' and 'ending', all three components are put together but a few root words without a

'suffix' also exist. For these special words it is rather problematic to identify the declension class as it is the 'suffix' sound which determines this (Rubenbauer, 1995). In this instance he therefore also refers to the form of the genitive plural in which the sound of a 'suffix' always comes out clearly. According to this, the simplest systematic of showing the Latin inflection is by presenting a word split into its root and a case marker which consist of both parts that Rubenbauer calls 'suffix' and 'ending'. In usual words, when both come together, transformations occur in which a clear separation is no longer possible. As theoretically outlined, a fusion of the case marker takes place. This differentiation is also used in a third grammar of Latin by Touratier (2013) and corresponds to the way in which the Latin declension is taught to students. Therefore, Latin case markers as indicated in these two grammars are considered.

In terms of inflectible parts of speech, nouns, adjectives, numerals and pronouns need to be analyzed. Latin does not know any kind of article (Panhuis, 2015). Usually, the double number of singular and plural is distinguished. Dual forms like *ambo* 'both' decline like respective adjectives or numerals. Latin also distinguishes three grammatical genders. Both categories have a major influence on the shape of the case marker. However, it is only number which needs to be distinguished in the following. In terms of gender, masculine and neuter often coincide.

When it comes to concrete cases, Latin distinguishes between the nominative, vocative, genitive, dative, accusative and the ablative. Of these, the vocative, also called salutation case, is striking. Except for the form of the o-declension class it resembles with the nominative. Particularly interesting in terms of meaning is the merging of the archaic instrumental and the locative that are now expressed by the ablative. Partly, the latter one is also taken over by the genitive (Panhuis, 2015). Additionally, the accusative is used to state the function of destination. In this context it is then governed by prepositions like *in* (Blake, 1994). Correspondingly, analyzing these cases plays a key role to detect semantic relations in the structure of Latin sentences. However, this is complicated by the coincidence of the forms of the dative and ablative plural in all declensions.

Latin Nouns

As mentioned throughout the previous description, Latin nouns are grouped into five different declension classes that are named according to the five stem vowels or, if this ends on a consonant, generally 'consonantal'. This last class is being grouped together with the i-declension. Each declension class also has a certain characteristic in terms of gender. However, this category is summed up in the following paradigm displayed in Table 4.1. In particular, this shows the Latin nominal case markers that follow the regular five declension classes as given in Touratier (2013). All forms are revised by using Rubenbauer (1995) as a second source. Possible markers that build on ancient Greek loan words are left out of the paradigm. It becomes apparent that even though case markers are broadly defined in both grammars, an unmarked form appears for nouns of the consonantal declension in the nominative singular. For all other forms, concrete case markers are provided. The vocative corresponds to the nominative for all declensions except for masculine nouns in the o-declension in which it is *-e*. At a closer look, the relation Rubenbauer (1995) describes as root words that do not have a 'suffix' emerges. For example, the dative or ablative plural *-bus* lacks such a sound that is present in the usual endings *-ibus* and *-ebus*.

Latin Adjectives

In this section, the case markers of Latin adjectives are under observation. Touratier (2013) differentiates two groups of adjectives according to their resemblance to the nominal a- and o- as well as the i- and consonantal declension classes. This is also referred to as the first and the second class, a term that is necessary for describing the case marking

Case	Singular	Plural
Nominative	-a, -us, -um, -is, -s, -e, -o, -u, -es	-ae, -i, -a, -es, -ia, -us, -ua
Vocative	-a, -e, -um, -is, -s, -us, -u, -es	-ae, -i, -a, -es, -ia, -us, -ua
Genitive	-ae, -i, -is, -us, -ei	-arum, -orum, -um, -ium, -m, -uum, -erum
Dative	-ae, -o, -i, -ui, -u, -ei	-is, -ibus, -bus, -ebus
Accusative	-am, -um, -em, -e, -im, -u	-as, -os, -a, -es, -is, -ia, -us, -ua
Ablative	-a, -o, -e, -i, -u	-is, -ibus, -bus, -ebus

Table 4.1: Table of Latin Nominal Case Markers

of pronouns later on. Overall, the adjectives follow the rules and case markers of Latin nouns according to the given declension class. This also includes the vocative. The special marker *-e* does appear due masculine nouns of the o-declension but also because it is an ending of the nominative singular in the i-declension. Whilst members of this second class show two different endings in terms of gender, the first class indicates all three genders. However, as for the paradigm in Table 4.2, these formes are summed up. All forms of the regular declension indicated by Touratier (2013) can also be found in Rubenbauer (1995).

Case	Singular	Plural
Nominative	-us, -a, -um, -is, -e, -s	-i, -ae, -a, -es, -ia
Vocative	-us, -a, -um, -is, -e, -s	-i, -ae, -a, -es, -ia
Genitive	-i, -ae, -is	-orum, -arum, -ium, -um
Dative	-o, -ae, -i	-is, -ibus
Accusative	-um, -am, -em, -e, -s	-os, -as, -a, -es, -is, -ia
Ablative	-o, -a, -i, -e	-is, -ibus

Table 4.2: Table of Latin Adjectival Case Markers

Latin Numerals

When it comes to Latin numerals, these are to be differentiated in cardinal and ordinal numbers (Rubenbauer, 1995). Both are similarly inflected as based on the adjectives of the a- and o-declension. In the following, their paradigm in Table 4.3 is built from these adjectival markers and the concrete forms given in Rubenbauer (1995). Whilst *unus, -a, -um*, 'one', represents the forms of the singular, *duo, -ae*, 'two', and *tres, -ia*, 'three', are added to the plural column. All of these forms are also presented in Panhuis (2015).

Latin Pronouns

This section aims to build a processable paradigm of Latin pronominal case markers. It roughly follows the order given in Touratier (2013) and is being checked on or also possibly expanded by forms given in both Rubenbauer (1995) and Panhuis (2015). Starting with the personal pronouns *ego*, *tu*, *nos*, *vos* and the reflexive forms *se* in the third person, a problem concerning the definition of case markers as suffixes arises. Especially due to the differences in the nominative and dative forms, for example between *ego* and *mihi* in

Case	Singular	Plural
Nominative	-us, -a, -um	-i, -ae, -a, -o, -es, -ia
Vocative	-e, -a, -um	-i, -ae, -a, -o, -es, -ia
Genitive	-i, -ae, -ius	-orum, -arum, -ium
Dative	-o, -ae, -i	-is, -obus, -abus, -ibus
Accusative	-um, -am	-os, -as, -a, -o, -es, -ia
Ablative	-o, -a	-is, -obus, -abus, -ibus

Table 4.3: Table of Latin Numeral Case Markers

the first person, a clear regularity of splitting the personal pronouns into a basic root and suffixing case markers cannot be found. Therefore, all distinct word forms, are added to the paradigm in Table 4.4. Similarly, varied spellings that for example concern the second person plural of the personal pronoun are incorporated. For these, Touratier (2013) still displays the outdated form *uos* in comparison to the modern notation *vos* being used in the two other works. Concerning the ablative, the latter grammars also mention the possibility of the post-position *cum* being attached to personal pronouns. These forms are also added to the table whilst the preposition *a* which can precede the ablative as indicated in Rubenbauer (1995) is left out. As theoretically indicated, prepositions may indicate a certain case but are not part of the inflectable word classes as being examined here.

Comprehensively, the pronominal vocative forms are not displayed in the declension tables of all three grammars; only existing special variants, as for the personal pronouns *tu* or *vos* in the second person singular respectively plural, are given separately (Panhuis, 2015). Therefore, no distinction of the vocative will be made for Latin pronouns. All three grammars also mention the possibility of reinforced personal pronouns with attached particles like *-met*, *-te*, *-pte* or *-se*. As these special forms are not fully described for all cases, they are also left out of the paradigm.

The possessive pronouns *meus, -a, -um*, *tuus, -a, -um*, *suus, -a, -um* in the singular and *noster, -tra, -trum* and *vester, -tra, -trum* in the plural inflect like the previously described adjectives of the a- and o-declension class (Rubenbauer, 1995). Due to this regularity, only these case markers are added to Table 4.4. Rubenbauer (1995) also indicates *mi* in the masculine singular as a further special form of the vocative.

When moving on to the Latin relative pronoun *qui, quae, quod*, 'which', a similar regularity in terms of case marking as with the adjectival declension can be seen. However, the genitive form *cuius* and the dative *cui* in the singular fall off the grid. Therefore, the whole paradigm is added to Table 4.4. The forms of the interrogative determinative *qui, quae, quod* perfectly agree with these. The interrogative pronouns *quis*, 'who', *quid*, 'what', limited to the singular, only vary in these two nominative forms which are therefore added to Table 4.4 as well. At the same time, these forms correspond to the enclitic indefinite pronoun *quis, quid*. However, the enclitic indefinite determinative *qui, quae, quod* shows a new form, *qua*, in the nominative singular as well as the nominative and accusative plural. Having added these forms to Table 4.4, regular endings for the indefinite pronoun *aliquis, aliquid* and the indefinite determinative *aliqui, aliqua, aliquod* could be found in the so far created paradigm on a possible root *ali-*. However, these forms represent a fixed word and not, as needed, a suffixed marker. Therefore, they are added to Table 4.4 marked as suffixes.

A special situation of Latin case marking appears when it comes to a further group of indefinite pronouns respectively indefinite determinatives. When looking at the nominative forms of *quisquam* 'anoyone', *quicumque/quaecumque/quodcumque* 'whoever', *quidam/*

quaedam/quoddam 'some', *quilibet/quaelibet/quidlibed/quodlibet* 'everyone', *quiuis/quaeuis/quiduis/quoduis* 'anyone', and *quisque/quidque/quaeque/quodque* 'everyone', it becomes apparent that their case is marked almost congruent with the indefinite pronoun *quis/quid* respectively the indefinite determinative *qui/quae/quod*. However, this is now realized as a prefix. In order to include this special case to the paradigm, these pronominal forms and some special variants are added to Table 4.4 as prefixes. The indefinite pronoun *quisquam*, 'everyone', exclusively used in negated sentences and the singular, has an additional neuter form *quicquam* used in the nominative and accusative. The accusative singular of *quidam* is *quendam* or *quandam* and the genitive plural similarly *quorundam* or *quarundam*. The dublication of *quis* and *quid*, *quisquis* and *quidquid*, can now be derived from the paradigm as being inflected either by prefixes or suffixes. However, only a further genitive form, *cuiuscuius* is given in (Touratier, 2013). A double special case, *unusquisque*, 'all', is indicated by (Rubenbauer, 1995).

When it comes to the Latin demonstrative pronouns *hic/haec/hoc*, *iste/ista/istud* and *ille/illa/illud*, all 'this', a possible common prefix over all declension forms can be found with *ist-* and *ill-* for the two latter ones. Whilst some endings resemble those of the adjectival declension already in the paradigm, new markers for the nominative, genitive, dative and accusative singular are added. For *hic/haec/hoc*, the question arises whether the letter *h* can be seen as the stem or not. However, since the forms already differ in the nominative, all these pronominal forms are added to Table 4.4. By this, the suffixes of the pronoun *ipse/ipsa/ipsum* based on *ips* can already be found there. This is also due to the fact that, even though it is classified as a pronoun, it can pass into the class of adjectives (Touratier, 2013).

The pronoun *is/ea/id*, 'this/that' can either be used as the non-reflexive form of the personal pronoun in the third person but also as an anaphoric demonstrative pronoun (Touratier, 2013). Again, as no clear root can be separated, its complete forms are added to Table 4.4. In the nominative, dative and ablative plural, *ei/eis* can also be spelled as *ii/iis* or even be more contracted to only *i/is*. The pronoun of identity, *idem/eadem/idem*, 'the same', uses the forms just described almost equally as a prefix by simply attaching *-dem* in every case. Therefore, together with special cases as in the nominative and accusative singular or the genitive plural, they are added as such to Table 4.4.

