# Do certain word orders attract case marking?

A typological survery on the dependency of syntax and morphology.

Thesis submitted for the degree of Bachelor of Arts.

Eberhard Karls Universität Tübingen
Seminar für Sprachwissenschaft

Fabio Maion
Matriculation number: [redacted]
fabio.maion@student.uni-tuebingen.de
[redacted]
72072 Tübingen

supervised by
Dr. Christian Bentz

Date of submission: July 8, 2018

# Contents

# 1 Introduction

"Come, let us go down, and confuse their language there, so that they will not understand one another's speech" (Coogan, 2010, Gen 11:7). That languages of the world differ significantly with regard to one another has always been known. The quotation of the Bible proves that human beings reasoned about the variety among natural languages already more than two thousand years ago. It is clear that today's scientists do not longer believe in a God that has come down to earth and changed the humans' languages, but that it is by now known that language variety is a result of diachronic developments stretching across thousands of years. For linguists, it can be fascinating to explore different languages from different families to see how much and with regard to which points they actually differ. A question that directly follows from such research is according to which properties languages do not differ and whether there are properties that can be claimed to be alike in all the world's languages. Is it possible to find principles to which all natural languages obey?

The interest in this field of linguistic typology has probably had its breakthrough with Greenberg (1963). In his article, Greenberg listed a bunch of different statements claiming to have found rules that are valid for the languages of the world. Only a few of his claims stated absolute properties that are alike in all languages, one of them being the preference for subjects to occupy an earlier clause position than objects. What Greenberg mostly observed were dependencies between different properties and most of his claims are so called implicational universals. These state that if a languages exhibits a given property A, it should also exhibit another property B. Most interested in syntax and word order, Greenberg (1963) does nevertheless propose ideas how morphological properties could interact with syntactic properties. This interaction is of interest also to recent linguistic research where it has been claimed that we cannot really define terms as "morphology" and "syntax", but should rather consider them as some kind of unit (Haspelmath, 2011). Haspelmath argues that there is a problem with a precise definition of what a word is and that there can be great difficulties in finding cross-linguistic properties for elements on the borderline between clitics and affixes.

Whether one accepts Haspelmath's assertion or not, there are concrete proposals about how morphology and syntax could interact. It is for example believed that case marking occurs more frequently with certain word orders than with others. What will be especially relevant to this thesis is what Greenberg postulated in his universal 41: "If in a language the verb follows both the nominal subject and the nominal object as the dominant order, the languages almost always has a case system" (Greenberg, 1963, 96). This statement shall be evaluated given typological data on the structural properties of languages of the world. I will furthermore examine the idea of an interaction between a lack of a dominant word order and the presence of case marking, a proposal found in various papers (e.g. Fedzechkina et al. (2017), Sinnemäki (2008)).

The thesis will be subdivided as follows. In chapter 2, I will present the claim about a correlation of certain word orders with the presence of case marking in detail. In chapter 3, I will investigate the following hypotheses by statistical analyses based on data from the World Atlas of Language Structure (WALS). The following chapters will then be dedicated to different possible explanations for the statistical results. In

chapter 4, I will discuss the necessity of a clear encoding of syntactic roles. Chapter 5 will focus on the claim that the human parser prefers to receive important information as early as possible. Chapter 6 will be dedicated to the phenomenon of grammaticalization and its possible impact on language structure. In chapter 7, I will finally summarize the results and possible explanations.

## 2 The hypothesis

The idea that morphology and syntax interact with each other has for a long time been a topic of linguistic research. At least since Greenberg (1963) there are strong reasons to think of a mutual dependency. We have already seen that Greenberg asserted that languages with verb-final order should have case marking. An example of such a language is Turkish. In sentence (1), the two arguments of the verb, *arkadaş* and *ev*, are case marked for nominative and definite accusative respectively. Furthermore, *baba* bears genitive case marking. The verb is inflected for person and tense and occupies the final position in the clause:

(1) Arkadaş-ımız-Ø baba-m-ın ev-i-ni al-dı-Ø
friend-POSS.1.PL-NOM father-POSS.1.SG-GEN house-POSS.3.SG-ACC buy-PAST-3.SG
"Our friend bought my father's house"

Before a thorough examination of Greenberg's universal can take place, it is necessary to shortly explain why a reanalysis of his finding is needful. When he published his study, Greenberg based his claims on a sample of 30 languages he was himself acquainted with. Although the sample was genetically and geographically balanced, basing universal statements about all the world's languages on an analysis of just 30 languages is not a stable base. Greenberg was himself conscious of this fact and considered his results as preliminary, encouraging further work on the topic: "The tentative nature of the conclusions set forth here should be evident to the reader" (Greenberg, 1963, 73).

In fact, some of his in total 45 proposed universals have not withstood more detailed statistical analyses. Dryer (1988) showed that there is no correlation between the order of object and verb and the order of noun and adjective - contrarily to what was claimed by Greenberg and believed to be correct until then. However, not all of Greenberg's universals have been disproved, some others are still quite vivid and have received further confirmation. Culbertson et al. (2012) confirmed Greenberg's statement about the correlation of noun-adjective and noun-numeral order, known as universal 18: "When the descriptive adjective precedes the noun, the demonstrative and the numeral, with overwhelming more than chance frequency, do likewise" (Greenberg, 1963, 86).

This shows that despite the methodological problem of the small sample, one should not throw the baby out with the bathwater and discard Greenberg's universals. Concerning his universal 41, which will be one main topic in this thesis, various other authors have accepted it as being true (e.g. Blake (1994, 15), Bentz and Christiansen (2013)). Still, I do not know of any study in which the correlation between verb-final word order and the presence of case marking in a language is tested with statistical methods. This is what I aim to do in chapter 3. Moreover, Greenberg did not propose any explanation for this supposed correlation. I will therefore, following the statistical testing, summarize some studies that come from different linguistic areas and all can contribute to a theoretical explanation of universal 41.

What is true for the correlation between verb-final order and the presence of case marking, namely that there is no study explicitly testing it, is not true for another correlation concerning word order and case marking. It is widely believed that languages lacking a dominant word order should use case marking to

encode the subject and the object in a clause (see Blake (1994, 15), Velupillai (2012, 176 and 282)). An example of such a language is modern Greek:

(2) O pater-as mou diavaz-i Kazantzak-i.
ART.NOM father-NOM my read-3.SG.PRES Kazantzakis-ACC
"My father is reading Kazantzakis"

(3) Diavaz-i o pater-as mou Kazantzak-i.
read-3.SG.PRES ART.NOM father-NOM my Kazantzakis-ACC
"My father is reading Kazantzakis"

(Mackridge, 1985, 235)

In the sentences above, it can be seen that Modern Greek has case marking on both the verb's arguments. The two sentences present the two word orders that are claimed to be pragmatically neutral in Modern Greek, SVO and VSO. However, other word orders are not impossible, but may be used if one of the constituents is to be accented.

Sinnemäki (2008) used a phylogenetically and geographically balanced sample of 50 languages to check whether a significant statistical correlation can be established. His main goal was to test whether low complexity in one domain (e.g. in syntax, absence of dominant word order) yields high complexity in another domain (e.g. in morphology, presence of case marking).

To check his hypothesis, Sinnemäki (2008) calculated what he called the functional load (FL) of the three possible encoding strategies (head marking, dependent marking and word order). Functional load was defined as "the measure of how often a contrast – a particular meaning distinction made with particular forms – is employed in a language" (Sinnemäki, 2008, 72). Its values could range from 1-4, depending on how extensively the different encoding strategies can be employed. Hungarian for example received the value 1 for word order and the value 4 for dependent marking, head marking of the agent and head marking of the patient (Sinnemäki, 2008, 80). These values meant that word order is never used in Hungarian to keep the verb's arguments apart. On the other hand, case marking and head marking might be used in potentially every sentence to keep the arguments apart.

Thereafter, Sinnemäki performed an analysis in R using Kendall's tau nonparametric correlation test to find out which correlations could be found in his sample between the functional loads of the different encoding strategies. A chi-square test was used to double-check the correlations for which Kendall's tau had yielded statistical significance. Both test showed a significant p-value only for one pair of functional loads, namely the one of word order and of dependent marking (Sinnemäki, 2008, 83). Sinnemäki therefore concluded that there is a trade-off in complexity between word order and case marking.

What the review of relevant literature so far has shown is that there are ideas on how word order and case marking interact with each other. It is supposed that case marking is more likely to be present in languages that either lack a dominant word order or have verb-final word order. In the following chapter, I will use data from the WALS to examine these hypotheses statistically. Most attention will be dedicated to the presence of

case marking in verb-final languages since no statistical testing has ever been done for this claim. Sinnemäki (2008) already provides strong arguments for the presence of case marking in free word order languages. I will however try to replicate also his results.

# 3 Word order and case marking - Empiric evidence

## 3.1 Data

Having discussed the theoretical background for my thesis, I will now turn to the statistical analyses. The data that will be used has been acquired from the WALS chapters 49 (Iggesen, 2013), 51 (Dryer, 2013c) and 81 (Dryer, 2013b). The chapters 49 and 51 give information about the presence or absence of case marking in the world's languages, whereas chapter 81 presents data on dominant word order. Information on case marking could also be acquired through the chapters 28 and 50 in WALS. However, they are not included for different reasons. Chapter 50 is not included because it uses the same sample as chapter 49, and chapter 28 is disregarded because it contains a category "Inflectional case marking is absent or minimal", merging languages with no case marking and language with a small case system. Consequently, its data does not allow to draw conclusions about absence or presence of a case system in the languages that belong to this category.

Since chapter 49 provides information not only on the presence of case marking, but also on the number of cases in a given language I will work with two different samples: a larger one (Sample A) containing 890 languages whose values on case marking are binary (present or absent) and a smaller one (Sample B) containing 217 languages and giving 8 different values for the variable case marking (no cases, 2 cases, 3 cases, 4 cases, 5 cases, 6-7 cases, 8-9 cases, 10 cases or more). Thereby, in Sample A data from the chapters 49, 51 and 81 is merged and Sample B contains the merged data from the chapters 51 and 81. Of course, this sampling procedure poses a problem: the resulting samples are convenience samples and might therefore be biased. To control for phylogenetical effects, I will perform the analyses in the following way. First the sample will be analyzed as a whole. Subsequently the analyses will be repeated considering the phylogenetical problem. How this is exactly done will be further explained in section 3.2.

Before it is possible to perform statistical analyses on the correlation of dominant SOV word order[1] and the presence of a case system as it is done in the next section, one first has to clarify two things. What is understood by a "case system"? And what is the definition of "dominant word order"?

In both cases, I will follow the definition given in the respective WALS chapters. In chapter 49, a marker is counted as being a case marker if there is "a sufficient degree of bondedness (phonological integration)

[1]The same correlation is of course expected to hold true for OSV languages. The only reason they are not mentioned separately is their extreme rareness. For the following analyses, there was data available from just two OSV languages: Tobati (Austronesian; Indonesia) and Wik Ngathana (Pama-Nyungan; Australia). Data on the number of cases was not available for these languages such they appear only in Sample A.

with its host noun in basic syntactic constructions – i.e. in non-expanded, head-only NPs" (Iggesen, 2013). A language is thus counted as having case markers even if it is not clear whether the markers in question are affixes or clitics. This implies a wider definition of case marking than often assumed. What might seem disputable in the first moment, is strongly encouraged by the fact that it is highly doubtful whether one can seriously tell apart affixes from clitics (Haspelmath, 2011). An example of a language which has case marking according to this broader definition is Japanese. The case markers in Japanese are similar to postpositions, but differ in the fact that they may be omitted in casual speech. This is why they are probably best analyzed as clitics:

(4) Taroo=ga tosyokan-de hon=o yon-de-ir-u.
Taroo=NOM library-POSTP book=ACC read-GERUND-be-NONPAST
"Taroo is reading a book in the library"

(Tsujimura, 1996, 135)

It was nevertheless not possible to use all the data from the WALS sample. All languages that were coded as having "exclusively borderline case-marking" in chapter 49 had to be excluded from both test samples. In those languages, it is not clear whether certain morphological markers are best interpreted as case-markers or as derivational suffixes. An example for such a language is Plains Cree which has a suffix *-ehk* that might be interpreted as a locative case-suffix. But since it then would be the only case-suffix in Plains Cree, it may also be analyzed as a suffix deriving locative adverbs (Iggesen, 2013). One can therefore not surely assert whether languages like Plains Cree have a case system. I therefore decided not to include any of the languages with the values "exclusively borderline case-marking". It was additionally necessary to exclude 21 languages from Sample A, because the chapters 49 and 51 provided divergent information. An example is French that was encoded as not having cases in chapter 49, but encoded as having prepositional clitics in chapter 51.

It still has to be defined what is understood by the term "dominant word order". I deliberately use the term "dominant" instead of "basic", a term which is also often found in the literature. Why the term "dominant" is preferred should get clear if one takes a look at the definition. A language is assigned a dominant word order if a specific order is at least twice as common as the second most frequent word order (Dryer, 2013a). This definition has several advantages. First, it makes the present study independent of any specific syntactic theory. To understand what is meant, one can consider for example the case of German. Whereas generative grammar claims that the basic word order in German is SOV (Müller, 2016, 99), German might also be analyzed as SVO language (Greenberg, 1963). The criterion of frequency allows to disregard this problematic and to analyze German as a language lacking a dominant word order, since both SOV and SVO word order occur frequently. The word order in main clauses will be SOV if an auxiliary is used to form the predicate (the main verb standing at the end of the sentence), otherwise the word order is SVO.

Moreover, Dryer (1997) found out that a language might change its word order properties if a change in the order of subject, object and verb occurs. He showed that Papago (Uto-Aztecan; USA/Mexico) had dominant VS and VO order. The reason is the high frequency of definite noun phrases that always follow the

noun in Papago. This, however, seemed to be a new phenomenon since word order used to be OV and SV (the second being less certain) in Papago due to other discourse principles governing word order. Dryer now claimed that the word order change had led to an increasing number of prepositions and to a higher frequency of the genitive following the noun, both being typical properties of VO languages. The higher frequency of VO order would therefore have led to a change in other properties of the language which provides a strong argument for using the frequency criterion in this thesis.

## 3.2 Results

To check whether SOV word order and the presence of a case system interacted with each other, the data of Sample A was analyzed with a $\chi^2$-test. The calculation for the statistical test was performed in R. Table 1 shows how the data is distributed in a contingency table.

Table 1: SOV and case marking

| | Case system | No case system | |
|---|---|---|---|
| SOV + OSV | 321 | 76 | 397 |
| Other word orders | 214 | 279 | 493 |
| | 535 | 355 | 890 |

The test gives $\chi^2 = 124.73$ and $p < 0.0001$ which is a statistically significant result. One can therefore assume that SOV word order and the presence of a case system are not independent from each other. However, as I named in section 3.1. above, until now no efforts have been made to balance the sample. A brief examination of the languages in the sample reveal that of the 890 languages in total 101 are Austronesian. This means that the largest language family is responsible for roughly a ninth of the data. A proposition that has been put forward to prevent such phylogenetical biases is to count genera instead of counting languages (Dryer, 1992). Each genus would then be assigned one value for word order and one value for case marking. Unfortunately, this poses a huge problem with the data in the sample. To understand this problem, let us take a look at Celtic languages. In the sample Breton is found as SVO languages with a case system, Cornish as SVO language without a case system, Irish as VSO language with a case system, and Welsh as VSO language without a case system. If one now had to assign one value for word order and one value for case marking to the genus "Celtic", it is obvious that the choice would have to be arbitrary.