(Touratier, 2013) concludes with a last group of pronominal adjectives that are partly indefinite pronouns or adjectives. Apart from the genitive and dative singular on *-ius* and *-i*, their other forms correspond again to those of the adjectival declension in the first class. Accordingly, all case markers for *alter, -a, -um* 'other', *uter, -tra, -trum* 'which', *neuter, -tra, -trum* 'none', *alius, -a, -ud* 'another', *nullus, -a, -um* 'none', *totus, -a, -um* 'whole', *solus, -a, -um* 'alone' and *unus, -a, -um*, 'one', are already part of Table 4.4.

Finally, the declension of the pronouns *nemo* and *nihil* is characterized by various special cases in which forms of *nullus, -a, -um* have been adapted. Whilst *nihil* only appears in the nominative and accusative singular, parallel forms of *nemo* and *nullus, -a, -um* are possible in the genitive, dative and accusative. All these special forms are added to Table 4.4 in their full version as being found in Rubenbauer (1995) and cTouratier (2013).

Case	Singular	Plural
Nominative	ego, tu, -us, -a, -um, qui, quae, quod, quis, quid, qua, -quis, -quid, -qui, -qua, -quod, quis-, quid-, qui-, quae-, quod-, quic-, -e, -ud, hic, haec, hoc, is, ea, id, i-, ea-, uter-, utra-, utrum-, nihil, nemo	nos, vos, uos, -i, -ae, -a, qui, quae, qua, -qui, -quae, -quod , qui-, quae-, hi, hae, haec, ei, ii, i, eae, ea, i-, ii-, eae-, ea-, utri-, utrae-, utra-
Genitive	mei, tui, sui, -i, -ae, cuius, -cuius, cuius-, -ius, huius, eius, eius-, utrius-, neminis	nostri, nostrum, vestri, vestrum, uestri, uestrum, -orum, -arum, quorum, quarum, -quorum, -quarum, quorum-, quarum-, quorun-, quarun-, horum, horunc, harum, harunc, eorum, earum, eorun-, earun-, utrorum-, utrarum-
Dative	mihi, tibi, sibi, -o, -ae, cui, -cui, cui-, -i, huic, ei, ei-, utri-, nemini	nobis, vobis, uobis, -is, quibus, -quibus, quibus-, his, eis, iis, is, eis-, iis-, is-, utris-
Accusative	me, te, se, -um, -am, quem, quam, quod, -quem, -quid, -quam, -quod, quem-, quid-, quam-, quod-, quic-, quen-, quan-, -um, -ud. hunc, hanc, hoc, eum, eam, id, eun-, ean- i-, utrum-, utram-, nihil, neminem	nos, vos, uos, -os, -as, -a, quos, quas, quae, qua, -quos, -quas, -qua, quos-, quas-, quae-, -ud, hos, has, haec , eos, eas, ea, eos- , eas-, ea-, utros-, utras-, utra-
Ablative	me, te, se, mecum, tecum, secum, -o, -a, quo, qua, -quo, -qua, quo-, qua-, hoc, hac, eo, ea, eo-, ea-, utro-, utra-	nobis, vobis, uobis, nobiscum, vobiscum, uobiscum, -is, quibus, -quibus, quibus-, his, eis, iis, is, eis- iis-, is-, utris-

Table 4.4: Table of Latin Pronominal Case Markers

4.1.2 Modern Greek

Besides Cypriot Greek, Modern Greek represents the only member of the Greek language genus in the Indo-European family. It derives from ancient Greek, one of the two classical languages on which most of the theoretical foundations of the presented case theory are based. That is why Modern Greek shows inflectional endings that indicate the categories case, number and gender as one suffix, representing a synthetic, in more detail a fusional language (Holton et al., 2016). As for most of these languages, some nouns, pronouns and determiners only exist in the singular or the plural. The majority, however, covers both numbers. From this it becomes apparent that Modern Greek lost the dual. Furthermore, in comparison to ancient Greek, modern Greek only differentiates four cases. These are applied to the parts of speech used in noun phrases, i.e. nouns, adjectives, numerals, pronouns and determiners (Holton et al., 2016). The functions of the dative have mostly been taken over by the genitive and its form resembles the accusative (Metger, 1998). However, the genitive that is used for possession, dimensions and as an attribute is used only infrequently. As in Latin, the vocative is mostly consistent with the nominative that indicates the subject and therefore the agent in a sentence. Next to it, the accusative denoting the patient is the most frequent case.

In the following analysis of case markers, the comprehensive grammar by Holton et al. (2016) takes the key position. Additionally, the short grammar by Metger (1998) is taken for an overview whilst the extracted case markers are reviewed with the help of the more detailed work by Ruge (2001). This does not clearly display case endings but full word forms from which they are to be derived.

Modern Greek Nouns

Instead of clear declension classes like given in Latin, Modern Greek nouns are primarily classified according to their gender and the form of their nominative singular which indicates the correct declension in the most instances (Holton et al., 2016). From this, several further rules regarding the two mentioned criteria may be set up to explain at least a rough system. From this should be highlighted that, with exceptions, the same markers are used in the genitive and accusative singular of masculine nouns and in the nominative and accusative singular and plural of feminine and neuter nouns (Holton et al., 2016).

In order to create the paradigm of Greek nominal case markers, the final and comprehensive declension table in Holton et al. (2016) was used as a basis. The missing forms of the vocative of feminine and neuter nouns have been supplemented according to the information in the text. They correspond predominantly to the nominative forms. Although Holton et al. (2016) specifies that the genitive plural of all Greek nouns ends in -ων, the paradigm in Table 4.5 shows several finer grained markers that are extended to the stem. These can also be found in a more simpler paradigm of Greek nominal declension given in Metger (1998). When the forms were finally reviewed by using Ruge (2001), it becomes clear that the more extensive forms result from alternations in the final sound of the stem. In order to follow the definition of case markers as grammatical morphemes that are added to the stem, these forms were added completely to Table 4.5. By adding these diverse forms it is aimed to refine the extraction of a specific case.

Modern Greek Adjectives

As in ancient Greek and Latin before, Modern Greek adjectives modify nouns and are used either attributively or predicatively (Holton et al., 2016). Thereby, they also need to agree with the noun to be described in case, number and gender. In comparison to German, it is also possible to place the adjective after the noun to give it a special emphasis. If the noun is preceded by an article, this needs to be put before the adjective as well. Due to the inflection for the three possible categories, the paradigm of their case markers in Table 4.6 is correspondingly extensive. Furthermore, since no comprehensive table of the

Case	Singular	Plural
Nominative	-ος, -ης, -ας, -ες, -ους, -εας, -ου, -η, -α, -ω, -ι, -ο, -μα, -ιμο	-οι, -ες, -εις, -ηδες, -αδες, -εδες, -ουδες, -α, -ια, -η, -ματα, -ιματα
Accusative	-ου, -ο, -η, -α, -ε, -εα, -ω, -ι, -ο, -μα, -ιμο	-ους, -ες, -εις, -ηδες, -αδες, -εδες, -ουδες, -α, -ια, -η, -ματα, -ιματα
Genitive	-ου, -ο, -η, -α, -ε, -εα, -ας, -ης, -εως, -ους, -ως, -ιου, -ματος, -ιματος	-ων, -εων, -ηδων, -αδων, -εδων, -ουδων, -ιων, -ματων, -ιματων
Vocative	-ου, -ο, -η, -α, -ε, -εα, -ω, -ι, -μα, -ιμο	-οι, -ες, -εις, -ηδες, -αδες, -εδες, -ουδες, -α, -ια, -η, -ματα, -ιματα

Table 4.5: Table of Modern Greek Nominal Case Markers

adjectival declination is given in all three consulted grammars, the paradigm consists of the manually extracted markers from the example tables of complete word forms in Holton et al. (2016). The grammar follows the order of presenting the adjectives according to their endings in the nominative singular. At first, the category of adjectives on *-ος, -η, -o* which includes numerous words and also follows the nominal declension is given and added to Table 4.6. Adjectives on *-ος, -a/-ıa, -o* only differ from these in the feminine singular. Such new markers are now and in the following always automatically added to the paradigm. This is consequently done for adjectives on *-υς, -ıa, -υ* that have slightly different masculine and neuter forms. Similar to these are adjectives on *-ης, -ıa, -ı* that are spelled slightly different. Adjectives on *-ης, -ϵς* have one set of forms for masculine and feminine but their markers are already part of the paradigm. A further group of adjectives can be classified according to their neuter ending on *-ıκo*. Even though some cases follow the basic nominal declension, it shows various new forms that are consequently added to Table 4.6. A rather small group of adjectives ends on *-ωυ, -ουσα, -ου*. Though, their forms are added to the paradigm as their representatives are frequently used. Finally, Modern Greek also knows some adjectives that do not decline. As in other languages, they are predominantly loan words (Holton et al., 2016). The resulting paradigm is conclusively reviewed by using Metger (1998) and Ruge (2001) as far as possible. While no different forms were found, Ruge (2001) points out that most of the more extensive markers inserted at the end are taken over from ancient Greek participles. In both grammars, they are part of the adjective declension and therefore also in Table 4.6.

Modern Greek Numerals

Modern Greek declines cardinal numerals. Table 4.7 builds on the markers shown by $\epsilon\nu\alpha\varsigma$, $\tau\rho\epsilon\iota\varsigma$ ανδ $\tau\epsilon\sigma\sigma\epsilon\rho\iota\varsigma$ 'one', 'three' and 'four' as presented in Holton et al. (2016). These basic numerals cover all case markers of further higher numbers. However, no distinction between singular and plural is made in all three grammars and no vocative is given. According to their semantic meaning, the singular forms are therefore only covered by $\epsilon\nu\alpha\varsigma$ 'one' whilst the plural forms result from $\tau\rho\epsilon\iota\varsigma$ ανδ $\tau\epsilon\sigma\sigma\epsilon\rho\iota\varsigma$'three' and 'four'. The forms of all the hundreds and 'a thousand' follow the plural declension of Modern Greek adjectives on *-ος* (Holton et al., 2016). Therefore these are also added to Table 4.7. Thousands, millions and billions, however, are written by the cardinal numbers already in the paradigm connected with the plural nouns denoting these sizes. As supported by Ruge (2001), they are consequently not added to the paradigm.

Case	Singular	Plural
Nominative	-ος, -η, -ο, -α, -ια, -υς, -υ, -ης, -ι, -ες, -ας, -ου, -ικο, -αδικο, -ων, -ουσα, -ον	-οι, -ες, -α, -ιοι, -ιες, -ια, -εις, -η, -ηδες , -ικα, -αδες, -ουδες, -αδικα, -οντες, -ουσες, -οντα
Accusative	-ο, -η, -α, -ια, -υ, -η, -ι, -ες, -ου, -ικο, -αδικο, -οντα, -ουσα, -ον	-ους, -ες, -α, -ιους, -ιες, -ια, -εις, -η, -ηδες , -ικα, -αδες, -ουδες, -αδικα, -οντες, -ουσες, -οντα
Genitive	-ον, -ης, -ας, -ιας, -ιου, -υ, -η, -ι, -ονς, -ικον, -αδικου, -οντος, -ουσας	-ωυν, -ιων, -ηδων , -ικων, -αδων, -ουδων, -αδικων, -οντων, -ουσων
Vocative	-ε, -η, -ο, -α, -ια, -υ, -η, -ι, -ικο, -αδικο, -ων, -ουσα, -ον	-οι, -ες, -α, -ιοι, -ιες, -ια, -ηδες , -ικα, -αδες, -ουδες, -αδικα, -ουτες, -ουσες, -ουτα

Table 4.6: Table of Modern Greek Adjectival Case Markers

Case	Singular	Plural
Nominative	-ας, -α	-εις, -ις, -ια, -α, -οι, -ες
Accusative	-α, -αν	-εις, -ις, -ια, -α, -ους, -ες
Genitive	-ος, -ας	-ων

Table 4.7: Table of Modern Greek Numeral Case Markers

Modern Greek Articles

The paradigm of modern Greek definite and indefinite articles is based on the overview given in the three grammars mentioned above. In Table 4.8, the three genera to which they can be declined are again not differentiated but summarized. However, it has to be pointed out that the genus distinction of the definite article also happens in the plural as compared to other languages like German (Ruge, 2001). In terms of case, there are no vocative forms of the Greek article (Ruge, 2001). The Greek direct article is only used with proper nouns, i.e. names of persons, cities or countries and together with some pronouns that are being followed by a noun. No article is used for the main group of predicate nouns and such that describe clear causes (Metger, 1998). Otherwise, the article needs to agree in case, number and gender with the noun it precedes (Holton et al., 2016).