It was therefore necessary to look for another method to balance the sample phylogenetically. This was done by artificially reducing the number of data points for each language family to not more than 27. Thereby, the proportions for each field of the contingency table were maintained. Table 2 shows how the Austronesian languages were distributed in the original sample, whereas table 3 shows how the reduced data is distributed.

Table 2: Austronesian

| | Case system | No case system | |
|---|---|---|---|
| SOV | 7 | 5 | 12 |
| Not-SOV | 5 | 84 | 89 |
| | 12 | 89 | 101 |

Table 3: Austronesian reduced

| | Case system | No case system | |
|---|---|---|---|
| SOV + OSV | 2 | 1 | 3 |
| Others | 1 | 22 | 23 |
| | 3 | 23 | 26 |

This reduction was done for the following families: Afro-Asiatic, Austronesian, Eastern Sudanic, Indo-European, Niger-Congo, Pama-Nyungan, Sino-Tibetan and Trans-New Guinea. This resulted in a new sample with 674 data points. In this sample, no language family represents more than 4% of the data. It is thus better balanced than the original sample. The new data is shown in table 4.

Table 4: SOV and case marking - phylogenetically balanced

| | Case system | No case system | |
|---|---|---|---|
| SOV + OSV | 261 | 60 | 321 |
| Others | 174 | 178 | 352 |
| | 435 | 238 | 673 |

Also for this data a $\chi^2$-test was performed. It gave $\chi^2 = 71.79$ and $p < 0.0001$ yielding again a statistically significant result. Let us now take a look at the data from Sample B where additional information on the number of cases for each language is given. Figure 1 shows how many cases there are in the sample's SOV languages and in its Not-SOV languages. This second group was created by adding all other word orders. The sample does not contain any OSV language which would have to be grouped with the SOV languages for a category "verb-final languages". Sample B contains 88 SOV languages and 129 Not-SOV languages. The data from Figure 1 seems to confirm the original hypothesis. The vast majority of the Not-SOV languages does not have case marking, whereas the number of SOV languages without a case system is much lower.

Figure 1: Case marking and word order

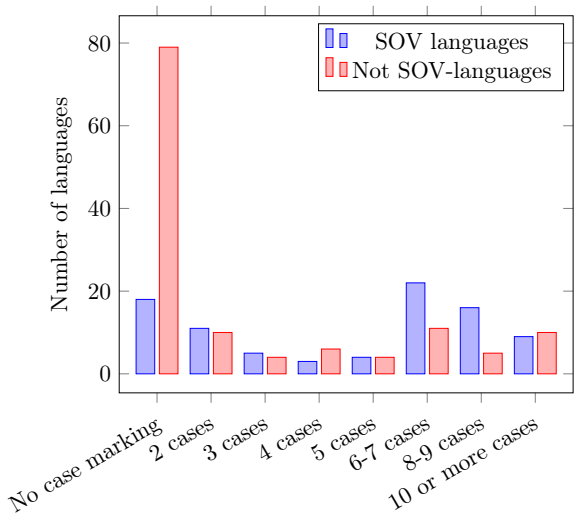



Figure 2: Ratio (Not-SOV)/SOV

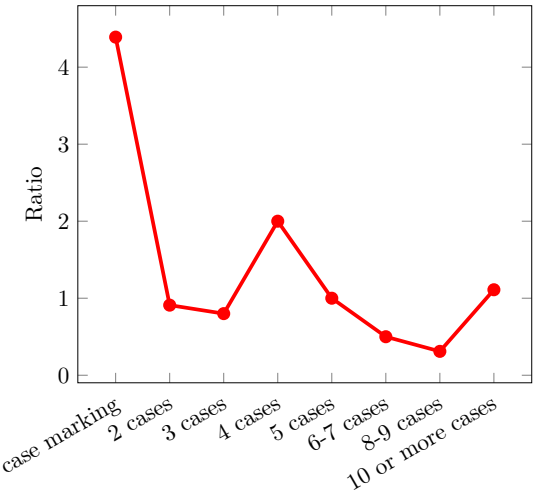


Figure 2 shows the ratio of Not-SOV languages and SOV languages with respect to the different number of cases. Although there is no uniform decrease, we can observe a tendency towards lower values for the ratio with an increasing number of cases in the case system. The values in the middle of the diagram (3 cases, 4 cases, 5 cases) should be taken with a grain of salt. Since we do not have many languages for these categories (see Figure 1), the small amount of data does not allow us to draw serious conclusions. What is however interesting is the rise in the last category (10 or more cases). It is difficult to think of an explanation why the ratio raises when languages with 10 or more cases are considered.

To check whether there is a correlation between the number of cases and the ratio shown in Figure 2, I calculated Spearman's Rho. For the calculation the values for case marking were transformed in cardinal numbers from 1 to 8. The calculation was again performed in R. The test gave $\rho = -0.4$ and $p = 0.32$ showing a negative correlation which does not reach significance. However, languages with no dominant word order were included in the analysis. Since these languages are also expected to have case marking, it might be worth to exclude them and to have a look at the evolution of the data. How the new data is distributed is shown in Figure 3 and 4 below.

Figure 3: Case marking and word order

No dominant word order languages excluded

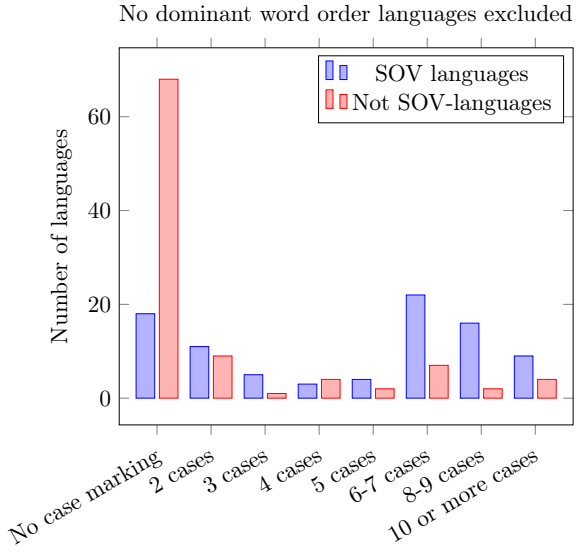
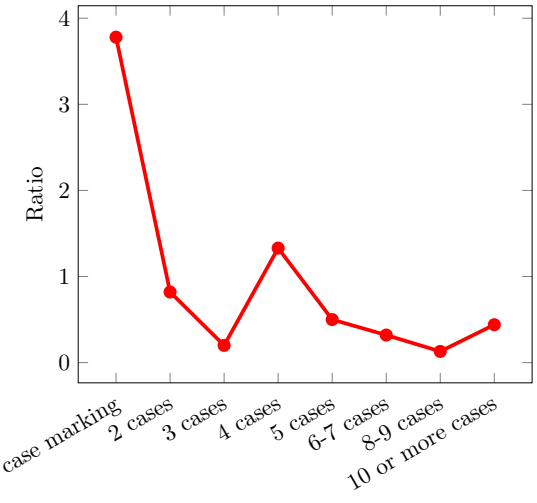



Figure 4: Ratio (Not-SOV)/SOV

No dominant word order languages excluded




In comparison with Figure 1 and 2 above, we do not see any big changes at first glance. However, there is a tendency for the values in Figure 4 to come closer to the x-axis for high number of cases. In general, the decrease seems to be a bit sharper than above. A correlation test seems to corroborate this impression. The calculation of Spearman's Rho gave $\rho = -0.62$ and $p = 0.115$. We now have a very strong negative correlation that is not far from reaching significance. These changed results are a hint that languages with no dominant word order might indeed have case marking. This will be investigated in more detail in the following section. For the remaining analyses of this chapter, I decided however to remove languages with no dominant order so that they cannot influence the results.

We can thus conclude that there is not only an overall correlation between SOV word order and the presence of a case system, but there also seems to be an increase in probability of SOV word order the more cases a language has. Again, it needs to be checked whether the observed correlation is due to a phylogenetical bias of the sample.

Unfortunately, this could not be done in the same way as with Sample A. The reason for this is quite simple. Since Sample B contains less languages and therewith less data points, but more categories (case marking is no longer a binary variable), a reduction similar to the one for Sample A would result in absurd values, many of them ranging between 0 and 1. I therefore decided to eliminate the language families with the most languages from the sample and to look at them separately. I removed all language families that made up more than 3% of the data (Indo-European (28 languages), Austronesian (22), Niger-Congo (14), Trans-New Guinea (10), Afro-Asiatic (9), Uralic(8)). These language families contain 91 languages of which only 24 had SOV as dominant order. Figure 5 shows the amount of languages with respect to case marking and word order. For Figure 6, again the ratio of Not-SOV languages and SOV-languages was calculated. Since the percentage of SOV-languages is quite low ($\frac{24}{91} = 26.4\%$) we expect higher values compared to the figures above.

Figure 5: Case marking and word order

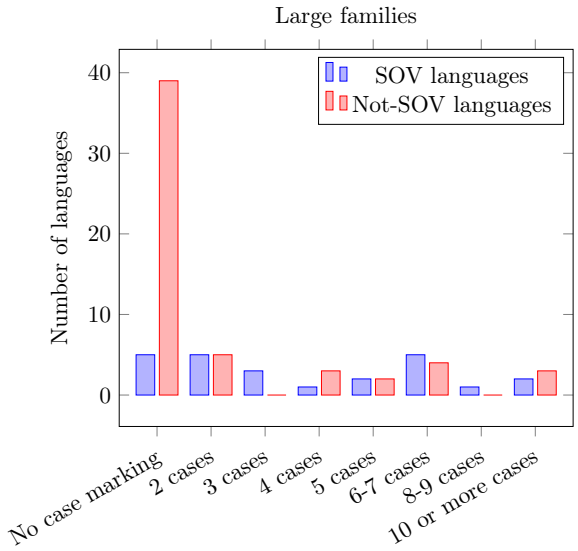
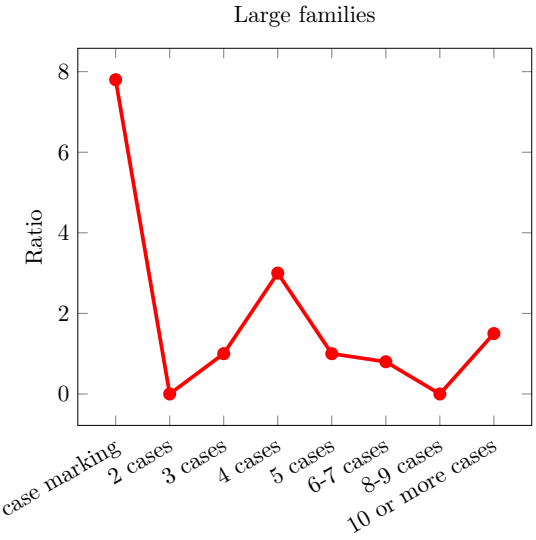



Figure 6: Ratio (Not-SOV)/SOV




Looking at Figure 5, we can easily recognize the high number of Not-SOV languages lacking case marking. The behavior of SOV languages becomes clearer if we look at Figure 6. The values are indeed visibly higher than in the preceding graphs. But what is more important is the tendency for the ratio to decrease with an increasing number of case. It is true that there is an increase in the middle of graph, but this might be due to chance effects resulting from the low number of languages in these categories. A correlation testing was omitted due to the low number of SOV languages. In general, both graphs behave as one would expect if the hypothesis is correct.

We need however to have a further look at the remaining languages in the sample and to check how the data behaves after removing the largest families. This is done in the two graphs below. Figure 7 shows the amount of SOV languages and the amount of other language types for each value for case marking. For Figure 8, the ratio (Other languages)/SOV languages is plotted.

Figure 7: Case marking and word order
Without large families

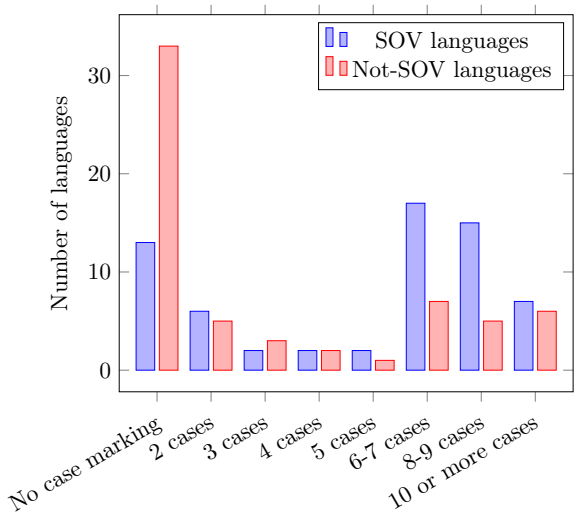
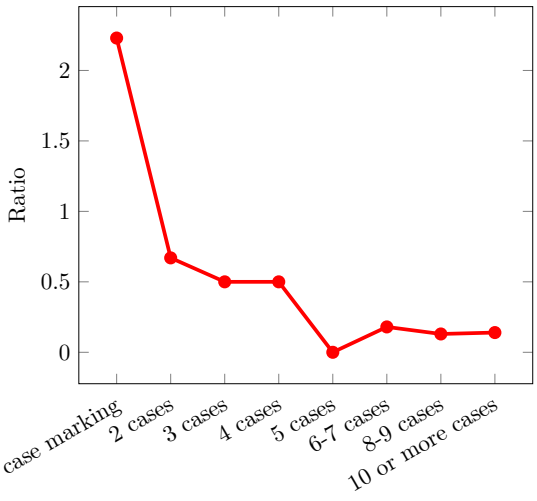



Figure 8: Ratio (Other languages)/SOV languages - Without large families




What is visible in Figure 7 reminds of what we saw in Figure 1. Again we have many languages not being SOV and lacking a case system, only a few language with a moderate case system (3-5 cases) and many SOV languages with a large case system (more than 5 cases). The high values for the other languages in these categories might be due to the effect of the languages without a dominant word order. They are examined more thoroughly in the following section. The graph in Figure 8 shows a very interesting slope. The low values for the ratio ranging between 0 and 2.23 are explained by the higher amount of SOV languages (this time 64 out of 126 languages had SOV as dominant word order). The tendency for the ratio to decrease is obvious in Figure 8. What is new is the lacking of runaway values. We have a sharp decrease in the beginning, then three data points with similar values and finally, after another sharp decrease, four data points with similar values. Anew, Spearman's correlation test was performed and the calculation gave $\rho = -0.82$ and $p = 0.01$. This time, we observe a very strong negative correlation reaching significance.

Summarizing the results, we can say that statistical testing revealed that SOV word order and the presence of case marking are not independent from one another. This effect was visible both in the overall Sample A and in a reduced sample balancing possible phylogenetical biases. A more in-depth examination revealed that there is not only a significant correlation between SOV word order and the presence of case marking, but that also the probability of a language being SOV increases the more cases a language has. This effect was not only visible in the overall Sample B, but also if the largest language families were excluded from the analysis. The effect became even significant after excluding the largest families.

### 3.3 A closer look at free word order languages

After we have seen the clear correlation effects between SOV word order and the presence of a case system, this chapter will focus on languages with no dominant word order. According to Sinnemäki (2008), these languages should have a case system significantly more often than other languages. Although his results are clear and his explanations are reasonable, it makes sense to try to replicate his findings. The reason for this lies in the sampling procedure. As described in section 2, Sinnemäki (2008) used a sample containing 50 languages from 50 different genera. The problem with this methodology is that the selection of genera represented in the sample and the selection of a language for each genus can have a strong effect on the final result (see also Section 3.2.). A reasonable alternative to this procedure is to perform a statistical analysis as I did for SOV word order and case marking. Therefore, the data of Sample A is shown in the contingency table in Table 5.

Table 5: No dominant word and case marking

| | Case system | No case system | |
|---|---|---|---|
| No dominant word order | 99 | 39 | 138 |
| Dominant word order | 436 | 316 | 752 |
| | 535 | 355 | 890 |

I performed a $\chi^2$-test for this data. Again, the calculation was done in R. The result ($\chi^2 = 8.64$ and $p = 0.0032$) shows a statistical significant effect that allows to conclude that languages having no dominant word order and languages having a case system are not statistically independent. Nonetheless, the problem is the same as in the section above: the sample is a convenience sample that might be strongly phylogenetically biased. Therefore, the sample was reduced in the same way as described in section 3.2. The values for the reduced samples are shown in the contingency table in Table 6.

Table 6: No dominant word and case marking - phylogenetically balanced

| | Case system | No case system | |
|---|---|---|---|
| No dominant word order | 87 | 27 | 114 |
| Dominant word order | 348 | 212 | 560 |
| | 435 | 239 | 674 |

Again, independence of languages with no dominant word order and languages with a case system was tested with a $\chi^2$-test. Also in this balanced samples, the calculation showed a statistically significant correlation ($\chi^2 = 7.7$ and $p = 0.0055$).

Subsequently, a look was taken at Sample B and at how the languages with no dominant word order are distributed over the different values for the variable case marking. Figure 9 below shows the number of languages with no dominant word order for each value for case marking. In total, the sample contains

32 languages with no dominant word order. Figure 10 below shows the probability for languages with no dominant word order for each value for case marking. The probabilities were calculated by diving the number of languages with no dominant word order in each category by the total number in the respective category. There are for example six languages with no dominant word order that have a case system with ten or more cases. In total, Sample B contains 19 languages in this category. The probability was thus calculated in the following way: $\frac{\text{Languages with no dominant word order and ten or more cases}}{\text{Languages with ten or more cases}} = \frac{6}{19} = 0.32$

Figure 9: Languages with no dominant word order

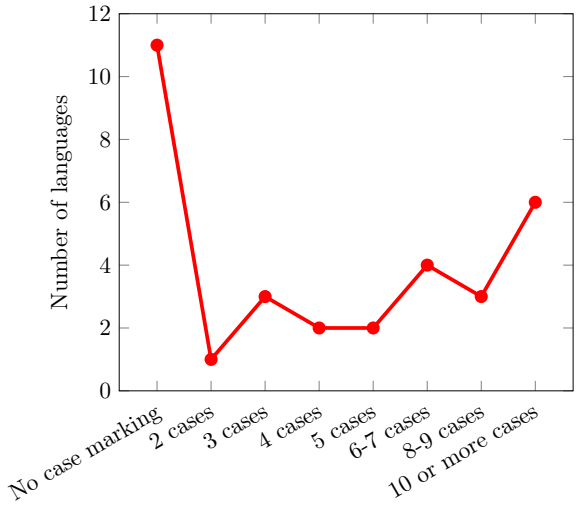
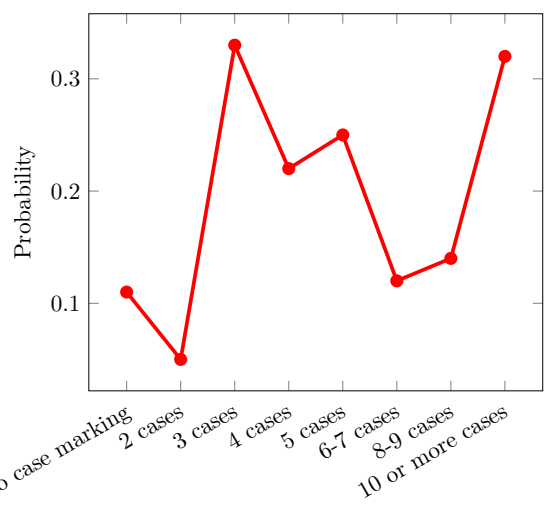



Figure 10: Probabilities for languages with no dominant word order




What we see in Figure 9 is a somewhat unexpected trend. The highest value is reached in the first category for languages with no morphological case marking. Still, the $\chi^2$-test showed that there is a significant correlation for languages with no dominant word order to have case marking. A possible explanation for the unexpected trend in Figure 9 could be that the trend is due to a phylogenetical bias of the sample. Of the eleven languages with no dominant word order that lack a case system, more of the half, namely six, are Austronesian. Also a look at Figure 10 provides arguments for the hypothesis that languages with no dominant word order tend to have a case system more often than other languages. The curve of the probability for a language not to have a dominant word order shows an overall tendency to rise with an increasing number of cases. Again the high value for the first category is partly due to the Austronesian languages. A correlation test was performed on the data of Figure 10, but this time no significant correlation could be established ($\rho = 0.429$ and $p = 0.3$). However, one sees that the correlation is clearly positive which might provide another argument for the initial hypothesis. If we try to balance the sample as it was done in the section above, excluding the most prominent language families, we obtain the following results:

Figure 11: Languages with
no dominant word order
Large families excluded

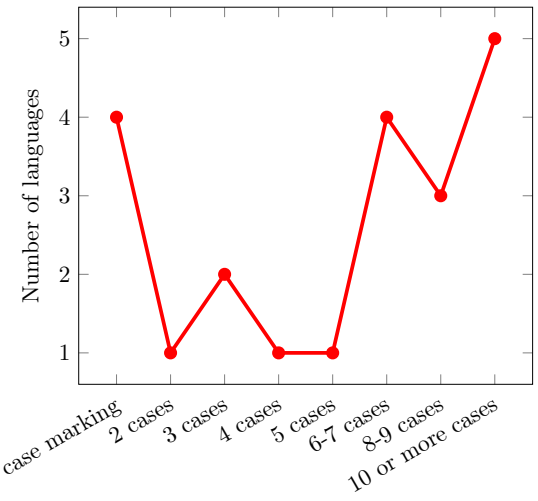



Figure 12: Probabilities for languages
with no dominant word order
Large families excluded

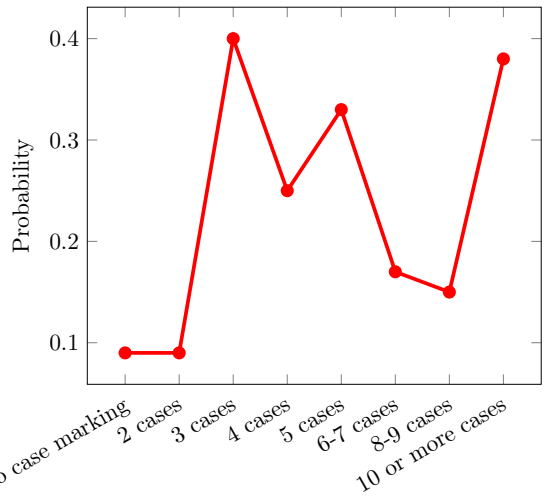



Figure 11 shows a similar trend to Figure 9 above. Figure 12 has, however, lost the initial peak that could be observed in Figure 10. In Figure 12, we can now observe a more regular increase and, more importantly, no value is any longer lower than the value for the first category.

What this section has therewith shown is that there is an association between languages with no dominant word order and languages with a case system. The findings from Sinnemäki (2008) could be replicated by showing statistical independence with a $\chi^2$-test. What is interesting is that the correlation was by far not as strong as the correlation between SOV word order and the presence of a case system, that was described in Section 3.2. Neither could there be established a correlation for an increasing probability of no dominant word order for an increasing number of cases, which constitutes an interesting difference to the analysis from the previous chapter. Nevertheless, what this chapter should have made clear is that both languages with SOV word order and languages with no dominant word order have a case system significantly more often than other languages.

# 4 Robustness of communication

In chapter 3, we could observe a clear tendency for both SOV languages and languages with no dominant word order to have a case system. Of course, this observation requires an explanation. In the following chapters, I will present three different factors that might cause the observed phenomena. These factors are not mutually exclusive, but do probably all contribute. The content of the present chapter is thought to be an explanation for the presence of a case system in both SOV languages and in languages with no dominant word order. The following chapters will then be dedicated to SOV languages only.

In the beginning, the question why languages with no dominant word order tend to have a case system will be of interest. Already Sinnemäki (2008) proposed an explanation in his paper. He explained the correlation by a trade-off between the different strategies that languages can use to encode what is subject and what is object in a sentence (word order, case marking and verbal agreement). A language with a fixed word order would not need a case system since the ordering of the constituents would allow to infer what is subject and what is object in a sentence. This is of course not possible in a language that allows free ordering of the constituents. Such a language would thus need to use other means in order to make sure that sentences are not misunderstood.

Further arguments for this hypothesis come from (Fedzechkina et al., 2017). They performed a study in which participants learned artificial languages. The participants were divided in two groups: one half learned an artificial language with rigid SOV word order, the other half learned a language in which 50% of the sentences had SOV word order and 50% OSV word order. Crucially, in this second language word order did not provide any clues for telling subject and object apart. Both languages contained six nouns, four verbs and a case marker. Case marking was however optional in both languages, in two thirds of the sentences the object was marked by the case marker "kah" (Fedzechkina et al., 2017, 6).

All participants were then exposed to the language for three days. During these days they performed different production and comprehension tests. What is most important in this context, is the data from the production tests. After three days of learning, learners of the fixed word order language had lowered case marking to an occurrence of 41%. On the other hand, the learners of the flexible word order language used case marking in 71% of the sentences. There could thus be observed a tendency to use case marking more frequently in the language where constituent order was not fixed.

The explanation Fedzechkina et al. propose is a trade-off between production effort and conditional entropy, an idea very similar to what was proposed in Sinnemäki (2008). Fedzechkina et al. (2017) observed that 17 of the 20 learners of the language with flexible word order lowered the conditional entropy in the output language. Sentences with case marking and/or fixed constituent order had an entropy of 0 bits since no uncertainty about the syntactic roles could arise. The only sentences with a higher value for the conditional entropy were the sentences of the flexible word order language that did not have case marking; these sentences had an entropy of 1 bit. It was further observed that the mean production effort (defined as the average number of syllable per sentence) increased in the language with flexible word order, whereas it

was lowered in the language with fixed word order. This confirms the idea of the trade-off: in a language with no uncertainty, the overall production effort is lowered, but it might in contrast be increased (e.g. by the use of case marking) if such an increase helps to lower uncertainty. Hence, the avoidance of redundancy could be the mainspring for the correlation between flexible constituent order and the presence of case marking.

Network simulations seem to confirm this idea. Lupyan and Christiansen (2002) performed a simulation with five different artificial languages varying according to flexibility of their word order. Each of these five grammars was combined with different frequencies of case marking, ranging from no case marking at all to case marking in every sentence without any syncretism. The results of their simulation showed that the more flexible a word order was, the more it benefited from case marking, in terms of correctly interpreted words. On the other hand, (nearly) no improvement in correct interpretation was achieved in the languages having a more fixed word order. Speaking in terms of Fedzechkina et al. (2017), no entropy reduction was achieved, but production effort did clearly increase. It makes thus sense to assume that languages with fixed word order will lower production effort (by not having case marking) since this does not result in an increase in entropy. Communicative goals can still be successfully achieved.

This explanation may in part also be helpful for the presence of case systems in SOV languages. It might be that SOV languages allow more flexibility in word other than other languages. Recall that in SOV languages, SOV is just the dominant, but not the only possible word order. Indeed, Bentz and Christiansen (2013) have shown that at least in Latin SOV was flexible, whereas Modern Romance languages show quite rigid SVO word order. More research would be needed to clarify whether languages of other language families show a similar tendency. What can surely be asserted is that there are also other reasons to assume that robustness of communication can be an explanation for the frequent presence of case systems in SOV languages.

Bartsch and Vennemann (1983, 178) claimed that topicalization of the object could be a problem in SOV languages if they lack case marking. Topicalization is thereby understood as bringing focus on the object by putting it at the beginning of a sentence. This process would result in an OSV order with the object being the topic of the sentence. Bartsch and Vennemann argued that this process of topicalization would lead to confusion about what is subject and what is object if case marking is absent. To make this point clearer let us take a look at the following examples of German subordinate clauses:

(5) dass der Junge den Mann schläg-t.
that the.NOM boy.NOM the.ACC man.ACC beat-3.SG
"that the boy beats the man"

(6) dass den Mann der Junge schläg-t.
that the.ACC man.ACC the.NOM boy.NOM beat-3.SG
"that the boy beats *the man*"

Sentence (5) has canonical SOV word order that is typical for German subordinate clauses. Case marking on both arguments allows to infer what is subject and what is object even in sentence (6) where the object *den Mann* is put in front, yielding a pragmatically marked OSV order. Still, this order does not provoke any confusion about the syntactic roles. Now in German, only masculine nouns distinguish nominative and

accusative in the singular. Feminine and neuter nouns show case syncretism which makes nominative and accusative indistinguishable. Having said that, let us have a look at the following sentence:

(7) dass das Kind die Frau schläg-t.
that the.NOM child.NOM the.ACC woman.ACC beat-3.SG
"that the child beats the woman"

For the interpretation of sentence (7) one would have to think of *das Kind* as being the subject and of *die Frau* as being the object. If one now tries to topicalize the object paralleling what was done in sentence (6), the result is the following:

(8) dass die Frau das Kind schläg-t.
that the.NOM woman.NOM the.ACC child.ACC beat-3.SG
"that the woman beats the child"
*not*: "that the child beats *the woman*"

This interpretation differs radically from the interpretation of the one of sentence (6) above. Here, topicalization was not achieved. To place *die Frau* in front did rather provoke an inversion of the syntactic and semantic roles. The constituent *das Kind* is no longer the subject of the sentence and the result of topicalizing the object ended up in transforming it into the subject of the sentence. According to (Bartsch and Vennemann, 1983, 178) this phenomenon might be a reason for caseless SOV languages to shift to SVO word order. In those languages, the verb intervenes between the subject and the object and bringing the object to the beginning of the clause would not yield confusion about the syntactic roles.