A special case concerning the possbility of expressing the dative is mentioned in Metger (1998). As compared to nouns in which the function of the dative is taken over by the genitive, Modern Greek articles can be used to take over this function through a combination of the preposition σε and the article in the accusative (Metger, 1998). In the case of the definite article, both parts merge to one word form whilst in the indefinite expression, both parts stay separated. This contradicts with the initial definition of morphological case markers as affixes or single word forms. Additionally, the forms are also not mentioned in bot Holton et al. (2016) and Ruge (2001). However, as the indication of a dative function makes an important contribution to extract semantic roles from case markers, these forms are to be added to the paradigm of the dative. Therefore, the preposition is omitted and only the accusative forms as given in Metger (1998) are added to this column in Table 4.8.

Case	Singular	Plural
Nominative	o, η, το, ενας, μια, ενα	οι, τα
Accusative	το, τον, τη, την, ενα, εναν, μια, μιαν	τους, τις, τα
Genitive	τον, της, ενος, μιας	των
Dative	οτο, οτον, οτη, οτην, ενα, εναν, μια	οτους, οτις, οτα

Table 4.8: Table of Modern Greek Articles

Modern Greek Pronouns

In this section, Modern Greek pronouns and determiners are analyzed for their case marking behaviour according to the given order in Holton et al. (2016). Initially, this grammar states that pronouns can stand instead of a whole noun phrase whilst determiners are words that modify nouns but do not belong to the class of adjectives, articles or numerals. Apart from the personal pronoun, most other pronouns can be used like this. Personal pronouns also have another distinctive feature. They can be separated into weak forms which are used like clitics and emphatically used strong forms. Weak forms are fare more frequently but only used in a close connection to other words and even pronounced together with them. The nominative of the first and second person is taken over by the verb and therefore vacant. The genitive forms are commonly used as possessive pronouns. Table 4.9 is filled beginning with those weak forms followed by their strong manifestations. Both are added as full word forms as no clear stem may be separated. Besides the strong form's function as subject or object, their primary usage is to distinguish persons from one another.

With *αυτος, τουτος* ανδ *εκεινος*, Modern Greek uses three demonstratives that may be used as strong pronouns of the third person or as true demonstratives that translate to

'this' and 'that'. Since they are declined like adjectives on *-oς, -η, -o*, these suffix markers are added to Table 4.9. Apart from the weak form in the genitive which is used as the possessive pronoun, Modern Greek knows the emphatic possesive *δικος* which can be both a pronoun as well as a determiner. In its strong form, it is connected with the genitive of the weak form to state a certain emphasis. Overall, however, it also declines like adjectives on *-oς, -η, -o*. Since an alternative set of feminine forms that decline like adjectives on *-oς, -ια, -o* is possible, these markers are also added to the paradigm. When it comes to the three Modern Greek interrogatives, *τι* 'what' is indeclinable and *ποσος* 'how much' declines again like adjectives on *-oς, -η, -o*. Roughly, *ποιος* 'who' also declines like adjectives on *-oς, -α, -o* but shows some minor changes which are added to Table 4.9.

The Modern Greek indefinite pronouns and determiners can be split into specific and non-specific ones. Specific *κατι* 'something' is indeclinable and denotes 'some' when used as a determiner of plural nouns. In contrast to this, *καποιος* 'someone' declines like adjectives on *-oς, -α, -o* and agrees with nouns when it accompanies them. The marker of the special form *καποιον* of the accusative singular has already been added to the paradigm. Similarly, the non-specific indefinites *πιποτα* 'any(thing)' and *πιποτε* 'no(thing)' are indeclinable. However, *κανεις* 'any(one)' and *κανεας* 'no(one)' can be declined. Their markers, which only exist in the singular, are also added to Table 4.9. The basic Modern Greek relative pronoun *που* 'who' is indeclinable and does not differentiate a denotation between humans or non-human objects. Besides this pronoun, relative clauses may also be introduced via the phrase *ο οποιος*. Since the declension of the pronoun *οποιος* follows the one of adjectives on *-oς, -α, -o*, its case markers are already part of the paradigm. Apart from some further indeclinable correlative pronouns, *οποιος, -α, -o* 'whoever' and *οσος, -η, -o* 'as much as' also belong to that class but are declinable. However, their declension again follows the one of those adjectives whose markers are already part of the paradigm. Both may be used as pronouns or together with a noun as determiners.

Of the Modern Greek universal pronouns, Holton et al. (2016) mentions *καθενας* 'every one, each one' as being inflected. Again, its case markers are already present in Table 4.9. The same holds true for the intensive pronouns *ιδιος* 'same' and *μονος* 'only, alone' as well as for the contrastive pronoun *αλλος* 'other, next' that follow adjectival declensions. Finally, Holton et al. (2016) also mentions several quantifiers like *ολος* 'all' and *ολοκληρος* 'the whole' that all decline like adjectives or demonstratives and are therefore also part of the final paradigm in Table 4.9. As far as possible, this was revised with the given pronouns and determiners in Ruge (2001).

Case	Singular	Plural
Nominative	τος, τη, το, εγω, εσυ, αυτος, αυτη, αυτο, -ος, -η, -ο, -ια, -α, -ις, -ας , -να	τοι, τες, τα, εμεις, εσεις, αυτοι, αυτες, αυτα, -οι, -ες, -α
Accusative	με, σε, τον, τη, την, το, εμενα, εσενα, αυτον, αυτη, αυτην, αυτο, -ο, -η, -ια, -ν, -ον, -α, -αν, -να, -ναν	μας, σας, τους, τις, τες, τα, εμας, εσας, αυτους, αυες, αυτα, -ους, -ες, -α
Genitive	μου, σου, του, της, εμενα, εσενα, αυτου, αυτης, αυτουν, αυτην, -ου, -ης, -ιας, -ανου, -ας, -ανης, -νος	μας, σας, τους, εμας, εσας, αυτων, -ων, -ανων

Table 4.9: Table of Modern Greek Pronominal Case Markers

4.1.3 German

One of the biggest language genera in the Indo-European language family is Germanic (Dryer and Haspelmath, 2013). Being eponymous to its parent category, German also represents a prototype amongst its related languages when it comes to morphological case marking. At first glance, as other Indo-European languages use up to nine cases, the German case inventory of only four is rather clear. However, the main reason for that is the power of the dative that takes over several functions that need to be expressed by using, for example, the ablative, the instrumental or the locative in other languages (Hentschel and Vokel, 2009). Apart from that, the dative is mainly used to describe the recipient or the goal. When it comes to the other cases, the syntactic function and semantic meaning they express follows the usual conditions. Whilst the nominative indicates the subject and therefore the argument or agent, the accusative describes the direct object which is the theme or patient as well as a reference on direction. As always, the genitive indicates possession.

Concerning the realization of German cases, the nominative singular as *casus rectus* is unmarked. In most instances, it is referred to as showing the stem of a word (Sahel, 2018). Since this does not hold true for foreign words that have been incorporated into the German language, see Elsen (2014) and the example of *Atlas* whose stem changes to *Atlant-* in the plural, Eisenberg (2020) introduces the notion of a 'basic form' of a word that is contrasted with stems. To basic forms, suffixes are attached in an agglutinating manner whilst the stem leads to a fusional behaviour of case marking. From this, it becomes apparent why German is characterized as a mixed form between these two categories of classical language typology. However, the usage of a root as the basic morpheme to which case is suffixed is less common. Mostly, case markers are attached to the stem represented by the nominative singular as typical for fusional languages. In the following, this very fine grained morphological distinction will not have a major impact, though. Since the essential morphological mean of indicating case is done via suffixation, this fits well into the (computational) linguistic definition of case markers that is followed (Sahel, 2018). In more detail, this holds true for the main German inflectional parts of speech, namely nouns, adjectives, numerals and some pronouns (Eisenberg, 2020). Of several other pronouns and the articles, no clear stem is shown which is why the whole word is considered as a case marker in these instances.

German Nouns

In Table 4.10, the case markers of German nouns are summed up to a paradigm of the regular German nominal declension. Foreign words have been left out of this collection as their declension is mostly irregular. Similarly, as the stem alternation to umlauts in the plural of several nouns does not affect their case marking, this is excluded as a process purely concerned with number (Sahel, 2018). Nevertheless, the case markers of German nouns needs to be considered separately between singular and plural (Thieroff and Vogel, 2012). Further differences occur according to the gender of German nouns by which rough declension classes are indicated. In the singular, feminine nouns do not mark any case. Similarly, the masculine singular endings of accusative and dative correspond to each other apart from a special marker that needs to be mentioned for the dative. In this, the suffix *-e* may be used for non-feminine nouns that do also not end on a vowel. This form is optional and becomes more and more outdated. Nevertheless, it is used in classical writings and therefore was added to the paradigm in Table 4.10. This follows the concretely extracted case markers that are initially based on the description of German flexion in (Thieroff and Vogel, 2012) and then refined by (Eisenberg, 2020). The genitive singular on *-s* results from proper names whilst the most prominent plural suffix is, together with its variants, *-n*.

When looking at the whole paradigm, it becomes apparent that it consists of several

fields of syncretism. It is quite clear as not many different forms of case markers are to be distinguished. Sahel (2018) therefore denotes the German nominal case as rudimentary when it is compared to other languages. Eisenberg (2020) indicates that the nominal declension is actually declining. This shows that surface cases and therefore also deep cases cannot be recognized clearly by using the case markers of nouns only. It is necessary to extract those of their attached articles and pronouns as well.

Case	Singular	Plural
Nominative		-e, -er, -en, -n, -s
Genitive	-s, -n, -es, -en, -ns	-e, -er, -en, -n, -s
Accusative	-en, -n	-e, -er, -en, -n, -s
Dative	-e, -en, -n	-n, -en, -ern, -s

Table 4.10: Table of German Nominal Case Markers

German Adjectives

German adjectives are either prepended to the nouns they describe attributively or they stay on their own in predicative and adverbial use (Thieroff and Vogel, 2012). However, it is only in the first of these instances that adjectives inflect for gender, number an case. Since their gender is only differentiated in the singular, all distinct forms according to this category will be summed up in in the following paradigm of adjectival case markers. A similar condition to which adjectives are differentiated is also summarized. This refers to their possibly strong or weak inflection that results from the definiteness of the preceding article or pronoun. In order to collect the concrete case markers of German adjectives, the presented tables in Thieroff and Vogel (2012) are again used as a primary source. It becomes apparent that the adjectival case markers are regularly appended to their stem. When reviewing them in Eisenberg (2020), it soon becomes apparent that all forms are similar. As for the nouns before, their paradigm in Table 4.11 is quite clear. Additionally, Thieroff and Vogel (2012) point out to several syncretism fields that can be found. Most of these refer to gender but, for example, the markers of the genitive singular also correspond exactly to those of the plural. Finally it should be indicated that there is also adjectives that are used attributively but are not inflected. For these, a zero marking would be conceivable.

Case	Singular	Plural
Nominative	-e, -er, -es	-e, -en
Genitive	-er, -en	-er, -en
Accusative	-e, -es, -en	-e, -en
Dative	-en, -em, -er	-en

Table 4.11: Table of German Adjectival Case Markers

German Numerals

When it comes to German numerals, this inflectable word form is not explicitly mentioned in the previously consulted grammars. Eisenberg (2020) only states in a side note that some numerals take adjectival forms. However, most of them are not inflected as has been described for some attributively used adjectival forms before as well. Elsen (2014)

makes this more clear by classifying both cardinal and ordinal numbers as adjectives. Additionally, forms like *drittel*, 'third', to which she refers to as number adjectives are not at all or only rarely declined. Therefore, no extra paradigm is created for these. The dative ending *-en* as well as *-e* for the nominative and accusative that is used with numbers is also already covered by the adjectival paradigm.