The observation is of course a good hint to understand the necessity of a clear assignment of syntactic roles, but is not strong enough to account for the frequent presence of case systems in SOV language. Even Bartsch and Vennemann (1983) have to admit that the subject occurs in sentence-initial position most frequently because it is the topic of the sentence in the majority of the cases. Topicalization of the object might thus not occur often enough to provoke a shift in word order. However, it has been claimed that even in canonical SOV sentences, word order alone might not be strong enough to assure a clear assignment of syntactic roles because the hearer cannot know whether the speaker tries to topicalize a constituent not equal to the subject.

This is for example shown in studies with Japanese native speakers. Yamashita (1997) showed that Japanese native speakers use case marking in scrambled sentences to infer what is the subject and what is the object of a sentence. Even if a constituent other than the subject occurs at the beginning of a clause, native speakers will be able to infer the syntactic role from the case marking the constituent received. This would of course not be possible if case marking was not available, which might lead to an increase of misunderstandings.

Also, Kurumada and Jaeger (2015) performed a study with Japanese native speakers. They conducted different experiments to see which semantic and pragmatic clues condition optional case marking in Japanese. Optional case marking is a phenomenon where (a) case marker(s) can be left out in casual speech. This phenomenon can be illustrated in the following sentence:

(9) Sensei-ga shobosha(-o) mi-ta.
teacher-NOM fire.engine(-ACC) see-PAST
"The teacher saw a fire engine"

(Kurumada and Jaeger, 2015, 155)

In Japanese, it is possible to omit the accusative case marker -o on *shobosha* because the sentence is still understandable. From the other constituents it is clear that *shobosha* has to be in the accusative. Hence, it is not necessary to repeat this piece of information on the object.

Kurumada and Jaeger (2015) presented sentences to Japanese native speakers both with and without the accusative case marker and asked them to repeat the sentences. This was done by showing two sentences to the participants that they were asked to remember. Afterwards, they were given the verb of one of the sentences and were asked to repeat the sentence belonging to that very verb. Crucially, the objects of the two sentences differed in terms of animacy. They were either inanimate as in sentence (8) above or animate and human. What Kurumada and Jaeger (2015) found was a clear preference for using the accusative case marker if the object of the sentence was animate (e.g. *seito-o - student-ACC*).

The same experiment was repeated with sentences with two animate arguments. In this case, the sentences differed with regard to plausibility. A sentence with low plausibility was for example:

(10) Hannin-ga keisatsu(-o) yonaka-ni taihoshi-ta-yo.
criminal-NOM police.officer(-ACC) night-at arrest-PAST-SFP
"The criminal arrested the police officer in the middle of the night"

(Kurumada and Jaeger, 2015, 159)

The reversion of the arguments in sentence (10) resulted in a sentence with high plausibility. The experiment was conducted in the same way as the one with the non-animate objects. Also the results were similar: it was found that in sentences with low plausibility, marking of the object with the accusative case marker was more frequent than in sentences with high plausibility. Both findings can be explained with a trade-off between production effort and a robust conveyance of information. In sentences where the object could theoretically also be the subject, case marking is needed to distinguish the two arguments. It seems thus sensible to assume that case marking is often found in SOV languages due to phenomena like topicalization and scrambling. Additionally, some SOV languages may have optional case marking (as Japanese) where case markers can be left out in special circumstances.

This explanation accounts already better why we find case marking in SOV languages, since it does not only look at sentences with a topicalized object. However, if in sentences where both of the verb's arguments could in theory be the subject of the sentence, case marking is needed to encode the syntactic roles, one has to ask why case marking is significantly less frequent in SVO and VSO languages. Let us first have a look at SVO languages, VSO languages will be discussed in further detail in the following chapter.

A good reason for why SVO languages often might lack case marking is given by Gibson et al. (2013). They performed gesture experiments with native speakers of English, Japanese and Korean. All participants were

shown short videos whose actions they were thereafter asked to gesture. The results showed an interesting picture: English participants used SOV order in 68% of the sentences with an inanimate patient and SVO order in 71% of the sentences with an animate patient. These latter sentences are reversible in the sense that both arguments could be the subject of the clause. Since pragmatics is not reliable for assignment of the syntactic roles, one would expect morphosyntactic features as word order or case marking to play a greater role and this seems indeed to be the case.

Japanese and Korean speakers used SOV order in all sentences, be the patient of the action animate or inanimate. However, a second gesture task provided insightful results for these groups of participants. They were asked to gesture a more complex action, including an embedded clause. The verb of the embedded clause was transitive, whereas the verb of the main clause was intransitive. What is interesting is that in the embedded clauses, Japanese native speakers used SVO order in 34% of the cases and Korean native speakers in 43% if the object of the clause was animate. Remarkably, Gibson et al. (2013) is not the only study noticing this change of word order. Futrell et al. (2015) repeated the findings with native speakers of English, Irish, Tagalog and Russian. Also in their study, SOV was the most frequent word order for non-reversible events. On the other hand, SVO order had a significant increase in frequency for reversible events where it was the most frequent word order. This shows that also native speakers of VSO languages tend to gesture events in the manner described above.

Now, what does this tell us? The use of SOV order without case marking was unproblematic as long as it was possible to rely on pragmatic information such as animacy to distinguish subject and object. When both arguments were animated and case marking was not available, the participants switched to SVO order to avoid misunderstandings. This implies that SVO is more robust in communicative terms than SOV if a language does not have case marking. A possible reason for this observation is the fact that the verb intervenes between the two arguments.

Another point that is worth mentioning is that Gibson et al. (2013) found that although no codified case markers were available in the gesture tasks, participants sometimes tried to add case marking to the arguments of the verb. For this purpose, they used "location in space" (Gibson et al., 2013, 1085). For all three groups of native speakers, English, Japanese and Korean, a tendency could be observed to use spatial clues with a higher probability if word order was SOV. The exact numbers are shown in Table 7 below.

Table 7: Spatial clues in gesture tasks

| | Experiment 1 - English speakers | | Experiment 2 - Japanese speakers | | Experiment 2 - Korean speakers | |
|---|---|---|---|---|---|---|
| Word order | SOV | SVO | SOV | SVO | SOV | SVO |
| Spatial clues present | 23 | 13 | 40 | 11 | 18 | 10 |
| Spatial clues absent | 15 | 94 | 17 | 18 | 32 | 28 |

(Gibson et al., 2013, 1086)

Furthermore, Futrell et al. (2015) try to establish a connection with the so called "noisy-channel-hypothesis". This hypothesis takes into account that people never use language in a completely silent environment. Non-linguistic noise might corrupt the signal and hamper successful communication. If both arguments are case marked in SOV languages, the loss of one of the arguments due to environmental noise is not a big problem. Thanks to the case marker, the hearer can still infer the syntactic role of the received constituent. In SVO languages, the hearer does not need case marking as he will know that the received constituent is the subject if it precedes the verb and the object if it follows the verb.

This chapter has shown that case marking in languages with no dominant word order and in SOV languages plays an important role to assure successful communication. If a languages does not have a fixed word order that allows to infer what is the subject and what is the object, it will need other means. Case marking may then be used. In SOV languages, due to topicalization or scrambling, confusion might arise about the syntactic roles of the arguments. This holds especially for reversible events when pragmatics cannot rule out a misleading interpretation. Moreover, case marking can minimize the negative effects for communication that may arise from loss of one constituent due to a noisy environment. We thus see that case marking has a bunch of different functions ensuring that the hearer can understand the message of the speaker correctly. These functions play a great role in SOV languages and in languages with no dominant word order and might hence account for the frequent use of case marking in these languages.

# 5 Early accessibility of information

We have seen in chapter 4 how communicative reasons might be responsible for languages with no dominant word order and SOV languages to have case marking more frequently than other languages. This became particularly evident with regard to SVO languages. However, the robustness of communication described in the chapter above might not be the only factor related to communicative reasons that could lead to a preference for case marking in SOV languages. In this chapter, I want to present another explanation for the fact that SOV languages so often have a case system. This chapter will especially focus upon the question why the percentage of SOV languages having a case system is significantly higher than the one of VSO languages. In this chapter and in the following one, I will disregard languages with no dominant word order. These have been discussed in the previous chapter and are generally much more discussed in the literature than SOV languages (Sinnemäki (2008), Fedzechkina et al. (2017)).

This chapter will concentrate on the idea that the world's languages tend to present the most important information as early as possible. This claim has not been developed in the present thesis, but is adopted from John Hawkins' Maximize On-line Processing (MaOP): "The human processor prefers to maximize the set of properties that are assignable to each item X as X is processed, thereby increasing O(n-line) P(roperty) to U(ltimate) P(roperty) ratios. The maximization difference between competing orders and structures will be a function of the number of properties that are unassigned or misassigned to X in a structure/sequence

S, compared with the number in an alternative" (Hawkins, 2004, 51). The On-line Properties are those that can be assigned when a word is parsed, the Ultimate Properties are the sum of all On-line Properties. Hawkins' framework is quantitative and allows to calculate which structures should be preferred within a certain grammar and which types of grammar should be preferred over others.

Let us look at an example to see how MaOP predicts structures within a certain grammar. Hawkins contrasts the sentences "I realize that the small young boy knows the answer" and "I realize the small young boy knows the answer" and compares whether the variant with or without the complementizer would be favored by the MaOP principle. A look at the proceeding will be helpful to understand how the quantitative nature of MaOP looks like:

| **Sentence A** | the | small | young | boy | knows |
|---|---|---|---|---|---|
| **Catgories** | Det | Adj | Adj | N | $V_2$ |
| **Phrases** | $NP_2$ | | | | $VP_2$,$S_2$ |
| **Attachments** | np2[Det] | np2[Adj] | np2[Adj] | np2[N] | vp2[v],s2[$VP_2$], vp1[$S_2$],s2[$NP_2$] |
| **Relations** | | | | | $S_2$=OBJ-V1, $NP_2$=SUBJ-$V_2$ |
| | 3 | 2 | 2 | 2 | 9 |
| **OP/UP ratio** | $\frac{3}{18} = 17\%$ | $\frac{5}{18} = 28\%$ | $\frac{7}{18} = 39\%$ | $\frac{9}{18} = 50\%$ | $\frac{18}{18} = 100\%$ |

| **Sentence B** | that | the | small | young | boy | knows |
|---|---|---|---|---|---|---|
| **Catgories** | Comp | Det | Adj | Adj | N | $V_2$ |
| **Phrases** | $S_2$ | $NP_2$ | | | | $VP_2$ |
| **Attachments** | s2[Comp] vp1[$S_2$] | np2[Det] s2[$NP_2$] | np2[Adj] | np2[Adj] | np2[N] | vp2[v] s2[$VP_2$] |
| **Relations** | $S_2$=OBJ-V1 | | | | | $NP_2$=SUBJ-$V_2$ |
| | 5 | 4 | 2 | 2 | 2 | 5 |
| **OP/UP ratio** | $\frac{5}{20} = 25\%$ | $\frac{9}{20} = 45\%$ | $\frac{11}{20} = 55\%$ | $\frac{13}{20} = 65\%$ | $\frac{15}{20} = 75\%$ | $\frac{20}{20} = 100\%$ |

| **Unassignment** | Word #1 | #2 | #3 | #4 | #5 | #6 |
|---|---|---|---|---|---|---|
| **Difference:** | | | | | | |
| **ZERO** | 17% | 28% | 39% | 50% | 100% | 100% |
| **THAT** | 25% | 45% | 55% | 65% | 75% | 100% |
| | +8 | +17 | +16 | +15 | -25 | |

→ THAT preferred: 56-25=31

(Hawkins, 2004, 57)

In the table above, we see how the calculation of MaOP works. For each words, the properties that can be assigned when the word is parsed are given below. These properties are subdivided as follows: categories (which part of speech does the word belong to?), phrases (which phrases can be constructed when the word is parsed?), attachments (what is the word's function within the phrase?) and relations (what function does the phrase have in the sentence?). We see that the construction of a phrase does not necessarily mean construction of the respective relation. For example, parsing of *the* in the sentence A allows to analyze the parsed word as a determiner, to construct a noun phrase and to attach the determiner to this new noun phrase. Observe that $NP_2$ can be constructed when *the* is parsed, but constructing $NP_2$ as the subject of the verb *knows* is delayed until the verb is parsed and the parser knows of the existence of a subordinate clause. An earlier assignment is not possible since $NP_2$ could theoretically also be the object of the main clause.

After the syntactic analysis, for each word the number of properties (construction of phrases, attachments, relations) that can be assigned after parsing the word is divided by the overall properties that will be assigned in the course of the clause. If we remain at the example *the* from sentence A, we count three properties that could be assigned. Since at the end of the sentence a total number of eighteen properties has been assigned, we have to divide three by eighteen. This proceeding is repeated for each word in the sentence. The comparison between the two sentences is done by comparing the percentage of assigned properties for each word and calculating the difference between the sentences for each word. MaOP predicts the sentence with a positive margin to be preferred, in this case the second one.

According to Hawkins, the crucial difference between the two sentences above is the fact that the hearer understands earlier that the subordinate clause is the object of the main clause if the complementizer is present. If the complementizer is missing, the hearer can construct the structure only when he receives the verb, and thereby much later. Moreover, the presence of a subordinate clause as such will become clear at an earlier stage if the complementizer is there. Attachments and relations can be assigned earlier which is shown in the calculation. This is why Hawkins argues that the second sentence should be favored.

This example should have made clear how MaOP works. For the purpose of this thesis, is not necessary to focus too much on the calculation of MaOP, but more instructive to have a look at its effects. Hawkins claims that MaOP could be an explanation to the preference for subjects to occur before objects. He argues that objects depend on subjects for a variety of assignments (e.g. anaphorical binding in a sentence like $\text{John}_i$ washed $\text{himself}_i$). Some of the object's properties could thus only be assigned if the subject is already known. Otherwise, these properties would have to remain temporarily unassigned. But since MaOP predicts that sentences with an earlier property assignment are favored, this would mean that subjects should indeed tend to appear before objects (Hawkins, 2007, 103).

The essential point for the present thesis is that MaOP can also be related to the correlation between SOV word order and case marking. Since the first phrases that will be received in SOV languages are noun phrases, MaOP would predict that those should be mophologically marked so that properties and relations can be assigned as early as possible. Let us recall that the possibilities for distinguishing the arguments of

the verb were basically three: word order, case marking and verbal agreement. We have seen in chapter 4 that word order is a somehow problematical means to use in SOV languages. If we now consider the idea behind MaOP, we understand that verbal agreement is not so helpful either. The verb occurs too late in the sentence and the hearer wants to construct the sentence structure as early as possible. Case marking is therefore the most functional means and, as we have seen in chapter 3, it is indeed very widespread in SOV languages.

Interestingly, MaOP allows us to explain why VSO languages are less likely to have case marking than SOV languages. If the verb does not occur last but first in the clause, then it makes sense to mark the verb and to use verbal agreement as coding strategy. In fact, it has been claimed that VSO order would correlate with rich verbal morphology (Nichols, 1992, 105). This symmetry seems to provide further arguments that the idea of MaOP is on the right track.

The preference for hearers to prefer an early occurrence of information has not only been observed in natural languages' grammars. Fedzechkina et al. (2015) tested 56 monolingual English participants to see whether they would show this preference in artificial language learning. Participants were divided in two groups: one group learned a language in which only the subject was case marked, whereas the other group learned a language in which only the object was case marked. Both languages had flexible order, SOV and OSV occurring equally often. The languages had furthermore optional case marking which was present in 67% of the sentences in the input language.

Sentence production tests revealed that case marking was more frequent in the learners' output than in the given input. This was valid for both languages. What was also found was a tendency to mark the first constituent for case more frequently that the second constituent which was significant in the object-marking language. The reason might be the preference of disambiguating sentence structures as early as possible and thus of providing case marking already on the first constituent.

Network simulations seem to further enforce this idea. Lupyan and Christiansen (2002) tested a network on its ability to correctly analyze sentence structures. In their simulations, they used different word orders both with and without case marking. Hardly surprisingly, they found that SOV grammars without case marking performed poorly. What is more interesting is that they attribute this result to the fact that the network was probably not able to distinguish the two unmarked nouns prior to the verb. They argue that the problem for caseless SOV languages is that the most informative constituent, the verb, is received last.

Van Everbroeck (2003) performed a more detailed network simulation coming to similar results. She trained a network by showing it sentences from an artificial language corpus. After 30 training sessions, the network was tested on its ability to correctly analyze sentences. Thereby, it had also to generalize to sentences it had never seen before. Not surprisingly, the network obtained satisfying results for SOV languages in the test sessions only if the languages had case marking. For SOV languages lacking case marking, the results were on the contrary pretty poor. The network was not able to reliably assign the correct syntactic roles to the constituents in the clause even in simple grammars containing only subject, object and verb. The results

became slightly better if head marking on the verb was added, but were still far from being satisfying. This is understandable since information on syntactical roles came late in the clause and the network did not have the possibility to rectify misassignments that had occurred earlier.

A particularly instructive aspect in Van Evenbroeck's study is that she did not only simulate simple grammars containing subjects, objects and verbs. In a second simulation she introduced a genitive and locative phrases into the grammars. Locative phrases were built up of a locative noun (in the locative case if case marking was present) and an adposition. SOV grammars without case marking now had difficulties to correctly analyze the second noun in a clause if case marking was missing. Since there was no case morpheme providing a clue for the noun's syntactic function, it could be both an object or a locative phrase. Let us have a look at a natural language example from Turkish to understand this problem:

(11) Hasan-Ø araba-yı bekli-yor-Ø
Hasan-NOM car-ACC.DEF wait-PRES-3.SG
"Hasan is waiting for the car"

(12) Hasan-Ø araba-da bekli-yor-Ø
Hasan-NOM car-LOC wait-PRES-3.SG
"Hasan is waiting in the car"

Turkish has case marking on nouns and this is crucial for the hearer to parse the sentences. When he receives the word *arabayı* in sentence (11) he can infer from the marker for definite accusative that it has to be the object of the sentence. On the other hand, in sentence (12) when he receives *arabada* with a locative case marker, he will immediately understand that it cannot be the object of the clause. If the hearer lacked this information as the network for caseless SOV languages in Van Everbroeck (2003), it would be temporarily impossible to assign a syntactic role to the second noun. He had to wait until the end of the clause when the verb is parsed. Grammatical relations would without case marking be unassigned for a longer period, what MaOP predicts to be unfavored. This prediction is born out in Van Everbroeck (2003).

In a third and last simulation, Van Evenbroeck also introduced relative clauses into the languages' grammars. For SOV languages she introduced relative clauses lacking a relative pronoun or particle, occurring before their heads and having the verbs in a non-finite form. She thereby tried to simulate what she claimed to be the typical patterns for SOV languages. Of course, this is only schematic since many SOV languages have the relative clause following the head noun. Still, this proceeding is an approximation and allows to see what happens if the relative clause precedes the noun. If it follows the noun, then the grammar should not work significantly different than the one for the second simulation.

Relative clauses preceding the noun again proved to be a severe problem for SOV languages without case marking. Van Everbroeck (2003) traces this difficulty back to an increase in ambiguity with respect to the former simulation. With relative clauses in the grammar, the first constituent of each clause could be either the subject of the main clause or the subject, the object or a locative phrase of the relative clause. Without case marking one would again have to wait until the verb is mentioned and could assign syntactic roles only at that point. This would again result in a delay in parsing the sentence.

What we have seen in this section is that in the communicative process not only clarity is clearly preferred (as seen in chapter 4), but that also grammars are favored that provide this clarity earlier than others. In a typological perspective, this could lead to the presence of case marking in SOV languages, that allows the hearer to assign the syntactic roles of the arguments at an earlier stage in the sentence. We have also seen that this principle does not only relate to core argument marking, but that early disambiguation might also be a factor favoring the presence of (a) locative case(s). This could be a first hint to the fact that in chapter 3 a correlation was found between an increasing number of cases and an increasing probability for a language to be SOV. This very correlation will be further discussed in the next chapter.

# 6 Grammaticalization processes

In the previous two chapters, I have tried to explain the results from chapter 3 mainly by means of communicative reasons. In this regard, the upcoming chapter will differ from the previous ones. I want to propose a further explanation for the high probability of SOV languages of having a case system by considering diachronic developments that can change the structure of a language over time. I will concentrate on grammaticalization theory and its impacts on grammar.

## 6.1 What is grammaticalization?

Before it is possible to look at how grammaticalization and the presence of case marking in SOV languages could interact, it is necessary to make clear what is understood by the term grammaticalization. Heine and Kuteva (2007, 32) propose the following definition: “Grammaticalization is defined as the development from lexical to grammatical forms, and from grammatical to even more grammatical forms.” They propose a scale along which grammaticalization processes work: free form $>$ clitic $>$ affix

This would mean that any free word, for example a noun or a verb, could develop over time to a more functional word (a clitic, a preposition, ...) and then end up in becoming an affix attached to another word. Heine and Kuteva (2007) claim that this diachronic process can be subdivided into four different stages: extension, semantic bleaching, decategorialization and phonological erosion. These stages follow a diachronic order. In the following, each of the stages will be briefly discussed to make clear what they comprise and what happens during grammaticalization.

Extension is the acquisition of new meanings for a specific linguistic construction. Those new meanings are acquired by an extended use of that very constructions in contexts that were not available for it before. Extension is thus a change in pragmatics. It is however inseparably connected to semantic bleaching or desemanticization, a semantic process. This second process involves the loss or change in meaning that arises from the usage of a linguistic construction in new contexts. Two examples from Heine and Kuteva (2007) should make clear what these terms mean. A good example for extension to new contexts is the use of the German verb "drohen" (literally "to threaten") as an auxiliary-like verb. If used as an auxiliary, it has a

future meaning announcing undesirable events in constructions like:

(13) Es droh-t zu regn-en.
it threaten-3.SG.PRES to rain-INF
"It is going to rain (and I dislike it)"

(Heine and Kuteva, 2007, 39)

The meaning of the verb "drohen" is abstract in the sentence above. It expresses the speaker's attitude towards what he thinks is likely to happen. Furthermore, we can observe semantic bleaching in this case because the literal meaning would demand an agent. As visible in the example above, the verb can be used in an impersonal construction with its new meaning. An English example of pragmatic extension and semantic bleaching would be the use of the verb "to keep" as an auxiliary in a sentence like: He kept going. Here, we see that the verb "to keep" occurs with a gerund following it and acquired a new meaning "continue to do something", which is induced by the usage in the new context.

The next step in grammaticalization would be decategorialization. This means that a given word will lose so much of its inherent meaning that it will give up its original category and develop into a more grammatical part of speech. A good example for this is the definite article in German "der, die, das" that has evolved from a demonstrative pronoun. In Old High German, it could still be used as a demonstrative (Braune and Reiffenstein, 2004, 247), whereas in modern German it is mostly used as a definite article. This development from a pronoun into an article is an example of how a word can come to acquire a more grammatical function over time.

The last stage in the grammaticalization model would be phonological erosion. Here, phonological material and/or stress are lost and what former was a free form develops into a bound one. Heine and Kuteva (2007, 44) present an example from Maninka (Western Mande; Sierra Leone/Guinea):

(14) à yé kàrán ná
3.SG PM learn at
"He is learning"

They claim that the marker *ná* has developed from a locative marker into a progressive marker. This process has resulted in the loss of phonological autonomy. The high tone on *ná* is lost and the word adopts the tone of the preceding word. It has also developed an allophone *lá* that is used as progressive marker if the preceding verb does not end in a nasal. Another, more extreme example of phonological erosion is the evolution of the German adverb *heute* probably deriving from Germanic *hiu dagu* (Kluge and Seebold, 2011, 414). In this case, much phonetic material has been lost and a noun phrase containing two words has developed into a single adverb.

An interesting example showing all four stages of grammaticalization is the development of the future in Romance languages. It evolved from an originally analytic construction in Latin with the infinitive plus the verb *habere*.[2] This construction was used to express an obligation, such that *cantare habeo* meant

[2]Latin also had a synthetic future, but this has been completely lost in all Romance languages.

something like *I have to sing*. Since this obligation inherently expressed an event that would occur in the future, pragmatic extension and semantic bleaching resulted in transforming this construction in an unmarked future. Morphosyntactic decategorialization transformed the auxiliary first into a clitic, as visible in Old-Portuguese *donar-lo-t'ai* meaning *I will give it to you*. The clitical pronouns *lo* and *t'* intervene between the main verb and the clitical auxiliary (Wolf and Hupka, 1981, 136). In a later stage, this clitic was transformed into a verbal affix so that no clitic pronouns can intervene between the main verb and the future marker (compare Modern Italian *te lo darò* where the two pronouns precede the inflected verb). This whole process was accompanied by phonological erosion of the future marker. In Modern Italian, the original Latin forms *habeo, habes, habet, habemus, habetis, habent* have been reduced to the affixes *-ò, -ai, -à, -emo, -ete, -anno*.

This example shows how the processes in grammaticalization can occur together, but it is not completely clarified in which chronological order they occur. It has to be admitted that their relative chronology is disputed. Bybee et al. (1994, 6) claim that phonological erosion is not the last process, but occurs simultaneously to semantic reduction. For the purpose of the present thesis, it is however not necessary to go that far in detail. It is just important to recognize that the evolution of the Romance future demonstrates very well how the four linguistic processes involved in grammaticalization can transform the nature of a construction (and even of a whole language if they occur with a high frequency) over time.

Grammaticalization does however not just mean that free words transform into affixes. It relates also to the process in which elements that already have a grammatical function evolve further towards having an even more grammatical function. This can for example be shown in the evolution of case systems. Case markers for those cases that are often called *grammatical* or *core* cases can arise from cases that are often referred to as *semantic* cases. It is for instance known that allative case markers can give rise to dative case markers. This is illustrated here with an example from Lezgian (Nakh-Daghestanian; Azerbaijan, Russia) in which the marker *-z* originally expressing the goal of a movement gave rise to a dative marker:

(15) Zun medinstitutdi-z fi-da
I.ABS medical.school-DAT go-FUT
"I will go to medical school"

(16) Ruša gadadi-z cük ga-na
girl.ERG boy-DAT flower give-AOR
"The girl gave a flower to the boy"

(Heine and Kuteva, 2010, 37)

Another important feature of grammaticalization that has not been mentioned so far is its unidirectionality. This term refers to the property of grammaticalization to change less grammatical forms into more grammatical forms, but not to allow the contrary. Admittedly, this rule allows some exceptions, a prominent one being the Modern Irish pronoun for the first person plural *muide*. Apparently, this pronoun evolved from a verbal suffix that was used to encode a subject in the first person plural and replaced the original pronoun *sinn* (Bybee et al., 1994, 14). It represents thus an example of a grammatical form that has developed into a less grammatical form. Notwithstanding, even strong critics of grammaticalization have to admit

that unidirectionality cannot be dismissed. Newmeyer (2000) tries to falsify the claim of unidirectionality by providing a bunch of counterexamples to it, but has to admit that changes from less grammatical forms into more grammatical ones occur at least ten times as often as the reverse.

Having explained the idea of grammaticalization and its most important properties, the following section will be dedicated to the question how this process may cause the results we could observe in chapter 3.

## 6.2 Why SOV-languages could tend to morphological complexity

Now that the properties of grammaticalization have been explained and it has been made clear how this phenomenon can change structures in language over time, its possible relation to the presence of case marking in SOV languages can be discussed. In order to find a link between these seemingly unrelated phenomena, one has to reflect on the evolution of case markers. Of course, case markers do not arise out of nothing, but evolve over time. Grammaticalization predicts that case affixes are the final result of a process that started with free words. In fact, Heine and Kuteva (2007, 76) claim: "Case affixes almost invariably [derive] from adpositions." This can easily be demonstrated with an example from Estonian where the adposition *kaas* used in earlier texts has developed into the Modern Estonian comitative marker *-ga*:

(17) Ta ela-b selle-s toa-s koos kahe õe-ga
she live-3SG.PRES this-INESS room-INESS together two.GEN sister-COM
"She lives in this room together with two sisters"

(Stolz et al., 2006, 364)

An essential point for the idea put forward in this chapter is the word order in SOV languages. It is known that SOV as dominant word order correlates with the existence of postpositions and the absence of prepositions (Dryer, 2000, 89). If now grammaticalization processes change those postpositions over time, one expects them to develop into suffixes since postpositions occur after the noun they modify. This is true for the Estonian example where the original form *kaas* was a postposition that has given rise to a comitative suffix. Furthermore, it can be found that 96,6% of the case markers in SOV languages are either suffixes or postpositional clitics (Dryer (2013b), Dryer (2013c)). This percentage is slightly higher than the one for the overall percentage of case markers. According to Dryer (2013c), 88,1% of all the case markers in the world's languages are either postpositional clitics or suffixes.

These numbers are in line with previous research. Hawkins and Gilligan (1988) claimed that the world's languages prefer suffixes over prefixes. Evidence from a bunch of different morphological categories is presented to prove that for all categories suffixes occur clearly more frequently than prefixes. Furthermore, they propose that the explanation for this observation is to be found within the domain of psycholinguistics. This explanation arises from a combination of previous research showing that the onset of words is more important for word recognition than the end, with other studies showing that both speakers and hearers prefer to process the stem of a word before its affixes. What they conclude is that word stems would occur at the onset of words for earlier word recognition and better sentence computation. Eventual affixes would

then follow the stems. Concerning prefixes, the problem is that they occur in the most important domain for word recognition, precede the stem and have thus to be processed before it. This delays word recognition and makes sentence structure comprehension more difficult. That is why suffixes should be preferred.

Further evidence for the preference of suffixes over prefixes comes again from studies on artificial language learning. St Clair et al. (2009) performed a study in which 24 English native speakers learned two different artificial languages. Both of these languages consisted of twelve words and two affixes. Words were divided into two different categories according to their phonological properties. Words in category A had consonant clusters, unrounded high vowels, nasals, and stops; words in category B had no consonant clusters rounded low vowels, and fricatives. Each affix could combine with exactly one of the two categories and consisted of an onset consonant and a lax vowel. The vowel was lax because word-internal vowels English tend to be lax. In this way, a phonological preference for prefixes should be induced since the final vowels of the affixes would occur word-initially with them. With suffixes, they would occur word-finally, a dispreferred phonological pattern in English. To induce a phonological preference for prefixes was done to make results more meaningful if the hypothesis of a preference for suffixes should indeed be born out.