German Articles

Traditionally, German grammar differentiates two kinds of articles, a definite form *der, die, das* 'the' and an indefinite form *ein, eine* 'a'. Since the latter form historically results from the numeral 'one', it lacks a plural form. Today, its main function is the introduction of a new referent or object to the discourse. The definite article is then used to directly refer to this. Its paradigm is equivalent to the demonstrative *der, die, das* which will be analyzed in the following. In colloquial usage, the article becomes reduced and often merges in this form with several prepositions. When it comes to the meaning of articles, Eisenberg (2020) states that the noun or noun phrase is actually almost depends on an article. Apart from syntactical issues, this also refers to the strength of the article to show the case more precisely. In a morphological discussion, Eisenberg (2020) leaves it open to see the letter *d* as the stem of the definite article. It has been decided not to follow this idea so that the final paradigm of articles consists of whole word forms as case markers. This can be found in Table 4.12. As before, the forms were taken from the two basic works Thieroff and Vogel (2012) and Eisenberg (2020).

Case	Singular	Plural
Nominative	der, das, die, ein, eine	die
Genitive	des, der, eines, einer	der
Accusative	den, das, die, einen, ein, eine	die
Dative	dem, der, einem, einer	den

Table 4.12: Table of German Articles

German Pronouns

In comparison to the rather clear paradigms of nominal and adjectival case markers, German shows most of its originally much richer inflectional behavior when it comes to pronouns (Hentschel and Vokel, 2009). Based on the order pursued in Thieroff and Vogel (2012), which group pronouns into classical ones and such that are similar to articles, this quality is being captured in the following by extracting the concrete pronominal case markers towards a paradigm displayed in Table 4.13.

Thieroff and Vogel (2012) start off by explaining the declension of German demonstrative pronouns like *dieser*, 'this', or *jener*, 'that'. As clear stems like *dies-* and *jen-* can be separated, these represent a prototypical pronominal flexion-pattern that is based on the declension of strong adjectives. Except for the genitive singular marker *-es* that is possible in masculine and neuter, the markers are identical. For certain words like *diesen Tages*, *-en* is possible in the genitive singular as well. All these markers are initially added to the paradigm. When looking at two further demonstrative pronouns, *derjenige*, 'the one' and *derselbe*, 'the same', a special feature arises concerning these composed words. Whilst the word ending is inflected like a weak adjective, the prefix *der-, die-, das-* is also inflected. This phenomenon is called internal flexion, *Binnenflexion* in German (Thieroff and Vogel, 2012). It contradicts with the initial statement of German case markers being almost exclusively realized as suffixes and presents a unique example of German prefixes. Consequently, both suffix and prefix forms are part of Table 4.13. At a closer look, these

prefixes also resemble the adnominal forms of the controversially debated demonstrative pronouns *der, die, das*, 'the' (Thieroff and Vogel, 2012). The debate on these is mainly based on their adnominal forms in which they are equal to the German definite article. Yet, when using them as pronominals, their forms of the Genitive singular *dessen, derer, deren* and plural *derer, deren* as well as the Dative plural *denen* differ from the corresponding articles in these cases (Thieroff and Vogel, 2012). Therefore, the forms of this demonstrative pronoun are also part of the pronominal paradigm.

When Thieroff and Vogel (2012) move on to indefinita and distinguish them from indefinite pronouns, the only relevant point concerning their case markers is their equivalent inflection as compared to the demonstrative pronouns or to adjectives. In more detail, indefinita like *aller, jeder*, 'all/everyone', or *mancher*, 'some' are inflected like *dieser* whereas *beide*, 'both' or *deartig*, 'such', are inflected like adjectives. (both strong and weak forms) Additionally, it has to be mentioned that of this set, *jeder* only exists as a singular word and *mehrere*, 'several', only in plural (Thieroff and Vogel, 2012).

When it comes to possessive pronouns *mein, dein, sein, ihr, unser, euer, ihr*, all possible suffixes on these nominative stem forms are already present in 4.13. They are again consistent with the previously mentioned forms of *dieser*. However, it has to be mentioned that they are unmarked in several cases like in the nominative or accusative singular (Thieroff and Vogel, 2012). The interrogatives *welcher*, 'which one', and *mancher*, 'some', are inflected like a demonstrative pronoun again (Thieroff and Vogel, 2012).

The German personal pronouns can be split into first, second and third person with the latter one additionally inflecting in the singular by gender (Thieroff and Vogel, 2012). Since there cannot be found any joint stem over all forms, these are all added to Table 4.13. Concerning the class of reflexive pronouns Thieroff and Vogel (2012) only mentions *sich*, 'itself' as a representative that does not change its form in the dative and accusative. In the genitive, it is represented by the possessive pronoun-.

Thieroff and Vogel (2012) clearly differentiates indefinite pronouns from the previously mentioned indefinita. The former only only occur autonomous and in singular. Concerning their inflexion form they are split into two groups according to their gender as masculine or neutral (Thieroff and Vogel, 2012). Whilst masculine forms like *jedermann*, 'anyone', have a genitive on *-s* or *-es, jemand*, 'someone', or *niemand*, 'noone', do also have a pronominal usage with case markers *-en* in the accusative and *-em* in the dative (Thieroff and Vogel, 2012). Neutral indefinite pronouns do not change. All these forms are part of the paradigm in Table 4.13

When it comes to interrogative pronouns, Thieroff and Vogel (2012) defines *wer*, 'who', as a inflectable masculine and *was*, 'what', as a non-inflectable neutral even though this basic form can also resemble an accusative. For *wer*, each case has a special yet often disputed form. In the genitive, it is *wessen*, in the accusative *wen* and in the dative *wem* (Thieroff and Vogel, 2012). Since these pronouns show no clear paradigm, they are not added to Table 4.13. Finally, German relative pronouns are under observation. For *der, die, das*, 'this', all forms are already part of the paradigm. The same applies to *welcher, welche, welches*, 'which one', whose markers resemble those of the demonstrative pronouns.

In order to verify the final paradigm, these pronominal case markers have been reviewed by using Eisenberg (2020).

Case	Singular	Plural
Nominative	-er, -es, -e, der-, das-, die- der, das, die, ich, du, er, sie, es	-e, -en, die-, die, wir, ihr, sie
Genitive	-es, -er, -en, -s, des-, der-, des, der, dessen, derer, deren, meiner, deiner, seiner, ihrer	-er, -en, der-, der, derer, deren, unser, euer, ihrer
Accusative	-en, -es, -e, den-, das-, die-, den, das, die, mich, dich, ihn, es, sie,	-e, -en, die-, die, uns, euch, sie
Dative	-em, -er, -en, dem-, der-, dem, der, mir, dir, ihm, ihr	-en, den, denen, uns, euch, ihnen

Table 4.13: Table of German Pronominal Case Markers

4.1.4 Russian

Next to Germanic, it is the Slavic language genus that represents several fusional languages of the Indo-European family (Dryer and Haspelmath, 2013). Out of these that are all similar amongst each other, Russian is selected as a further prototypical example for the extraction of multilingual case markers. This is mainly based on its large spatial and speaker related distribution. Furthermore, its Cyrillic script also differs to the Latin and Greek ones before, this language is more interesting than Polish, for example.

Initially, a basic morphological overview of the Russian word form theory is provided by Kohls (2009). An indispensable root may be complemented with a prefix, a suffix, with both or neither of these affixes to build a stem or, put differently, a word. By appending a set of case endings to inflectable words like nouns, adjectives, numerals and pronouns, the final grammatical forms of Russian words are constructed. As can be grasped from this, Russian knows neither definite nor indefinite articles as being used for example in German. Instead, this relation is transmitted by the context or the word order (Wade, 2020). The three Russian genders, masculine, feminine and neuter, are dictated by the noun and transmitted to the adjective or pronoun that qualifies for or stands for it. Even though gender only affects word forms in the singular, it is used as a factor to distinguish three Russian nominal declension patterns. Whilst most of the masculine and neuter nouns are part of the first and most feminine nouns belong to the second, the third declension class covers only feminine soft-sign nouns. Concerning this latter distinction as a second factor, the first and the second declension contain both hard- and soft-ending nouns. Other aspects that affect the shape of Russian declension endings are spelling rules and changes of stress. This distinction of case is also indicated by Blake (1994). However, as previously in Latin, stress will not be taken into account in the following subsections. Stress markers are usually not part of written texts and overall not part of the definition of case markers set up for this thesis.

In concrete, Russian differentiates six cases of which the initial four are similar to German. Their meaning is indicated by Wade (2020) as follows. The nominative denotes the subject of a state or an action and therefore the agent. It has also incorporated the vocative and its meaning of addressing persons. The accusative as the object refers simply to the patient. Next to the usual implications of possession and relationship, the Russian genitive may also be used partitively. As the indirect object, the dative refers to the recipient of an action. As indicated by its name, the instrumental describes the instrument. Finally, the prepositional is only used after prepositions. Historically, it emerged from the locative and still fulfills this function.

Russian Nouns

This section describes the collection of Russian nominal case markers as based on the comprehensive English speaking grammar by Wade (2020). The collected forms are being revised and complemented with the help of the German speaking standard grammar by Kohls (2009).

Initially, in terms of word formation, Russian nouns either represent an irreducible stem or a compounded noun (Wade, 2020). In both cases, the different word forms are created by attaching inflectional endings to these shapes. In a first step, the paradigm in Table 4.14 is filled based on a simplified declension chart that shows these markers given in Wade (2020). This shows that both the nominative and accusative singular as well as the genitive plural may be unmarked. In the accusative, Russian also differentiates between nouns that denote animate persons or animals and inanimate objects. According to the gender, it is either the nominative, the genitive or a separate marker that is used. Concerning the paradigm that is to be created, neither animacy nor gender are to be analyzed. Therefore, all possible markers are simply collected. In the following, some alternative and more detailed Russian case markers as indicated by Wade (2020) are added to the paradigm in Table 4.14. At first, the partitive genitive singular of nouns that denote a measurable quantity is expressed using the suffixes *-у* or *-ю*. The same but now stressed suffixes are used by some nouns in the propositional singular as well, denoting a location in combination with certain prepositions like *в* or *на*. Since stress is not taken into account, these are added to the paradigm in unstressed form. This issue show up a few more times but is always treated in this way. The possible instrumental marker *-ою* is also added to the paradigm as a literary form often found in poetry. Furthermore, the genitive plural *-ей* can also only be *-й*. Throught this, a seldom change of some nouns from *-ей* to *-ай* is also covered. Both forms can also be found in Kohls (2009).

In general, the paradigm in Table 4.14 is supplemented with several further endings given in this second grammar. Being described as more learner-friendly, it specifies further and more finely differentiated endings in contrast to the simplified declension table and the separately added forms in Wade (2020). In order to accurately distinguish different cases, all forms, although they are only phonologically expanded, have been included in the final paradigm.

Case	Singular	Plural
Nominative	-й, -ъ, -о, -е, -ё, -а, -я, -мя	-ы, -и, -а, -я, -ъя, -е, -ата, -ята, -ена
Accusative	-й, -ъ, -о, -е, -ё, -а, -я, -мя, -ы, -и, -у, -ю, -ени	-ы, -и, -а, -я, -ов, -ев, -ей, -ъя, -е, -ата, -ята, -ена, -й, -ёв, -ён
Genitive	-а, -я, -ы, -и, -у, -ю, -ени	-ов, -ев, -ей, -й, -ёв, -ён
Dative	-у, -ю, -е, -и, -ени	-ам, -ям, -енам
Instrumental	-ем, -ом, -ём, -ей, -ой, -ёй, -ъю, -ою, -енем	-ами, -ями, -енами
Prepositional	-е, -и, -у, -ю, -ени	-ах, -ях, -енах

Table 4.14: Table of Russian Nominal Case Markers

Russian Adjectives

When approaching Russian adjectives, a differentiation between two basic kinds is made. Long adjectives are mostly used as attributes preceding the noun. As in other Indo-Eurpoean languages, they agree with this noun in gender, case and number. While some adjectives only exist in a long form, most have both a long and a short one. However,

this short form is only used predicatively and therefore not of interest in terms of case marking (Wade, 2020). Kohls (2009) makes this more clear by stating that short forms are not declined at all. Therefore, Table 4.15 only depicts case markers of long Russian adjectives and follows the order of several declension tables given in Wade (2020).