Participants were exposed to 18 sentences in the training session. Each sentence was built up by two words and was presented to them twice. To each word, an affix from the respective category was added. Half of the participants were exposed to a language with suffixes, the other half to a language with prefixes.

After the training sessions, participants were given 24 sentences and told that half of them were similar to the language they had just learned. The test sentences were shown to them and participants were asked to tell whether they thought the sentence to be similar or dissimilar to the input language. Dissimilar sentences contained either the wrong sort of affixes or combined affixes with the wrong category of words. Every participant completed two training and two testing sessions. When participants were asked to tell whether the test sentences were similar to the sentences in the training session, their ability to recognize affixes and to group words into the different categories was tested. It was indeed found that learners of the suffixing languages answered correctly significantly more often than learners of the prefixing language. This does not only provide further evidence for Hawkins and Gilligan (1988) and their claim of an overall preference for suffixes, but also their proposed explanation that suffixes facilitate word recognition.

If one accepts that the languages of the world prefer suffixes over prefixes and that the onsets of words are more important for word recognition, one would expect phonological erosion to take place rather at the end than at the beginning of words. In the section above, it could be seen that phonological erosion is one of the parameters of grammaticalization. It is responsible for reducing the phonological form of a free word by loss of phonemes and/or stress. In the review above, it has furthermore been shown that affixes tend to occur more frequently in word-final position than in word-initial position. All this together would lead to the conclusion that grammaticalization takes place more easily where functional elements follow the words they modify. For adpositions, the most frequent source for case markers, it is true that they follow their head nouns in SOV languages. We might thus expect that adpositions evolve more often into case markers

in SOV languages than in languages with other word orders due to the respective placement of adpositions. Recalling the definition of case marking employed in this thesis, it can be stated that the claim about the preference of suffixes over prefixes can be enlarged and also include a preference for enclitics over proclitics, at least concerning case markers. This is not unexpected as these clitics function precisely as affixes would do and might not even be distinguishable from them (Haspelmath, 2011). Since unidirectionality is known to be a feature of grammaticalization, the evolution of those affixes back to free words is not expected to happen. The loss of the affixes is improbable due to the communicative reasons for case marking discussed in the previous chapters.

The diachronic process of grammaticalization might thus be another possible explanation for the frequent presence of case systems in SOV languages. The argumentation of this chapter is by no means contradictory to what has been said in the chapters before. It is true that the idea that the existence of case marking of SOV languages might be partly due to grammaticalization does not have a lot to do with communicative reasons, but it adds evidence from another field of linguistic research. What might be most interesting about this possible explanation is that it could not only account for the mere presence of case systems in SOV languages, but also for the increased probability of a language being SOV for a higher number of cases in the language. This could be due to the fact that SOV languages develop new cases more frequently and more easily than languages with other word orders.

# 7 Conclusion

It is a debated issue in linguistics whether certain properties within syntax correlate with certain other properties within morphology. This thesis has been an attempt to examine whether this claim can be empirically supported regarding certain types of word order and the presence or absence of a case system. The statistical results that were shown and discussed in chapter 3 provided indeed arguments for not considering word order and case marking as independent phenomena. It could be shown that languages with no dominant word order and languages with dominant SOV word order have case marking significantly more often than languages with other word orders. Moreover, one could observe a slight tendency for a higher probability of a language being SOV with an increased number of cases.

The issue of the following chapters was then to seek for explanations for these empirical observations. For languages with no dominant word order, success and robustness of communication were claimed to play a crucial role. Languages with fixed word order have the possibility of encoding subject and object by given word order patterns. Since this is not possible in languages with no dominant word order but identification of subject and object is crucial for a correct interpretation of the clause, it was argued that such languages need case marking to clearly distinguish the subject from the object.

Communicative reasons were also put forward as an explanation for the frequent use of case marking in languages with SOV word order. Such languages may allow more flexibility in word order than SVO languages, which might be a reason for the need of a case system in SOV languages (Bentz and Christiansen, 2013). Here, further research involving different language families would be needed to investigate whether it is in fact true that word order is more rigid if SVO is the dominant word order. Furthermore, in SOV languages problems for successful communication might originate from processes like topicalization and scrambling. It was claimed that these processes provide uncertainty especially in SOV languages and that case marking might function as a strategy that counteracts misunderstandings.

Also another factor regarding communication, namely early availability of information on the sentence structure, was examined and found to be a possible explanation for a case system in SOV languages. A theoretical framework, the MaOP by John Hawkins, has been combined with experimental studies, all of them pointing in the same direction. Grammatical patterns that present information on the syntactic structure of the sentence are preferred if they present this information early on. Case marking on nouns is thus expected in SOV languages, since the first words to be received are nouns and since case marking provides a clue for their syntactic role.

In the last chapter, communicative reasons were laid aside and the focus shifted to a diachronic explanation, discussing the phenomenon of grammaticalization. It was claimed that due to an overall preference for suffixes over prefixes in the languages of the world, SOV word order might favor the productivity of grammaticalization processes. This hypothesis would predict a higher morphological complexity of SOV languages, as free word forms might change more easily into clitics or affixes. It would then also explain why the probability for a language to be SOV increased with a higher number of cases. The reason would be that

SOV languages develop morphological cases more productively.

As far as I know, this hypothesis has not been suggested in any paper so far. It must be said that it is somehow speculative since we cannot be sure about the provenience of all case markers. Recall for example that case is also reconstructed for languages that we cannot trace further back in time, as for example Proto-Indo-European. However, the following quote from Heine and Kuteva (2007, 52) illustrates why the idea presented in the thesis can still be valid: "To be sure, there are a number of functional items for which no reliable etymology exists; but we do not see any reason why they should behave diachronically any differently from those for which there is sufficient diachronic information."

It would thus be an interesting task to try to verify or falsify the ideas that have been proposed in the last chapter. This could be done by more statistical analyses on the possible correlation of SOV languages and morphological complexity. If the idea that the presence of case systems and more productive grammaticalization in SOV languages is on the right track, then we should see more overall complexity in such languages. This means that morphological complexity could also show up in other parts of speech. It is for example known that auxiliaries follow the verbs in most SOV languages (Dryer, 2000, 90) and that they are a common source for verbal affixes as past markers (Heine and Kuteva, 2010, 137, 148) or evidentials Aikhenvald (2003). An analysis of the complexity of verbal systems in languages with different word orders would be an interesting research topic relating to what has been claimed regarding SOV word order and morphological complexity.

Whatever is the correct way of explaining the statistical correlations, this thesis has shown that case marking is found significantly more often in languages with either SOV word order or with no dominant word order. This shows that it is indeed worth looking for interactions of syntax and morphology.

# 8 Bibliography


Aikhenvald, A. Y. (2003). Evidentiality in typological perspective. In A. Y. Aikhenvald (Ed.), *Studies in Evidentiality*, Amsterdam/Philadelphia, pp. 1–32. John Benjamins.

Bartsch, V. and T. Vennemann (1983). *Grundzüge der Sprachtheorie. Eine linguistische Einführung*. Tübingen: Niemeyer.

Bentz, C. and M. H. Christiansen (2013). Linguistic adaptation: The trade-off between case marking and fixed word orders in Germanic and Romance languages. In *East Flows the Great River: Festschrift in Honor of Prof. William SY. Wang's 80th Birthday*, pp. 45–61.

Blake, B. J. (1994). *Case*. Cambridge: Cambrdige Univesity Press.

Braune, W. and I. Reiffenstein (2004). *Althochdeutsche Grammatik I.* Tübingen: Max Niemeyer Verlag.

Bybee, J. L., R. D. Perkins, and W. Pagliuca (1994). *The evolution of grammar: Tense, aspect, and modality in the languages of the world*. Chicago [i.a.]: University of Chicago Press.

Coogan, M. D. (Ed.) (2010). *The new Oxford annotated Bible: new revised standard version with the Apocrypha; an ecumenical study Bible*. Oxford [u.a.]: Oxford Univ. Press.

Culbertson, J., P. Smolensky, and G. Legendre (2012). Learning biases predict a word order universal. *Cognition 122*(3), 306–329.

Dryer, M. S. (1988). Object-verb order and adjective-noun order: dispelling a myth. *Lingua 74*(2-3), 185–217.

Dryer, M. S. (1992). The Greenbergian word order correlations. *Language 68*(1), 81–138.

Dryer, M. S. (1997). On the six-way word order typology. *Studies in Language. International Journal sponsored by the Foundation "Foundations of Language" 21*(1), 69–103.

Dryer, M. S. (2000). Word order. In T. Shopen (Ed.), *Language Typology and Syntactic Description. Volume I: Clause Structure*, Cambridge, pp. 61–131. Cambridge University Press.

Dryer, M. S. (2013a). Determining dominant word order. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology.

Dryer, M. S. (2013b). Order of subject, object and verb. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/51. Accessed on 2018-05-02.).

Dryer, M. S. (2013c). Position of case affixes. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/51. Accessed on 2018-05-02.).

Fedzechkina, M., F. T. Jaeger, and J. C. Trueswell (2015). Production is biased to provide informative cues early: Evidence from miniature artificial languages. *Proceedings of the Annual Meeting of the Cognitive Science Society 37*, 674–679.

Fedzechkina, M., E. L. Newport, and F. T. Jaeger (2017). Balancing effort and information transmission during language acquisition: Evidence from word order and case marking. *Cognitive science 41*(2), 416–446.

Futrell, R., T. Hickey, A. Lee, E. Lim, E. Luchkina, and E. Gibson (2015). Cross-linguistic gestures reflect typological universals: A subject-initial, verb-final bias in speakers of diverse languages. *Cognition 136*, 215–221.

Gibson, E., S. T. Piantadosi, K. Brink, L. Bergen, E. Lim, and R. Saxe (2013). A noisy-channel account of crosslinguistic word-order variation. *Psychological science 24*(7), 1079–1088.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. *Universals of language 2*, 73–113.

Haspelmath, M. (2011). The indeterminacy of word segmentation and the nature of morphology and syntax. *Folia linguistica 45*(1), 31–80.

Hawkins, J. A. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.

Hawkins, J. A. (2007). Processing typology and why psychologists need to know about it. *New Ideas in Psychology 25*(2), 87–107.

Hawkins, J. A. and G. Gilligan (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua 74*(2-3), 219–259.

Heine, B. and T. Kuteva (2007). *The Genesis of Grammar: A Reconstruction*. Oxford: Oxford University Press.

Heine, B. and T. Kuteva (2010). *World lexicon of grammaticalization*. Cambridge [i.a.]: Cambridge University Press.

Iggesen, O. A. (2013). Number of cases. In M. S. Dryer and M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*, Leipzig. Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/51. Accessed on 2018-05-02.).

Kluge, F. and E. Seebold (2011). *Etymologisches Wörterbuch der deutschen Sprache*. Berlin/Boston: De Gruyter.

Kurumada, C. and T. F. Jaeger (2015). Communicative efficiency in language production: Optional case-marking in japanese. *Journal of Memory and Language 83*, 152–178.

Lupyan, G. and M. H. Christiansen (2002). Case, word order, and language learnability: Insights from connectionist modeling. *Proceedings of the Annual Meeting of the Cognitive Science Society 24*, 596–601.

Mackridge, P. (1985). *The modern Greek language*. Oxford [i.a.]: Oxford University Press.

Müller, S. (2016). *Grammatical theory: From transformational grammar to constraint-based approaches*. Berlin: Language Science Press.

Newmeyer, F. J. (2000). Deconstructing grammaticalization. *Language sciences 23*(2-3), 187–229.

Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago [i.a.]: University of Chicago Press.

Sinnemäki, K. (2008). Complexity trade-offs in core argument marking. In F. K. Matti Miestamo, Kaius Sinnemäki (Ed.), *Language complexity. Typology, contact, change*, Amsterdam [i.a.], pp. 67–88. John Benjamins.

St Clair, M. C., P. Monaghan, and M. Ramscar (2009). Relationships between language structure and language learning: The suffixing preference and grammatical categorization. *Cognitive Science 33*(7), 1317–1329.

Stolz, T., C. Stroh, and A. Urdze (2006). *On comitatives and related categories: A typological study with special focus on the languages of Europe*. Berlin: Walter de Gruyter.

Tsujimura, N. (1996). *An introduction to Japanese linguistics*. Cambridge, Mass.: Blackwell.

Van Everbroeck, E. (2003). Language type frequency and learnability from a connectionist perspective. *Linguistic Typology 7*(1), 1–50.

Velupillai, V. (2012). *An introduction to linguistic typology*. Amsterdam [i.a.]: Benjamins.

Wolf, L. and W. Hupka (1981). *Altfranzösisch. Entstehung und Charakteristik*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Yamashita, H. (1997). The effects of word-order and case marking information on the processing of Japanese. *Journal of Psycholinguistic Research 26*(2), 163–188.