Initially, the class of adjectives on hard endings, to which they can also be differentiated, is added to the paradigm. Wade (2020) additionally adds the marker *-ою* as a mostly poetic feminine one. For the mixed declension class, which depends on the final consonant of an adjective, their further forms are included in Table 4.15. Finally the same process is executed for soft ending adjectives that also show some new markers. Furthermore, Russian also knows possessive adjectives (Wade, 2020). These can be split into two groups which both show a regularity of case endings. Some also resemble those of the regular adjectives. Therefore, only the differing forms are added to Table 4.15. Finally, it is to mention that there are also some Russian adjectives that are indeclinable.

Overall, the paradigm elaborated from Wade (2020) was reviewed by using Kohls (2009). It was double checked that, apart from some distinct endings, most forms of the accusative singular resemble those of the nominative or genitive singular. So far, no differing forms have been encountered.

Case	Singular	Plural
Nominative	-ый, -ая, -ое, -ий, -ее, -ой, -яя, -я, -е, -а, -о	-ые, -ие, -и, -ы
Accusative	-ый, -ого, -его, -ую, -ое, -ая, -ий, -ее, -ой, -юю, -ю, -е, -у, -о	-ые, -ых, -ие, -их, -и, -ы
Genitive	-ого, -ой, -ей, -его	-ых, -их
Dative	-ому, -ой, -ему, -ей, -у	-ым, -им
Instrumental	-ым, -ой, -ою, -им, -ей, -ею	-ыми, -ими
Prepositional	-ом, -ой, -ем, -ей	-ых, -их

Table 4.15: Table of Russian Adjectival Case Markers

Russian Numerals

In this section it is aimed to create a paradigm of Russian numeral case markers that is as comprehensive as possible. It arises from the exemplary declension tables of individual numbers as given in (Wade, 2020) and (Kohls, 2009). In order to make the process traceable, the concrete numbers are being mentioned in the following. At first, Table 4.16 is filled with the case endings of the cardinal numerals *один* 'one', *два* 'two', *три* 'three' and *четыре* 'four' as displayed in (Wade, 2020). The singular markers originate from *один* 'one' in which this masculine form has no nominative case marker. However, the nominative singular feminine and neuter is marked. Both grammars do not indicate clearly whether the forms are singular or plural. Since all the other numerals semantically already indicate plurality, their markers are added to the plural column in Table 4.16.

The numbers *пять* 'five' - *двадцать* 'twenty' as well as *пятьдесят* 'fifty' - *восемьдесят* 'eighty' decline regularly like soft-sign feminine nouns. Therefore, these markers were taken over to this paradigm as well. The missing numbers, *сорок* 'fourty', *девяносто* 'ninety' and *сто* 'hundred' have the ending *-а* in the oblique cases. Additionally, the endings *-ам* for the dative, *-ами* for the instrumental and *-ах* for the prepositional, which are added to the cardinal numbers between two hundred to nine hundred, are included into the paradigm. At this stage, this already includes the case markers of *оба/обе*, 'both'. Furthermore, Russian also knows collective numerals in a series from *двое* 'two' to *десятеро* 'ten' of which the forms above seven are only rarely used. The case markers of these forms given

in Wade (2020) are also added to the plural column of Table 4.16.

When it comes to ordinal numerals, Wade (2020) states that these decline like hard adjectives and are also used like them. Consequently, the respective endings are also included in Table 4.16. When finally revising the paradigm by using Kohls (2009), some more finely grained endings emerge again. These include the genitive marker *-ух* that is also used in the accusative when it denotes animacy. Kohls (2009) also shows the declension of *тысяча* 'tousand', *миллион* 'million' and *миллиард* 'billion' which do have a singular form. These endings are added to the paradigm in Table 4.16 as well.

Case	Singular	Plural
Nominative	-а, -о, -ый, -ая, -ое	-и, -е, -а, -о, -ы, -ые
Accusative	-ого, -у, -о, -ый, -ую, -ое	-и, -их, -а, -е, -ух, -ёх, -ых, -ы, -ые
Genitive	-ого, -ой, -и, -а	-их, -ух, -ёх, -и, -а, -ых, -ов
Dative	-ому, -ой, -е, -у	-им, -ум, -ём, -и, -а, -ам, -ым
Instrumental	-им, -ой, -ою, -ей, -ом, -ым	-ими, -умя, -емя, -мя, -ю, -ью, -а, -ами, -ыми, -ьмя
Prepositional	-ом, -ой, -е	-их, -ух, -ёх, -и, -а, -ах, -ых

Table 4.16: Table of Russian Numeral Case Markers

Russian Pronouns

When analyzing the Russian pronouns for their case markers or rather case forms, the order in which they are presented in Wade (2020) will be followed. At the beginning, the forms of the personal pronouns, *я, ты, он/она/оно, мы, вы, онн* are added completely to Table 4.17. The governing prepositions of the prepositional case are left out as in Latin. Adding these contradicts the basic definition of morphological case markers. Special forms of the instrumental, *мною* and *тебою*, which are used in verse, passive constructions and colloquial texts, or *ею*, which is used in educated speech, are also included to the paradigm (Wade, 2020). With *себя*, Russian knows a reflexive pronoun which refers back to the subject or agent closest to it. Therefore, it does not exist in the nominative. In terms of number, it is both used in the singular and plural (Kohls, 2009). Due to a change of the stem in the instrumental, all complete forms are added to Table 4.17.

When it comes to the Russian possessive pronouns, a regularity can be found for *мой* 'my', *твои* 'your', *наш* 'our' and *ваш* 'your'. Wade (2020) indicates *мо* and *тво* as stems of the first two pronouns to which regular endings are attached. In order to fit the nominative forms into this pattern, *-й* is seen here as a pronominal case marker although it is not clearly indicated as a such in the grammars. However, as it is also a regular noun marker of the nominative, this is justifiable. The stem of the latter two pronouns is the given nominative form. All different endings are added as such to Table 4.17. The possessive pronouns of the third person, *его* 'his'/'its' and *её* 'her'/'its' in the singular and *их* in the plural resemble the genitive form of the personal pronoun and are invariable amongst all cases (Kohls, 2009). Therefore they are added to the paradigm as full forms if they are not yet present. Since the reflexive pronoun *свой, своя, своё, свои* declines like the regular possessive pronouns, its endings are already part of Table 4.17.

For the Russian interrogative pronouns, both grammars provide a declension table of *кто* 'who' and *что* 'what'. As these only exist in the singular, they are added completely to the paradigm of Russian pronouns since no clear stem can be separated. However, as can be seen from its three different forms, the interrogative pronoun *чей, чья, чьё* 'whose' declines for gender and also for number. Due to the same reason as before, all distinct

forms need to be added to the paradigm 4.17. Those pronouns may also be used as interrogatives or relative pronouns together with *какой* 'what' and *который* 'which', that are declined like hard adjectives (Kohls, 2009). Therefore, these markers are also added to Table 4.17.

Regarding the Russian demonstrative pronouns *этот* 'this', *тот* 'that', *сей* 'this' and *экий* 'what a', things are similar as for the relative pronouns before. The forms cannot be traced back to a clear stem which is why all distinct forms were put into the paradigm of Table 4.17. In contrast to Kohls (2009), Wade (2020) mentions the latter form as well but simultaneously describes it as mainly being used in conversational register and becoming archaic.

When looking at the declension of Russian determinative pronouns, it becomes apparent that in the case of *сам* 'self', this masculine nominative singular form can be separated as a stem throughout the whole paradigm. Additionally, most of the flexional endings attached to it are already part of Table 4.17. Those which are not are added together with the form *самоё* which represents a traditional literary form of the feminine accusative singular (Wade, 2020). When it comes to *весь* 'all' or 'the whole', a stem on *вс-* can be found throughout the whole paradigm except for the mentioned form of the masculine nominative and accusative singular. By classifying this change as a stem alternation, additional flexional endings of the further cases are added to the paradigm of Russian pronominal markers as before.

Of the further determinative pronouns, *каждый* 'each' and *самый* 'most' are declined like hard ending adjectives whose markers are already part of the paradigm (Wade, 2020). Since *всякий* 'any' and *всячецкий* 'all kinds' are declined following the mixed declension of Russian adjectives, the markers that differ from the already present hard ending ones are added to Table 4.17 (see also Kohls (2009)).

In Russian, negated pronouns are created via prefixation of the negation particle *ни* which can be translated as 'nor' (Wade, 2020). For *никакой*, 'none', there is no issue as it declines like a hard adjective. However, for *никто*, 'no one', *ничто*, 'nothing', and *ничей*, 'nobody', the full word forms need to be added to the paradigm as was done for the underlying interrogatives.

Finally, for *некоторый*, 'some', all hard ending adjectival case markers are already in the paradigm whilst for *некий*, similarly 'some', several finer grained markers need to be included. Overall, Kohls (2009) was used as a second source throughout the whole process. In this, the presentation of the pronouns was kept simpler.

Case	Singular	Plural
Nominative	я, ты, он, она, оно, -и, -я, -ё, -а, -е, -й, его, её, кто, что, чей, чья, чьё, -ой, -ое, -ая, -ый, зтот, зта, зто, тот, та, то, сей, сия, сие, зкий, зкая, зкое, -о, -ий, никто, ничто, ничей, ничья, ничьё	нас, вас, их, -и, чьи, -ые, -ие, зти, те, сии, зкие, -е, ничьи
Accusative	меня, тебя, его, её, оно, себя, -и, -его, -ю, -ё, -у, -е, -й, кого, что, чей, чьего, чью, чьё, -ой, -ое, -ая, -ый, -ую, -ого, зтот, зтого, зту, зто, тот, того, ту, то, сей, сего, сию, сие, зкий, зкоего, зкую, -оё, -о, -ий, никого, ничто, ничей, ничьего, ничью, ничьё, -оего	мы, вы, оны, себя, -и, -их, их, чьи, чьих, -ые, -йе, -ых, -их, зти, зтих, те, тех, сии, сих, зкие, зких, -е, -ех, -ие, ничьи, ничьих, -оих
Genitive	меня, тебя, его, её, оно, себя, -его, -ей, кого, чего, чьего, чьей, -ого, -ой, зтого, зтой, того, той, сего, сей, зкоего, зкой, -ого, -ой, никого, ничего, ничьего, ничьей, -оего, -оей	нас, вас, их, себя, -их, чьих, -ых, -их, зтих, тех, сих, зких, -ех, ничьих, -оих
Dative	мне, тебе, ему, ей, себе, -ему, -ей, его, её, кому, чему, чьему, чьей, -ому, -ой, зтому, зтой, тому, той, сему, сей, зкоему, зкой, -ому, -ой, никому, ничему, ничьему, ничьей, -оему, -оей	нам, вам, им, себе, -им, их, чьим, -ым, -им, зтим, тем, сим, зким, -ем, ничьим, -оим
Instrumental	мной, тебой, им, ей, ею, мною, тебою, собой, собою, -им, -ей, -ею, его, её, кем, чем, чьим, чьей, чью, -ым, -ой, -им, зтим, зтой, зтою, тем, той, тою, сим, сей, сею, зким, зкой, зкою, -ой, -ою, -ем, никем, ничем, ни-чьим, ничьей, ничью, -оим, -оей	нами, вами, ими, собой, собою, -ими, их, чьими, -ыми, -ими, зтими, теми, сими, зкими, -еми, ничьими, -оими
Prepositional	мне, тебе, нём, ней, себе, -ём, -ей, -ем, его, её, ком, чём, чьём, чьей, -ом, -ой, зтом, зтой, том, той, сём, сей, зком, зкой, -ом, -ой, ником, ничём, ничьём, ничьей, -оем, -оей	нас, вас, них, себе, -их, их, чьих, -ых, -их, зтих, тех, сих, зких, -ех, ничьих, -оих

Table 4.17: Table of Russian Pronominal Case Markers

4.2 An Altaic Language: Turkish

As for the first non Indo-European language, this section depicts the collection of Turkish case markers. In more detail, Turkish belongs to the Altaic language family and is considered the main representative of the Turkic language genus (Dryer and Haspelmath, 2013). Amongst its related languages such as Uzbek, Tatar, Kazakh or Kirghiz, it is the most widespread in terms of number of speakers (Göksel and Kerslake, 2005). Apart from that, Turkish represents a prototypical example of agglutinative languages to which all further Turkic languages equally belong (see Ersen-Rasch (2012) and Dryer and Haspelmath (2013)). Due to its rather simple inflectional structure, the Turkish nominal declension was also used by Blake (1994) to explain the clear separability of case markers as inflectional suffixes from stems in his introduction to *case*.

In the following, Ersen-Rasch (2012) will be used as the primary source for a more detailed analysis of Turkish inflectional morphology. The extracted case marking forms in this grammar will then be revised by consulting Göksel and Kerslake (2005) as a secondary source. Both grammars initially indicate that different nominal word forms are created by appending the respective suffixes of grammatical categories to a predominantly immutable word stem in a fixed order. For Turkish, this sequence is number, possession and case. Both grammars differentiate six Turkish cases. At first glance, apart from the vocative, the case system resembles that of Latin with an additional concrete locative. However, in terms of case meaning, several differences appear. At first, besides being the subject and therefore the agent, the nominative may also represent an unmarked direct object that is closely connected to the verb. In this case it represents the object that is talked about. Specifically, the Turkish object is indicated by the accusative, though. Furthermore, apart from the classical direct object, the dative is used to denote the relation 'where to'. Finally, the ablative implies the denotation 'from where'.

As compared to the previous Indo-European languages, Turkish does not differentiate any grammatical gender. Loanwords that do so in their root language have no influence on case marking. More importantly, no articles are used to indicate the definiteness of nouns. Accordingly, no paradigm of this part of speech can be created. Concerning its orthography, Turkish agrees to a large extent with the Latin script but knows some further special characters of which one, ɪ, appears frequently in the case morpheme.

Turkish Nouns

As described above, the case markers of Turkish nouns are generally to be extracted as the suffixes of the third inflectional category that is added to the nominal stem. Therefore, no further distinction with number has to be made at first glance. However, as Ersen-Rasch (2012) indicates a slightly different variant of the genitive and accusative case marker between singular and plural in a table of nouns inflected for number, possession and case. Consequently, and also to follow the general structure of the gold standard, a distinction between singular and plural endings will be made in the following. In the final paradigm, this will not have an effect, though, as the overall set of markers stays the same.

As signaled by implication, the paradigm in Table 4.18 is based on the general overview as well as on concrete examples given in (Ersen-Rasch, 2012). As a result of the theoretically mentioned phonological conditions, these show more detailed case suffixes after stems on final vocalic sounds. In order to attest completeness to the gold standard, these slightly different forms are all added to the paradigm. Similar to German, the nominative is unmarked by representing the infinitive form. A special situation appears with proper nouns. The case suffix is separated from the stem or the previous suffixes by an apostrophe and therefore represents a clitic (Ersen-Rasch, 2012). Yet, as the inflectional morpheme also stays the same, no differentiation is made for these cases. The final paradigm in Table 4.18 agrees with the given forms in (Göksel and Kerslake, 2005). This grammar also mentions a further inflectional marker of nominals that indicates an instrumental, comi-

tative or conjunctive meaning. It may be attached to noun phrases like dative, locative or ablative markers which thereby function as oblique objects. However, *-la, -yla, -le* are unstressed and therefor not seen as clear case suffixes (Ersen-Rasch, 2012).

In a further step, copula and person markers can additionally be attached to inflected word forms as suffixes to create predicates. This operation changes the status of the respective case suffixes to infixes. Both grammars describe this general process, but do not provide a clear paradigm to support it for all cases. From the given examples, it can only be seen that these additional suffixes are used in the locative. Therefore, these suffixes are also indicated as infixes in the final paradigm in Table 4.18.

Case	Singular	Plural
Nominative		
Accusative	-ı, -i, -yi, -yü, -yı, -yu, -ü, -u	-ı, -i, -yi, -yü, -yı, -yu, -ü, -u
Genitive	-ın, -nin, -nün, -nın, -nun, -in, -ün, -un	-ın, -nin, -nün, -nın, -nun, -in, -ün, -un
Dative	-a, -ye, -ya, -e	-a, -ye, -ya, -e
Locative	-da, -de, -te, -ta, -da-, -de-, -te-, -ta-	-da, -de, -te, -ta, -da-, -de-, -te-, -ta-
Ablative	-dan, -den, -ten, -tan	-dan, -den, -ten, -tan

Table 4.18: Table of Turkish Nominal Case Markers

Turkish Adjectives

When approaching Turkish adjectives, it becomes apparent that this part of speech is not being inflected at all (Göksel and Kerslake, 2005). Therefore, no paradigm of adjectival case markers can be compiled. Nevertheless, some special features of this part of speech should be mentioned briefly. In Turkish, several nouns may be used as adjectives without any special marking (Ersen-Rasch, 2012). The other way around, even more adjectives can be used as nouns. Therefore, in these instances and if the noun is not inflected, the word order is necessary to clearly distinguish between the adjective, that usually comes first, and the noun. Concerning their function, Turkish adjectives may be used attributively and predicatively as known from other languages. Additionally, some may also simultaneously be used adverbially (Ersen-Rasch, 2012). Finally, in contrast to the missing case inflection, the comparative forms of Turkish adjectives may be built.

Turkish Numerals

Concerning Turkish numerals, both grammars indicate a similar lack of inflection as with Turkish adjectives before. Ersen-Rasch (2012) as well as Göksel and Kerslake (2005) state that numerals are suffixed by -inci, -üncü, -uncu, -ıncı, -nci, -ncı without making a differentiation according to case.

Turkish Pronouns

As mentioned in the course of the description of Turkish nouns, pronominal suffixes that indicate possession may be directly attached to them (Ersen-Rasch, 2012). As these are marked before the case and do not fuse with it, they were therefore not taken into account (Blake, 1994). Besides, it is Turkish pronouns as independent parts of speech that are now under observation in this section. In order to extract the Turkish pronominal case markers, the order in which the respective pronouns are presented in Ersen-Rasch (2012) is followed.

This begins with the personal pronouns *ben, sen, o* 'I', 'you' and 'he/she/it' in the singular together with their now clearly differentiated plural forms *biz, siz, on* 'we', 'you' and 'they'. These nominative forms are again unmarked and resemble a clear stem to which case markers are attached. These are initially added to the paradigm in Table 4.19. In this context, a vowel alternation in the dative of the first and second person singular to *ban-* and *san-* should be mentioned. However, this does not affect the case endings that are similar to the nominal ones. By contrast, the genitive marker of the first person singular and plural shows a special form ending on *-im*. In general, the genitive forms of the personal pronoun are yet again identical to the ones of the possessive pronoun. Apart from these markers, that mainly resemble the nominal forms, further suffixes may be attached to personal pronouns. The ending *-ce* roughly denotes 'in my/your/... opinion', attaching *-le/-la* means 'with me/you/...' and the suffix *-ız/-uz* translates to 'without me/you/...' (Ersen-Rasch, 2012). However, as these are not classified as case markers, they are not added to the paradigm.

In order to create a possessive pronoun that can stand in place of a noun, the nominalizing suffix *-ki* is attached to the personal pronouns. Its case markers mainly follow those of the personal pronouns but some differing forms are included in Table 4.19. The same holds true for the Turkish demonstrative pronouns *bu, şu* 'this (one here)', 'this (one there)' and *o* 'that (one there)'. As before, the suffix *-la* can be attached to the demonstrative pronouns was well, denoting roughly 'so that' whilst *-ız/-uz* translates to 'without this/that'. Again these suffixes are not added to the paradigm of case markers although they may be used to indicate a similar relation.

According to Ersen-Rasch (2012) Turkish also knows four locative pronouns. Whilst *nere* may be translated as 'which place?', *bura, şura* and *ora* denote 'this place here/there/over there'. Most of their case marking suffixes are already part of Table 4.19. The new forms are added to the paradigm.

When it comes to the Turkish reflexive pronoun *kendi* that translates to 'own/self', all case markers are already part of the paradigm. The same holds true for the reciprocal pronoun *birbiri* 'each other'. Regarding the two Turkish interrogatives *kim* 'who' and *ne* 'what', the latter one shows some further, finer grained case markers like *-yin* in the genitive singular that are added to Table 4.19. Additional *hangi* or *hangisi* 'whose' may already be declined by using the paradigm.

Amongst the larger set of Turkish indefinite pronouns, Ersen-Rasch (2012) explicitly mentions *kimse*, '(some)one (person)'. Since this pronoun declines like a noun, its case markers are already covered by the paradigm. The same applies to the further declinable representatives of this class which complement the presentation of Turkish pronouns in this grammar.

Finally, to ensure the paradigm in Table 4.19 covers all pronominal case markers, it is revised by using Göksel and Kerslake (2005). This grammar does not show clear paradigms of pronominal case marking but mentions the general procedure and deviating forms. Based on the examples shown, no differing markers could be identified. In general, as due to the agglutinative character of Turkish, the pronominal case markers are very similar to those attached to nouns.

Case	Singular	Plural
Nominative	-si, -sı	-i, -ı
Accusative	-i, -nu, -ni, -u, -yı, -yi	-i, -ı, -ini
Genitive	-im, -in, -nun, -nin, -nın, -yin	-im, -in, -ın, -inin
Dative	-a, -na, -n, -e, -ya, -ne, -ye	-i, -ı, -e, -a, -ine
Locative	-de, -nda, -nde	-de, -da, -inde
Ablative	-den, -dan, -nden, -ndan, -yden	-den, -dan, -inden

Table 4.19: Table of Turkish Pronominal Case Markers

4.3 A Uralic Language: Finnish

Although Finland is bordered by countries in which primarily Indo-European languages are spoken, the Finnish language belongs to the Uralic language family that is quite different from these. In more detail, Finnish is part of the Finnic, or rather Finno-Ugric genus, being closely related to Estonian and Hungarian (see Dryer and Haspelmath (2013) and White (2008)). Like Turkish before, Finnish is an agglutinative language in which grammatical forms of words are crated by attaching suffixes to stems of nominals, i.e. nouns, adjectives, pronouns and numerals. This is described as being relatively regular by White (2008) which also indicates the presence of attributive agreement of adjectives, pronouns and numerals to their preceding noun. Karlsson (2018) highlights the extensive use of inflection in order to indicate meaning by different word forms.

Finnish uses fifteen different cases. In the following, these are four groups according to Putz (2002) whilst their function and possible translation is given according to Karlsson (2018). At first, four syntactic cases are differentiated. Of these, the nominative indicates the basic agent, the genitive possession and the accusative represents a specific object case. A further case of this group is the partitive that is used for indefinite disclosures and quantification that translate to 'some ...'. The two semi syntactic cases essive and translative signal a state or a change of a state. In order to display location, Finnish uses six cases which may be divided into each tree internal and external ones. The first group includes the inessive 'inside', the elative 'out of' as well as the illative 'into'. Of the latter ones, the adessive denotes 'on' plus possibly the instrument, the ablative 'off' and the allative 'onto'. Blake (2012) explains the origin of these meanings by the fusion of orientation markers with case markers. In particular, the locative cases essive and partitive have partly fused with forms for 'in' and 'on' to yield local cases. The three resembling cases are called 'marginal' by Putz (2002) which becomes apparent when considering their function. The comitative is used to state accompanying behavior as 'with ...' whilst the abessive denotes the contrary, 'without ...'. Finally, the instructive is used idiomatically. When it comes to the category of number, Finnish differentiates between singular and plural, the similar question of depicting these as in Turkish arises again, especially because the case endings of the plural are usually similar to the ones in the singular. This results from the agglutinative insertion of the affix *-i-* between the stem and the case ending for all cases apart from the nominative. Here, *-t* is used as no further case marker follows. However, as White (2008) mentions, the markers of the genitive, partitive and illative may be slightly different in both numbers. Therefore, the following paradigms differentiates between this category as well. Finally, Finnish does not differentiate any grammatical gender which can have an affect on its case markers.

Finnish Nouns

In an initial presentation of nominal Finnish stems, White (2008) shows that stem often but not always represents the basic form to which inflectional endings are attached. This is mainly based on an alternation in the stem. Therefore, the following paradigm of Finnish nominal case markers in 4.20 is based on the forms that White (2008) clearly indicates as such. Starting from the given summary table, further mostly phonologically conditioned differences are included. These are shown by concrete examples.

As for most of the previous languages, the Finnish nominative singular is unmarked as the basic form of a noun. This may also apply to its related case, the accusative singular. It is then again these two cases that are marked differently in the plural. As a sign of the basic form, -t is added to the stem and also to the paradigm. For most of the further cases, the similar markers of the singular and the plural need to be transferred. The differing ones are extracted from the examples in White (2008). A further special case concerns the commitative. This is only marked in the plural although its meaning is both singular or plural. It is also the forms of the commitative that indicate the eventuality of a further possession suffix being attached to Finnish nominal case forms. This issue is also highlighted by Blake (1994) who states that a connection of both possessive and case suffix would lead to an extremely extensive paradigm of case endings. Therefore, following the theoretical description of case markers as affixes, all nominal case suffixes similarly need to be seen as case infixes. Although White (2008) states that possessive suffixes are simply added after the case ending of words, no complete paradigm for all cases but only examples are given. Furthermore, it is indicated that case endings remain strong and both -t and -n as final sounds of case markers are dropped when a possessive suffix is added after them. Similarly, the transitive ending -ksi- changes to -kse-. Following these specifications, all suffixes in Table 4.20 are added accordingly as infixes.

Finally, this paradigm is revised by using Karlsson (2018) as a second source. As far as possible, all case markers were extracted from the general overview and the concrete examples in this grammar.

Finnish Adjectives

This section on Finnish adjectival case markers can be kept rather short. As White (2008) and the further grammars indicate, no formal difference exists between nouns and adjectives. Therefore, the paradigm in Table 4.21 corresponds to that of nouns in terms of the displayed suffixes. The possessive marker that changes these to infixes is not attached to adjectives. Therefore, the following forms are only those suffixes of the nominal paradigm above. This is supported by the second grammar, Karlsson (2018), as well. Again, no further suffixes are found in this source.

Finnish Numerals

As compared to adjectives, Finnish cardinal numerals have both singular and plural forms in all cases (White, 2008). According to Karlsson (2018), they decline like the nominals analyzed before. In concrete, the following paradigm in Table 4.22 consists of the same case markers as have been worked out for the Finnish nouns. Whilst the case of cardinal numbers from one to ten is indicated by case suffixes, it is indicated in numbers from eleven to nineteen by infixes. This results from the invariable form *toista* that is attached to the basic numbers. However, this also applies for higher numbers up to one hundred in which both parts are inflected for case. In concrete, the marking on the tens is realized as an infix whilst the one on the ones represents a suffix.

When it comes to ordinal numbers, White (2008) indicates that they are decorated by the ordinal sign -s. This applies for the nominative whilst the forms of the further cases as indicated in both grammars correspond to the ones already in Table 4.22.

Case	Singular	Plural
Nominative		
Genitive	-n	-en, -den, -ten, -tten, -e-, -de-, -te-, -tte-
Accusative	-n	-t
Partitive	-a, -ä, -ta, -tta, -tä, -ttä, -a-, -ä-, -ta-, -tta-, -tä-, -ttä-	-a, -ä, -ta, -tä, -a-, -ä-, -ta-, -tä-
Inessive	-ssa, -ssä, -ssa-, -ssä-	-ssa, -ssä, -ssa-, -ssä-
Elative	-sta, -stä, -sta-, -stä-	-sta, -stä, -sta-, -stä-
Illative	-on, -ön, -an, -än, -en, -in, -un, -yn, -han, -hän, -hin, -hen, -hun, -hon, -hön, -seen, -o-, -ö-, -a-, -ä-, -e-, -i-, -u-, -y-, -ha-, -hä-, -hi-, -he-, -hu-, -ho-, -hö-, -see-	-in, -hin, -siin, -i-, -hi-, -sii-
Adessive	-lla, -llä, -lla-, -llä-	-lla, -llä, -lla-, -llä-
Ablative	-lta, -ltä, -lta-, -ltä-	-lta, -ltä, -lta-, -ltä-
Allative	-lle, -lle-	-lle, -lle-
Essive	-na, -nä, -na-, -nä-	-na, -nä, -na-, -nä-
Translative	-ksi, -kse-	-ksi, -kse-
Comitative		-ne, -ne-
Abessive	-tta, -ttä, -tta-, -ttä-	-tta, -ttä, -tta-, -ttä-
Instructive	-n	-n

Table 4.20: Table of Finnish Nominal Case Markers

Case	Singular	Plural
Nominative		
Genitive	-n	-en, -den, -ten, -tten
Accusative	-n	-t
Partitive	-a, -ä, -ta, -tta, -tä, -ttä	-a, -ä, -ta, -tä
Inessive	-ssa, -ssä	-ssa, -ssä
Elative	-sta, -stä	-sta, -stä
Illative	-on, -ön, -an, -än, -en, -in, -un, -yn, -han, -hän, -hin, -hen, -hun, -hon, -hön, -seen	-in, -hin, -siin
Adessive	-lla, -llä	-lla, -llä
Ablative	-lta, -ltä	-lta, -ltä
Allative	-lle	-lle
Essive	-na, -nä	-na, -nä
Translative	-ksi	-ksi
Comitative		-ne
Abessive	-tta, -ttä	-tta, -ttä
Instructive	-n	-n

Table 4.21: Table of Finnish Adjectival Case Markers

Case	Singular	Plural
Nominative	-s	
Genitive	-n	-en, -den, -ten, -tten
Accusative	-n	-t
Partitive	-a, -ä, -ta, -tta, -tä, -ttä	-a, -ä, -ta, -tä
Inessive	-ssa, -ssä	-ssa, -ssä
Elative	-sta, -stä	-sta, -stä
Illative	-on, -ön, -an, -än, -en, -in, -un, -yn, -han, -hän, -hin, -hen, -hun, -hon, -hön, -seen	-in, -hin, -siin
Adessive	-lla, -llä	-lla, -llä
Ablative	-lta, -ltä	-lta, -ltä
Allative	-lle	-lle
Essive	-na, -nä	-na, -nä
Translative	-ksi	-ksi
Comitative		-ne
Abessive	-tta, -ttä	-tta, -ttä
Instructive	-n	-n

Table 4.22: Table of Finnish Numeral Case Markers

Finnish Pronouns

In order to create a paradigm of Finnish pronominal case markers, this passage follows the order and concrete declension tables of pronouns given in Karlsson (2018). As opposed to White (2008) that only broadly states that the Finnish pronouns have all case forms, this grammar does not provide markers in all cases, especially not in the marginal ones (comitative, abessive and instructive). Therefore, the paradigm in Table 4.23 builds on this rather restricted approach. As some of the pronouns function like nouns, their case markers will intersect. The same holds true for such that are used as premodifiers and thereby need to agree with the nominal head. However, as there are some differences, the declension table of the nouns will not be taken over as a general basis. Nevertheless, it is used to revise the following forms by Karlsson (2018).

At first, personal pronouns are under observation. Whilst for the singular forms *minä* 'I', *sinä* 'you' and *hän* 'he/she' the stem is *minu-, sinu-* and *häne-*, both basic form and stem coincide in the plural. In concrete, this is *me* 'we', *te* 'you' and *he* 'they'. The declension follows the basic nominal forms but adds the marker *-dät* for the accusative and *-dän* for the genitive plural. The two main Finnish demonstrative pronouns are *tämä* 'this' and *tuo* 'that'. In the plural, they are *nämä* 'these' and *nuo* 'those'. The forms *se* 'it' and *ne* 'they' refers to something previously mentioned. Concerning their case markers, *-män* and *-den* are used for the genitive singular and plural whilst all other forms decline regularly. Further differing forms appear when it comes to the interrogative pronouns *kuka* 'who' and *mikä* 'which, what'. The suffix *-kä* may be used to indicate the genitive and accuastive singular as well as the nominative, genitive and accusative plural. All further forms indicated by Karlsson (2018) correspond to the ones of the nominal paradigm. The same holds true for *kumpi* 'which of two' and *millainen* 'what kind of'. Concerning the Finnish indefinite pronouns, their case marking behavior strongly influences the paradigm in Table 4.23 to look even more like the nominal one. Of these, *joku* 'someone' is a two parted pronoun in which both parts *jo-* and *-ku* are inflected. Resulting from this, infix forms of its case markers are to be added to the paradigm. This also needs to be done for *jokin* 'something', in which *-kin* is added after the case marker as a clitic. The same holds true for the negative equivalents of *jokin*, *kukaan* 'no one, anyone' and *mikään* 'nothing, anything'. For both forms, differing infixes are added to Table 4.23. For *jompi=kumpi* 'either one, one or the other' both parts decline again. In *kumpi-kin* 'each of two, both', *kumpi-kaan* 'neither' and *kukin* 'each, everyone' again the clitics *-kin* and *kaan* are added. At this point, most infixes are already part of the paradigm. Though, further forms are still enclosed. When it comes to Finnish relative pronouns, the most common one is *joka* 'who, which, that'. For this form *-ka* is also appended in the genitive singular. All other markers, likewise for the second relative pronoun *mikä*, are already in the paradigm. At this point the impression arises that the paradigm is quite complete. When White (2008) is used to revise these findings, there were no apparent differences concerning the concrete case markers although some further, finer grained pronouns were presented. From this, it becomes apparent that the final paradigm in Table 4.23 is comprehensive. Most of the pronominal case markers resemble those of the other nominals. However, as several different forms emerged and not all cases are distinguished, building this paradigm from scratch was necessary.

Case	Singular	Plural
Nominative		-kä
Genitive	-n, -män, -kä, -nkä, -n-, -ka, -nka	-dän, -den, -kä, -den-, -en-
Accusative	-t, -kä	-dät, -kä
Partitive	-a, -ta, -tä, -ta-, -tä-, -a-	-tä, -ta, -a, -ta-, -tä-
Inessive	-ssa, -ssä, -nä, -ssa-, -ssä-	-ssä, -ssa, -ssa-, -ssä-
Elative	-sta, -stä, -tä, -sta-, -stä-	-stä, -sta, -sta-, -stä-
Illative	-un, -en, -hän, -hon, -hen, -hin, -an, -hun, -hon-, -en-, -hin-, -an-, -hun-	-hin, -in, -hin-, -in-
Adessive	-lla, -llä, -lla-, -llä-	-llä, -lla, -lla-, -llä-
Ablative	-lta, -ltä, -lta-, -ltä-	-ltä, -lta, -lta-, -ltä-
Allative	-lle, -lle-	-lle, -lle-
Essive	-nä, -na, -na-	-nä, -na, -na-
Translative	-ksi, -ksi-	-ksi, -ksi-

Table 4.23: Table of Finnish Pronominal Case Markers

5 Conclusion

At this point, this thesis has examined the theoretical foundations of case, case semantics and its connection to surface case markers. After a linguistically and computationally sensible definition of these forms was set out on the basis of a morphological approach, a methodology of extracting these morphemes from a multilingual set of languages was created and practically conducted. During the analysis of four Indo-European languages, that may all be classified as rather fusional, it became apparent that the theoretical separation of a nominal stem and the case marker is not always clearly possible. In these instances, speakers need to trust their feeling for language and linguists the set out declension paradigms in grammar works in order to extract concrete case markers. However, as these are mostly suffixes, they can be extracted from text by using computational methods quite well. Similarly, this applies to those suffixing markers of the two agglutinative languages that have been analyzed. Yet, their characteristic of additionally adding suffixes not related to case and therefore turning the case suffixes into infixes complicates the matter.

Overall, it was challenging to apply such a rather fixed system of case markers to the diversity of natural languages. In order to create a reliable and valid gold standard that may be used as a reference work, it was tried to be as coherent and detailed as possible. On the other hand, very fine grained or incongruent details that do not fit into the pattern had to be omitted. It was aimed to describe these decisions similarly as all those forms that lead towards the final paradigms.

In retrospect, it was initially expected to analyze and include more languages into the gold standard. However, during the editing process of the first languages, it became apparent that analyzing the complete case marking behaviour of languages was much more extensive than anticipated. Especially due to the fact that grammar books do not show clear paradigms for all distinct members of a certain part of speech, it was necessary to transfer several case markers between these. Additionally, it is mostly summarized declension tables that are given which lack special forms that are important for the gold standard, though. Such forms often needed to be extracted from given examples.

When it comes to an outlook for possible future work on this task, a short overview on further languages that would be interesting to be further analyzed should be given. As another Uralic-Ugric language, Hungarian shows an exceedingly diverse set of cases that is consequently also of interest when it comes to case semantics. Similarly, Polish as a further Slavic language that does not use the Cyrillic Alphabet would be on this list. These two are, for example, also analyzed is the contrastive grammar of Gunkel et al. (2017) that compares them and two further languages to German. Additionally, Tamil as the most representative member of Dravidian languages with a rather comprehensive case system could improve the gold standard. Together with Hindi, the most representative of the possible Indic languages, their script is rather challenging, though. The same holds true for Japanese that would also present an agglutinative language. Eventually, it was also thought of dealing with Hebrew or Arabic, two Semitic languages. However their present case system is rather unclear in present day English.

Further challenges that need to be addressed when continuing this task arise. A first example would be the possibility of double case marking in some languages. This was exemplary mentioned for some Finnish pronouns in which the case morpheme is represented as an infix and a suffix. A possible solution for this very fine grained analysis would be to initially check for suffixes and then, in a second step, look for in- and prefixes. Moreover, as being mentioned by Dryer (2013), languages without either adpositional or morphological

case markings are also widely distributed on this planet. Therefore, it would be important to try to incorporate case marking strategies that are executed via the word order. Blake (1994) additionally indicates on this that such analytic case marking methods are rather used for non-core or peripheral relations and are thereby especially important to detect semantic roles.

Finally, during the work on this thesis, UniMorph 4.0 (Batsuren et al., 2022) was introduced. This presents the succeeding approach of UniMorph by which the machine generate silver standard used in *CaMEL* was created. This new paper presents normalized morphological inflection tables for hundreds of diverse languages of the world, compiled by numerous linguists. However, it remains unclear whether and how exactly case is distinguished in the process and the final resulting tables. Therefore, this thesis may hopefully maintain its raison d'être for some more time.

Bibliography

Matthew Baerman. Case syncretism. In Andrej Malčukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford Handbooks in Linguistics. Oxford University Press, Oxford, 2012.

Khuyagbaatar Batsuren, Omer Goldman, Salam Khalifa, Nizar Habash, Witold Kieraś, Gábor Bella, Brian Leonard, Garrett Nicolai, Kyle Gorman, Yustinus Ghanggo Ate, Maria Ryskina, Sabrina J. Mielke, Elena Budianskaya, Charbel El-Khaissi, Tiago Pimentel, Michael Gasser, William Lane, Mohit Raj, Matt Coler, Jaime Rafael Montoya Samame, Delio Siticonatzi Camaiteri, Esaú Zumaeta Rojas, Didier López Francis, Arturo Oncevay, Juan López Bautista, Gema Celeste Silva Villegas, Lucas Torroba Hennigen, Adam Ek, David Guriel, Peter Dirix, Jean-Philippe Bernardy, Andrey Scherbakov, Aziyana Bayyr-ool, Antonios Anastasopoulos, Roberto Zariquiey, Karina Sheifer, Sofya Ganieva, Hilaria Cruz, Ritván Karahóǧa, Stella Markantonatou, George Pavlidis, Matvey Plugaryov, Elena Klyachko, Ali Salehi, Candy Angulo, Jatayu Baxi, Andrew Krizhanovsky, Natalia Krizhanovskaya, Elizabeth Salesky, Clara Vania, Sardana Ivanova, Jennifer White, Rowan Hall Maudslay, Josef Valvoda, Ran Zmigrod, Paula Czarnowska, Irene Nikkarinen, Aelita Salchak, Brijesh Bhatt, Christopher Straughn, Zoey Liu, Jonathan North Washington, Yuval Pinter, Duygu Ataman, Marcin Wolinski, Totok Suhardijanto, Anna Yablonskaya, Niklas Stoehr, Hossep Dolatian, Zahroh Nuriah, Shyam Ratan, Francis M. Tyers, Edoardo M. Ponti, Grant Aiton, Aryaman Arora, Richard J. Hatcher, Ritesh Kumar, Jeremiah Young, Daria Rodionova, Anastasia Yemelina, Taras Andrushko, Igor Marchenko, Polina Mashkovtseva, Alexandra Serova, Emily Prud'hommeaux, Maria Nepomniashchaya, Fausto Giunchiglia, Eleanor Chodroff, Mans Hulden, Miikka Silfverberg, Arya D. McCarthy, David Yarowsky, Ryan Cotterell, Reut Tsarfaty, and Ekaterina Vylomova. Unimorph 4.0: Universal morphology, 2022. URL `https://arxiv.org/abs/2205.03608`.

Balthasar Bickel and Johanna Nichols. Fusion of selected inflectional formatives. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013. URL `https://wals.info/chapter/20`.

Barry J. Blake. *Case.* Cambridge Textbooks in Linguistics. Cambridge Univ. Press, Cambridge, 1994.

Barry J. Blake. History of the research on case. In Andrej Malčukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford Handbooks in Linguistics. Oxford University Press, Oxford, 2012.

Miriam Butt. *Theories of Case.* Cambridge Textbooks in Linguistics. Cambridge Univ. Press, Cambridge, 2006.

Bernard Comrie. *Language Universals and Linguistic Typology.* Blackwell, Oxford, 1989.

Walter A. Cook. *Case Grammar Theory.* Georgetown Univ. Pr., Washington D.C., 1989.

Michael Cysouw. Inducing semantic roles. In Silvia Luraghi and Heiko Narrog, editors, *Perspectives on Semantic Roles*, pages 23–68. John Benjamins, Amsterdam, 2014.

Matthew S. Dryer. Position of case affixes. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013. URL `https://wals.info/chapter/51`.

Matthew S. Dryer and Martin Haspelmath, editors. *WALS Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013. URL `https://wals.info/`.

Peter Eisenberg. *Grundriss der Deutschen Grammatik: Das Wort*. J.B. Metzler, Stuttgart, 2020.

Hilke Elsen. *Grundzüge der Morphologie des Deutschen*. De Gruyter, Berlin, 2014.

David Embick and Morris Halle. On the status of *Stems* in morphological theory. In Twan Geerts, Ivo Van Ginneken, and Haike Jacobs, editors, *Romance Languages and Linguistic Theory*, pages 37–63. John Benjamins, Amsterdam, 2005.

Margarete I. Ersen-Rasch. *Türkische Grammatik*. Harrassowitz, Wiesbaden, 2012.

Charles J. Fillmore. *Fillmore's Case Grammar: A Reader*. Studies in descriptive linguistics. Groos, Heidelberg, 1987.

Scott Grimm. Semantics of case. *Morphology*, 21(3-4):515–544, 2011.

Lutz Gunkel, Christine Günther, and Ursula Hoberg. *Grammatik des Deutschen im europäischen Vergleich*. Schriften des Instituts für Deutsche Sprache. De Gruyter, Berlin, 2017.

Aslı Göksel and Celia Kerslake. *Turkish*. Routledge Comprehensive Grammars. Routledge, London, 2005.

Martin Haspelmath. Terminology of case. In Andrej Malčukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford Handbooks in Linguistics. Oxford University Press, Oxford, 2012.

Elke Hentschel and Petra M. Vokel, editors. *Deutsche Morphologie*. de Gruyter, Berlin, 2009.

David Holton, Peter Mackridge, and Irene Philippaki-Warburton. *Greek: An Essential Grammar*. Routledge Essential Grammars. Routledge, London, 2016.

Oliver A. Iggesen. Number of cases. In Matthew S. Dryer and Martin Haspelmath, editors, *The World Atlas of Language Structures Online*. Max Planck Institute for Evolutionary Anthropology, Leipzig, 2013. URL `https://wals.info/chapter/49`.

Fred Karlsson. *Finnish*. Routledge Comprehensive Grammars. Routledge, Abingdon, 2018.

Siegfried Kohls. *Langenscheidts Standardgrammatik Russisch*. Langenscheidt, Berlin, 2009.

Andrej Malčukov and Andrew Spencer, editors. *The Oxford Handbook of Case*. Oxford Handbooks in Linguistics. Oxford University Press, Oxford, 2012.

Thomas Mayer and Michael Cysouw. Creating a massively parallel Bible corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 3158–3163, Reykjavik, Iceland, May 2014. European Language Resources Association (ELRA).

Wilhelm Metger. *Neugriechische Kurzgrammatik*. Hueber, Ismaning, 1998.

Edith A. Moravcsik. The distribution of case. In Andrej Malčukov and Andrew Spencer, editors, *The Oxford Handbook of Case*, Oxford Handbooks in Linguistics. Oxford University Press, Oxford, 2012.

Don L. Nilsen. *Toward a Semantic Specification of Deep Case.* Janua Linguarum Series Minor 152. Mouton, The Hague, 1972.

Dirk G. J. Panhuis. *Lateinische Grammatik.* De Gruyter, Berlin, 2015.

Martin Putz. *Finnische Grammatik.* Edition Praesens, Wien, 2002.

Hans Rubenbauer. *Lateinische Grammatik.* Buchner, Bamberg, 1995.

Hans Ruge. *Grammatik des Neugriechischen.* Romiosini, Köln, 2001.

Said Sahel. *Kasus.* Kurze Einführungen in die germanistische Linguistik. Winter, Heidelberg, 2018.

Rolf Thieroff and Petra Vogel. *Flexion.* Kurze Einführungen in die germanistische Linguistik. Winter, Heidelberg, 2012.

Christian Touratier. *Lateinische Grammatik.* WBG, Darmstadt, 2013.

Terence Wade. *A Comprehensive Russian Grammar.* Wiley Blackwell, Chichester, 2020.

Leonie Weissweiler, Valentin Hofmann, Masoud Jalili Sabet, and Hinrich Schuetze. CaMEL: Case Marker Extraction without Labels. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5506–5516, Dublin, Ireland, May 2022. Association for Computational Linguistics. URL `https://aclanthology.org/2022.acl-long.377`.

Leila White. *A Grammar Book of Finnish.* Finn Lectura, Helsinki, 2008.

List of Tables

Contents of the Enclosed CD

The CD enclosed to this printed version contains a main directory named by the title of the thesis. In this, its original format in LaTeX is stored together with the pdf version and the further required files that are needed to compile it (clba.sty style sheet, bibl-casemarkers.bib BibTeX file and the CIS logo lmu_cis_logo.pdf). It should be noted at this point that the thesis was created by using the cloud-based LaTeX editor 'Overleaf'. Additionally, two further directories are added that contain digitally available sources that were used for the thesis and a presentation which was held in the accompanying colloquium. Finally, the created gold standard is appended in a processable python file. For this, the tables of case markers of each language in the preceding text were stored in a nested dictionary named after this language. In order to include the three dimensions of multilingual case markers, the dictionaries are nested in three levels. The value of the first key, the declinable parts of speech, is the two numbers. These again represent the key of a dictionary whose values are the cases used by the respective language. Lastly, each case is a key of a dictionary whose value is a set of case markers indicating this case. According to the specification in Weissweiler et al. (2022) instead of a hyphen, the $ sign is used here to indicate word boundaries of case markers. Thereby, a differentiate of the three types of affixes as well as complete case marking word forms can be made. Similarly, by differentiating part of speech and number, these categories can be included or excluded so that a comprehensive set of all case markers for each language can be created.