# 9 Appendix

## 9.1 Sample A

| Language | Word order | Genus | Family | Case marking |
|---|---|---|---|---|
| Abipón | SVO | South Guaicuruan | Guaicuruan | No |
| Abui | SOV | Greater Alor | Timor-Alor-Pantar | No |
| Abun | SVO | North-Central Bird's Head | West Papuan | No |
| Acehnese | No dominant order | Malayo-Sumbawan | Austronesian | No |
| Achagua | SVO | Inland Northern Arawakan | Arawakan | No |
| Acholi | SVO | Nilotic | Eastern Sudanic | No |
| Achumawi | SVO | Palaihnihan | Hokan | Yes |
| Acoma | No dominant order | Keresan | Keresan | Yes |
| Adang | SOV | Greater Alor | Timor-Alor-Pantar | No |
| Adioukrou | SVO | Kwa | Niger-Congo | No |
| Adyghe (Abzakh) | SOV | Northwest Caucasian | Northwest Caucasian | Yes |
| Adynyamathanha | No dominant order | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Agarabi | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Agta (Central) | VSO | Greater Central Philippine | Austronesian | No |
| Ajagbe | No dominant order | Kwa | Niger-Congo | No |
| Akan | SVO | Kwa | Niger-Congo | No |
| Akha | SOV | Burmese-Lolo | Sino-Tibetan | No |
| Ala'ala | SOV | Oceanic | Austronesian | Yes |
| Alamblak | SOV | Sepik Hill | Sepik | Yes |
| Alawa | No dominant order | Alawa | Mangarrayi-Maran | Yes |
| Albanian | SVO | Albanian | Indo-European | Yes |
| Aleut | SOV | Aleut | Eskimo-Aleut | Yes |
| Alune | SVO | Central Malayo-Polynesian | Austronesian | No |
| Alyawarra | SOV | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Amahuaca | No dominant order | Panoan | Panoan | No |
| Amanab | SOV | Border | Border | Yes |
| Ambae (Lolovoli Northeast) | SVO | Oceanic | Austronesian | No |
| Ambulas | SOV | Middle Sepik | Sepik | Yes |
| Amele | SOV | Madang | Trans-New Guinea | No |
| Amharic | SOV | Semitic | Afro-Asiatic | Yes |
| Anejom | VOS | Oceanic | Austronesian | No |
| Angas | SVO | West Chadic | Afro-Asiatic | No |
| Anggor | SOV | Senagi | Senagi | No |

| | | | | |
|---|---|---|---|---|
| Anguthimri | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Anindilyakwa | No dominant order | Anindilyakwa | Gunwinyguan | Yes |
| Anufo | SVO | Kwa | Niger-Congo | No |
| Anywa | No dominant order | Nilotic | Eastern Sudanic | No |
| Ao | SOV | Kuki-Chin | Sino-Tibetan | Yes |
| Apache (Western) | SOV | Athapaskan | Na-Dene | Yes |
| Apalaí | No dominant order | Cariban | Cariban | No |
| Apatani | SOV | Tani | Sino-Tibetan | No |
| Apinayé | SOV | Ge-Kaingang | Macro-Ge | No |
| Arabana | SOV | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Arabic (Egyptian) | SVO | Semitic | Afro-Asiatic | No |
| Arabic (Gulf) | SVO | Semitic | Afro-Asiatic | No |
| Arabic (Iraqi) | SVO | Semitic | Afro-Asiatic | Yes |
| Arabic (Modern Standard) | VSO | Semitic | Afro-Asiatic | Yes |
| Arabic (Syrian) | No dominant order | Semitic | Afro-Asiatic | Yes |
| Arapesh (Mountain) | SVO | Kombio-Arapesh | Torricelli | No |
| Arawak | SVO | Caribbean Arawakan | Arawakan | No |
| Archi | SOV | Lezgic | Nakh-Daghestanian | Yes |
| Armenian (Eastern) | No dominant order | Armenian | Indo-European | Yes |
| Armenian (Western) | SOV | Armenian | Indo-European | Yes |
| Arrernte (Mparntwe) | SOV | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Asmat | SOV | Asmat-Kamoro | Trans-New Guinea | No |
| Assamese | SOV | Indic | Indo-European | Yes |
| Athpare | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Atsugewi | No dominant order | Palaihnihan | Hokan | Yes |
| Au | SVO | Wapei-Palei | Torricelli | No |
| Avar | SOV | Avar-Andic-Tsezic | Nakh-Daghestanian | Yes |
| Awa | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Awa Pit | SOV | Barbacoan | Barbacoan | Yes |
| Awtuw | SOV | Ram | Sepik | Yes |
| Aymara (Central) | SOV | Aymaran | Aymaran | Yes |
| Azari (Iranian) | SOV | Turkic | Altaic | Yes |
| Azerbaijani | SOV | Turkic | Altaic | Yes |
| Babungo | SVO | Bantoid | Niger-Congo | No |
| Bachamal | No dominant order | Anson Bay | Anson Bay | Yes |
| Bagirmi | SVO | Bongo-Bagirmi | Central Sudanic | No |

| | | | | |
|---|---|---|---|---|
| Bai | SVO | Bai | Sino-Tibetan | No |
| Bajau (West Coast) | No dominant order | Sama-Bajaw | Austronesian | No |
| Baka (in South Sudan) | SVO | Bongo-Bagirmi | Central Sudanic | No |
| Bambara | SOV | Western Mande | Mande | No |
| Banoni | SVO | Oceanic | Austronesian | No |
| Bao'an | SOV | Mongolic | Altaic | Yes |
| Barai | SOV | Koiarian | Trans-New Guinea | Yes |
| Barasano | No dominant order | Tucanoan | Tucanoan | Yes |
| Baré | SVO | Inland Northern Arawakan | Arawakan | Yes |
| Bari | SVO | Nilotic | Eastern Sudanic | No |
| Barupu | SOV | Warapu | Skou | No |
| Basque | SOV | Basque | Basque | Yes |
| Batak (Karo) | No dominant order | Northwest Sumatra-Barrier Islands | Austronesian | No |
| Batak (Toba) | VOS | Northwest Sumatra-Barrier Islands | Austronesian | No |
| Baure | VOS | Bolivia-Parana | Arawakan | Yes |
| Bauzi | SOV | East Geelvink Bay | East Geelvink Bay | Yes |
| Beja | SOV | Beja | Afro-Asiatic | Yes |
| Bella Coola | VSO | Bella Coola | Salishan | No |
| Belorussian | No dominant order | Slavic | Indo-European | Yes |
| Bengali | SOV | Indic | Indo-European | Yes |
| Berber (Figuig) | VSO | Berber | Afro-Asiatic | Yes |
| Berber (Middle Atlas) | VSO | Berber | Afro-Asiatic | Yes |
| Berber (Rif) | SVO | Berber | Afro-Asiatic | Yes |
| Beria | SOV | Eastern Saharan | Saharan | Yes |
| Betoi | SOV | Betoi | Betoi | Yes |
| Betta Kurumba | SOV | Southern Dravidian | Dravidian | Yes |
| Bhojpuri | SOV | Indic | Indo-European | Yes |
| Biak | SVO | South Halmahera - West New Guinea | Austronesian | No |
| Biloxi | SOV | Core Siouan | Siouan | Yes |
| Bilua | No dominant order | Bilua | Solomons East Papuan | No |
| Bimoba | SVO | Gur | Niger-Congo | No |
| Binandere | SOV | Binanderean | Trans-New Guinea | Yes |
| Bininj Gun-Wok | No dominant order | Gunwinygic | Gunwinyguan | Yes |
| Biri | SVO | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Birom | SVO | Platoid | Niger-Congo | No |
| Bisu | SOV | Burmese-Lolo | Sino-Tibetan | Yes |

| | | | | |
|---|---|---|---|---|
| Bongo | SVO | Bongo-Bagirmi | Central Sudanic | No |
| Breton | SVO | Celtic | Indo-European | Yes |
| Budu | SVO | Bantoid | Niger-Congo | No |
| Buduma | SVO | Biu-Mandara | Afro-Asiatic | No |
| Bulgarian | SVO | Slavic | Indo-European | No |
| Bunu (Younuo) | SVO | Hmong-Mien | Hmong-Mien | No |
| Bunuba | No dominant order | Bunuban | Bunuban | Yes |
| Burarra | SOV | Burarran | Mangrida | Yes |
| Buriat | SOV | Mongolic | Altaic | Yes |
| Burmese | SOV | Burmese-Lolo | Sino-Tibetan | Yes |
| Burushaski | SOV | Burushaski | Burushaski | Yes |
| Cahuilla | SOV | California Uto-Aztecan | Uto-Aztecan | Yes |
| Camling | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Campa (Axininca) | No dominant order | Pre-Andine Arawakan | Arawakan | Yes |
| Camsá | No dominant order | Camsá | Camsá | Yes |
| Canela-Krahô | SOV | Ge-Kaingang | Macro-Ge | No |
| Cantonese | SVO | Chinese | Sino-Tibetan | No |
| Carib | SOV | Cariban | Cariban | No |
| Carib (De'kwana) | SOV | Cariban | Cariban | No |
| Cashibo | SOV | Panoan | Panoan | Yes |
| Catalan | SVO | Romance | Indo-European | No |
| Cavineña | No dominant order | Tacanan | Tacanan | Yes |
| Cayapa | SOV | Barbacoan | Barbacoan | Yes |
| Cèmuhî | VOS | Oceanic | Austronesian | No |
| Chácobo | No dominant order | Panoan | Panoan | No |
| Chai | No dominant order | Surmic | Eastern Sudanic | Yes |
| Chamorro | VSO | Chamorro | Austronesian | No |
| Chang | SOV | Northern Naga | Sino-Tibetan | Yes |
| Chatino (Yaitepec) | VSO | Zapotecan | Oto-Manguean | No |
| Chechen | SOV | Nakh | Nakh-Daghestanian | Yes |
| Chemehuevi | No dominant order | Numic | Uto-Aztecan | Yes |
| Chinantec (Comaltepec) | VSO | Chinantecan | Oto-Manguean | No |
| Chinantec (Lealao) | VOS | Chinantecan | Oto-Manguean | No |
| Chinantec (Palantla) | VSO | Chinantecan | Oto-Manguean | Yes |
| Chinantec (Quiotepec) | VSO | Chinantecan | Oto-Manguean | No |
| Chipaya | SOV | Uru-Chipaya | Uru-Chipaya | Yes |
| Chitimacha | SOV | Chitimacha | Chitimacha | Yes |
| Choctaw | SOV | Muskogean | Muskogean | Yes |

| | | | | |
|---|---|---|---|---|
| Cholón | SOV | Cholon | Cholon | Yes |
| Chontal Maya | SVO | Mayan | Mayan | No |
| Chrau | SVO | Bahnaric | Austro-Asiatic | No |
| Chukchi | No dominant order | Northern Chukotko-Kamchatkan | Chukotko-Kamchatkan | Yes |
| Chumash (Barbareño) | VOS | Chumash | Chumash | No |
| Chuvash | SOV | Turkic | Altaic | Yes |
| Coahuilteco | SOV | Coahuiltecan | Coahuiltecan | Yes |
| Cocama | SVO | Tupi-Guaraní | Tupian | Yes |
| Comanche | SOV | Numic | Uto-Aztecan | Yes |
| Coos (Hanis) | No dominant order | Coosan | Oregon Coast | Yes |
| Coptic | SVO | Egyptian-Coptic | Afro-Asiatic | Yes |
| Cornish | SVO | Celtic | Indo-European | No |
| Cree (Swampy) | No dominant order | Algonquian | Algic | Yes |
| Crow | SOV | Core Siouan | Siouan | No |
| Cua | SVO | Bahnaric | Austro-Asiatic | Yes |
| Cubeo | OVS | Tucanoan | Tucanoan | Yes |
| Cupeño | SOV | California Uto-Aztecan | Uto-Aztecan | Yes |
| Czech | SVO | Slavic | Indo-European | Yes |
| Dagur | SOV | Mongolic | Altaic | Yes |
| Dan | SOV | Eastern Mande | Mande | No |
| Dani (Lower Grand Valley) | SOV | Dani | Trans-New Guinea | Yes |
| Danish | SVO | Germanic | Indo-European | Yes |
| Dargwa | SOV | Lak-Dargwa | Nakh-Daghestanian | Yes |
| Darma | SOV | Bodic | Sino-Tibetan | No |
| Dâw | SOV | Nadahup | Nadahup | Yes |
| Day | SVO | Adamawa | Niger-Congo | No |
| Desano | SOV | Tucanoan | Tucanoan | Yes |
| Dhaasanac | SOV | Lowland East Cushitic | Afro-Asiatic | Yes |
| Dhargari | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Dharumbal | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Didinga | VSO | Surmic | Eastern Sudanic | Yes |
| Diegueño (Mesa Grande) | SOV | Yuman | Hokan | Yes |
| Digaro | SOV | Digaroan | Sino-Tibetan | Yes |
| Dime | SOV | South Omotic | Afro-Asiatic | Yes |
| Diola-Fogny | SVO | Northern Atlantic | Niger-Congo | No |
| Diyari | SOV | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Djabugay | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Djambarrpuyngu | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |

| | | | | |
|---|---|---|---|---|
| Djapu | SOV | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Djaru | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Djinang | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Djingili | SOV | Djingili | Mirndi | Yes |
| Domari | No dominant order | Indic | Indo-European | Yes |
| Dong (Southern) | SVO | Kam-Tai | Tai-Kadai | No |
| Donno So | SOV | Dogon | Dogon | Yes |
| Doyayo | SVO | Adamawa | Niger-Congo | No |
| Drehu | No dominant order | Oceanic | Austronesian | No |
| Duka | SVO | Kainji | Niger-Congo | No |
| Dumo | SOV | Western Skou | Skou | No |
| Duna | SOV | Duna | Trans-New Guinea | Yes |
| Dutch | No dominant order | Germanic | Indo-European | No |
| Dyirbal | No dominant order | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Efate (South) | SVO | Oceanic | Austronesian | Yes |
| Eipo | SOV | Mek | Trans-New Guinea | Yes |
| Émérillon | SOV | Tupi-Guaraní | Tupian | Yes |
| Emmi | No dominant order | Wagaydy | Western Daly | Yes |
| Enets | SOV | Samoyedic | Uralic | Yes |
| Engenni | SVO | Edoid | Niger-Congo | No |
| Enggano | SVO | Enggano | Austronesian | Yes |
| Epena Pedee | SOV | Choco | Choco | Yes |
| Ese Ejja | SOV | Tacanan | Tacanan | No |
| Estonian | SVO | Finnic | Uralic | Yes |
| Eudeve | SOV | Cahita | Uto-Aztecan | Yes |
| Even | SOV | Tungusic | Altaic | Yes |
| Evenki | SOV | Tungusic | Altaic | Yes |
| Ewe | SVO | Kwa | Niger-Congo | No |
| Ewondo | SVO | Bantoid | Niger-Congo | No |
| Fijian | No dominant order | Oceanic | Austronesian | No |
| Finnish | SVO | Finnic | Uralic | Yes |
| Folopa | SOV | Teberan | Teberan-Pawaian | Yes |
| Fore | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Frisian | No dominant order | Germanic | Indo-European | No |
| Fur | SOV | Fur | Fur | Yes |
| Futuna-Aniwa | SVO | Oceanic | Austronesian | No |
| Gadsup | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Gaelic (Scots) | VSO | Celtic | Indo-European | Yes |
| Gahuku | SOV | Eastern Highlands | Trans-New Guinea | Yes |

| | | | | |
|---|---|---|---|---|
| Galo | SOV | Tani | Sino-Tibetan | Yes |
| Gamo | SOV | North Omotic | Afro-Asiatic | Yes |
| Gapapaiwa | SOV | Oceanic | Austronesian | Yes |
| Garo | SOV | Bodo-Garo | Sino-Tibetan | Yes |
| Garrwa | No dominant order | Garrwan | Garrwan | Yes |
| Gbaya Kara | SVO | Gbaya-Manza-Ngbaka | Niger-Congo | No |
| Gbeya Bossangoa | SVO | Gbaya-Manza-Ngbaka | Niger-Congo | No |
| Georgian | SOV | Kartvelian | Kartvelian | Yes |
| German | No dominant order | Germanic | Indo-European | Yes |
| Gidabal | No dominant order | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Gimira | SOV | North Omotic | Afro-Asiatic | Yes |
| Gitksan | No dominant order | Tsimshianic | Penutian | No |
| Goajiro | VSO | Caribbean Arawakan | Arawakan | Yes |
| Goemai | SVO | West Chadic | Afro-Asiatic | No |
| Gogodala | SOV | Gogodala | Gogodala-Suki | Yes |
| Gokana | SVO | Cross River | Niger-Congo | No |
| Golin | SOV | Chimbu | Trans-New Guinea | No |
| Gondi | SOV | South-Central Dravidian | Dravidian | Yes |
| Gooniyandi | No dominant order | Bunuban | Bunuban | Yes |
| Great Andamanese | SOV | Great Andamanese | Great Andamanese | Yes |
| Grebo | SVO | Kru | Niger-Congo | No |
| Greek (Modern) | No dominant order | Greek | Indo-European | Yes |
| Greenlandic (West) | SOV | Eskimo | Eskimo-Aleut | Yes |
| Guaraní | SVO | Tupi-Guaraní | Tupian | No |
| Gude | VSO | Biu-Mandara | Afro-Asiatic | No |
| Gugada | SOV | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Guhu-Samane | SOV | Binanderean | Trans-New Guinea | Yes |
| Gula (in Central African Republic) | SVO | Bongo-Bagirmi | Central Sudanic | No |
| Gumawana | SOV | Oceanic | Austronesian | No |
| Gumbaynggir | No dominant order | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Gumuz | SVO | Gumuz | Gumuz | No |
| Gunin | No dominant order | Worrorran | Worrorran | Yes |
| Gününa Küne | No dominant order | Puelche | Chon | Yes |
| Gureng Gureng | No dominant order | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Gurr-goni | No dominant order | Burarran | Mangrida | Yes |
| Guugu Yimidhirr | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |

| | | | | |
|---|---|---|---|---|
| Halia | SVO | Oceanic | Austronesian | No |
| Hamtai | SOV | Angan | Trans-New Guinea | Yes |
| Hatam | SVO | Hatam | West Papuan | No |
| Hausa | SVO | West Chadic | Afro-Asiatic | No |
| Hawaiian | VSO | Oceanic | Austronesian | No |
| Hayu | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Hindi | SOV | Indic | Indo-European | Yes |
| Hixkaryana | OVS | Cariban | Cariban | No |
| Hmong Njua | SVO | Hmong-Mien | Hmong-Mien | No |
| Hoava | VSO | Oceanic | Austronesian | No |
| Hopi | SOV | Hopi | Uto-Aztecan | Yes |
| Hua | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Hualapai | SOV | Yuman | Hokan | Yes |
| Huastec | SVO | Mayan | Mayan | No |
| Huave (San Mateo del Mar) | SVO | Huavean | Huavean | No |
| Huichol | SOV | Corachol | Uto-Aztecan | Yes |
| Huitoto (Minica) | SOV | Huitoto | Huitotoan | Yes |
| Huitoto (Murui) | SOV | Huitoto | Huitotoan | Yes |
| Hunde | SVO | Bantoid | Niger-Congo | Yes |
| Hungarian | No dominant order | Ugric | Uralic | Yes |
| Hunzib | SOV | Avar-Andic-Tsezic | Nakh-Daghestanian | Yes |
| Hup | SOV | Nadahup | Nadahup | Yes |
| Hupa | No dominant order | Athapaskan | Na-Dene | Yes |
| Iaai | SVO | Oceanic | Austronesian | No |
| Iban | SVO | Malayo-Sumbawan | Austronesian | No |
| Icelandic | SVO | Germanic | Indo-European | Yes |
| Igbo | SVO | Igboid | Niger-Congo | No |
| Ijo (Kolokuma) | SOV | Ijoid | Ijoid | Yes |
| Ik | VSO | Kuliak | Eastern Sudanic | Yes |
| Ika | SOV | Arhuacic | Chibchan | Yes |
| Ilocano | VSO | Northern Luzon | Austronesian | No |
| Inanwatan | No dominant order | South Bird's Head | Marind | Yes |
| Indonesian | SVO | Malayo-Sumbawan | Austronesian | No |
| Innamincka | SOV | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Iquito | SVO | Zaparoan | Zaparoan | Yes |
| Iranxe | SOV | Iranxe | Iranxe | No |
| Irarutu | SVO | South Halmahera - West New Guinea | Austronesian | No |
| Irish | VSO | Celtic | Indo-European | Yes |
| Italian | SVO | Romance | Indo-European | No |
| Itelmen | SOV | Southern Chukotko-Kamchatkan | Chukotko-Kamchatkan | Yes |
| Izi | SVO | Igboid | Niger-Congo | No |

| | | | | |
|---|---|---|---|---|
| Jabêm | SVO | Oceanic | Austronesian | Yes |
| Jakaltek | VSO | Mayan | Mayan | No |
| Jamsay | SOV | Dogon | Dogon | Yes |
| Japanese | SOV | Japanese | Japanese | Yes |
| Jebero | No dominant order | Cahuapanan | Cahuapanan | Yes |
| Jingpho | SOV | Jingpho | Sino-Tibetan | No |
| Jino | SOV | Burmese-Lolo | Sino-Tibetan | No |
| Jivaro | SOV | Jivaroan | Jivaroan | Yes |
| Ju\|'hoan | SVO | Ju-Kung | Kxa | No |
| Jukun | SVO | Platoid | Niger-Congo | No |
| Jur Mödö | SVO | Bongo-Bagirmi | Central Sudanic | No |
| Kabardian | SOV | Northwest Caucasian | Northwest Caucasian | Yes |
| Kachari | SOV | Bodo-Garo | Sino-Tibetan | Yes |
| Kadazan | VSO | North Borneo | Austronesian | No |
| Kairiru | SOV | Oceanic | Austronesian | No |
| Kalispel | VSO | Interior Salish | Salishan | Yes |
| Kalkatungu | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Kaluli | SOV | Bosavi | Bosavi | Yes |
| Kamaiurá | SOV | Tupi-Guaraní | Tupian | Yes |
| Kamasau | SOV | Marienberg | Torricelli | No |
| Kamass | SOV | Samoyedic | Uralic | Yes |
| Kamoro | SOV | Asmat-Kamoro | Trans-New Guinea | Yes |
| Kamu | No dominant order | Eastern Daly | Eastern Daly | Yes |
| Kana | SVO | Cross River | Niger-Congo | Yes |
| Kanakuru | SVO | West Chadic | Afro-Asiatic | No |
| Kanembu | SOV | Western Saharan | Saharan | Yes |
| Kannada | SOV | Southern Dravidian | Dravidian | Yes |
| Kanoê | SOV | Kapixana | Kapixana | Yes |
| Kanuri | SOV | Western Saharan | Saharan | Yes |
| Kara (in Central African Republic) | SVO | Bongo-Bagirmi | Central Sudanic | No |
| Karachay-Balkar | SOV | Turkic | Altaic | Yes |
| Karakalpak | SOV | Turkic | Altaic | Yes |
| Karankawa | No dominant order | Karankawa | Karankawa | No |
| Karen (Bwe) | SVO | Karen | Sino-Tibetan | No |
| Karen (Sgaw) | SVO | Karen | Sino-Tibetan | No |
| Karimojong | VSO | Nilotic | Eastern Sudanic | No |
| Karó (Arára) | SOV | Ramarama | Tupian | No |
| Karok | No dominant order | Karok | Karok | Yes |
| Kashmiri | SVO | Indic | Indo-European | Yes |
| Katcha | SVO | Kadugli | Kadu | No |
| Katla | SVO | Katla-Tima | Kordofanian | No |

| | | | | |
|---|---|---|---|---|
| Kawaiisu | No dominant order | Numic | Uto-Aztecan | Yes |
| Kayah Li (Eastern) | SVO | Karen | Sino-Tibetan | No |
| Kayapó | SOV | Ge-Kaingang | Macro-Ge | Yes |
| Kayardild | No dominant order | Tangkic | Tangkic | Yes |
| Kele | SVO | Oceanic | Austronesian | No |
| Kemant | SOV | Central Cushitic | Afro-Asiatic | Yes |
| Kenga | SVO | Bongo-Bagirmi | Central Sudanic | Yes |
| Keresan (Santa Ana) | No dominant order | Keresan | Keresan | Yes |
| Ket | SOV | Yeniseian | Yeniseian | Yes |
| Kewa | SOV | Engan | Trans-New Guinea | Yes |
| Khalaj | SOV | Turkic | Altaic | Yes |
| Khaling | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Khalkha | SOV | Mongolic | Altaic | Yes |
| Kham | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Kham (Dege) | SOV | Bodic | Sino-Tibetan | Yes |
| Khanty | SOV | Ugric | Uralic | Yes |
| Kharia | No dominant order | Munda | Austro-Asiatic | Yes |
| Khasi | SVO | Khasian | Austro-Asiatic | No |
| Khmer | SVO | Khmer | Austro-Asiatic | No |
| Khmu' | SVO | Palaung-Khmuic | Austro-Asiatic | No |
| Kilivila | No dominant order | Oceanic | Austronesian | No |
| Kiliwa | SOV | Yuman | Hokan | Yes |
| Kinnauri | SOV | Bodic | Sino-Tibetan | Yes |
| Kiowa | SOV | Kiowa-Tanoan | Kiowa-Tanoan | Yes |
| Kipea | VOS | Kariri | Kariri | No |
| Kiribati | VOS | Oceanic | Austronesian | No |
| Kisi | No dominant order | Mel | Niger-Congo | No |
| Kiwai (Southern) | SOV | Kiwaian | Kiwaian | Yes |
| Klamath | No dominant order | Klamath-Modoc | Penutian | Yes |
| Koasati | SOV | Muskogean | Muskogean | Yes |
| Kobon | SOV | Madang | Trans-New Guinea | No |
| Koegu | SVO | Surmic | Eastern Sudanic | No |
| Koiali (Mountain) | SOV | Koiarian | Trans-New Guinea | Yes |
| Koita | SOV | Koiarian | Trans-New Guinea | Yes |
| Kokborok | SOV | Bodo-Garo | Sino-Tibetan | Yes |
| Kokota | VSO | Oceanic | Austronesian | No |
| Kolami | SOV | Central Dravidian | Dravidian | Yes |
| Kombai | SOV | Awju-Dumut | Trans-New Guinea | No |
| Kombio | SVO | Kombio-Arapesh | Torricelli | Yes |
| Komi-Zyrian | SVO | Permic | Uralic | Yes |

| | | | | |
|---|---|---|---|---|
| Komo | SVO | Koman | Koman | No |
| Kongo | SVO | Bantoid | Niger-Congo | No |
| Konni | SVO | Gur | Niger-Congo | No |
| Korana | No dominant order | Khoe-Kwadi | Khoe-Kwadi | Yes |
| Korean | SOV | Korean | Korean | Yes |
| Koreguaje | VSO | Tucanoan | Tucanoan | Yes |
| Korku | SOV | Munda | Austro-Asiatic | Yes |
| Koromfe | SVO | Gur | Niger-Congo | No |
| Korowai | SOV | Awju-Dumut | Trans-New Guinea | Yes |
| Koryak | SOV | Northern Chukotko-Kamchatkan | Chukotko-Kamchatkan | Yes |
| Kosraean | SVO | Oceanic | Austronesian | No |
| Koya | SOV | South-Central Dravidian | Dravidian | Yes |
| Koyra Chiini | SVO | Songhay | Songhay | No |
| Koyraboro Senni | SOV | Songhay | Songhay | No |
| Kresh | SVO | Kresh | Central Sudanic | No |
| Krongo | VSO | Kadugli | Kadu | Yes |
| Kuku-Yalanji | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Kuman | SOV | Chimbu | Trans-New Guinea | No |
| Kumauni | SOV | Indic | Indo-European | Yes |
| Kunama | SOV | Kunama | Kunama | Yes |
| Kunimaipa | SOV | Goilalan | Trans-New Guinea | Yes |
| Kuot | VSO | Kuot | Kuot | No |
| Kurdish (Central) | SOV | Iranian | Indo-European | No |
| Kutenai | No dominant order | Kutenai | Kutenai | No |
| Kuuk Thaayorre | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Kuvi | SOV | South-Central Dravidian | Dravidian | Yes |
| Kwaza | No dominant order | Kwaza | Kwaza | Yes |
| Kwoma | SOV | Middle Sepik | Sepik | Yes |
| Kwomtari | SOV | Kwomtari | Kwomtari-Baibai | Yes |
| Kyaka | SOV | Engan | Trans-New Guinea | Yes |
| Ladakhi | SOV | Bodic | Sino-Tibetan | Yes |
| Lahu | SOV | Burmese-Lolo | Sino-Tibetan | No |
| Lai | SOV | Kuki-Chin | Sino-Tibetan | No |
| Lakhota | SOV | Core Siouan | Siouan | No |
| Lalo | SOV | Burmese-Lolo | Sino-Tibetan | No |
| Lamaholot | SVO | Central Malayo-Polynesian | Austronesian | No |
| Lamang | VSO | Biu-Mandara | Afro-Asiatic | No |
| Lamani | SOV | Indic | Indo-European | Yes |
| Lampung | SVO | Lampungic | Austronesian | No |
| Lango | SVO | Nilotic | Eastern Sudanic | No |

| | | | | |
|---|---|---|---|---|
| Lao | SVO | Kam-Tai | Tai-Kadai | No |
| Latvian | SVO | Baltic | Indo-European | Yes |
| Leko | No dominant order | Leko | Leko | Yes |
| Lele | SVO | East Chadic | Afro-Asiatic | No |
| Lelemi | SVO | Kwa | Niger-Congo | No |
| Lenakel | SVO | Oceanic | Austronesian | No |
| Lendu | No dominant order | Lendu | Central Sudanic | No |
| Lewo | SVO | Oceanic | Austronesian | No |
| Lezgian | SOV | Lezgic | Nakh-Daghestanian | Yes |
| Limilngan | No dominant order | Limilngan | Darwin Region | Yes |
| Linda | SVO | Ubangi | Niger-Congo | No |
| Lisu | SOV | Burmese-Lolo | Sino-Tibetan | No |
| Lithuanian | SVO | Baltic | Indo-European | Yes |
| Logoti | No dominant order | Moru-Ma'di | Central Sudanic | No |
| Longgu | VOS | Oceanic | Austronesian | No |
| Loniu | SVO | Oceanic | Austronesian | No |
| Lotha | SOV | Kuki-Chin | Sino-Tibetan | Yes |
| Lucazi | SVO | Bantoid | Niger-Congo | Yes |
| Lugbara | No dominant order | Moru-Ma'di | Central Sudanic | No |
| Luiseño | No dominant order | California Uto-Aztecan | Uto-Aztecan | Yes |
| Lunda | SVO | Bantoid | Niger-Congo | Yes |
| Luo | SVO | Nilotic | Eastern Sudanic | No |
| Lusi | SVO | Oceanic | Austronesian | No |
| Luvale | SVO | Bantoid | Niger-Congo | No |
| Maasai | VSO | Nilotic | Eastern Sudanic | Yes |
| Maba | SOV | Maban | Maban | Yes |
| Macushi | No dominant order | Cariban | Cariban | Yes |
| Ma'di | No dominant order | Moru-Ma'di | Central Sudanic | No |
| Madimadi | No dominant order | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Mae | SVO | Oceanic | Austronesian | No |
| Magahi | SOV | Indic | Indo-European | Yes |
| Magar | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Maidu (Northeast) | SOV | Maiduan | Penutian | Yes |
| Maipure | SVO | Alto-Orinoco | Arawakan | No |
| Maisin | SOV | Oceanic | Austronesian | Yes |
| Majang | VSO | Surmic | Eastern Sudanic | Yes |
| Makah | VSO | Southern Wakashan | Wakashan | No |
| Makasae | SOV | Makasae-Fataluku-Oirata | Timor-Alor-Pantar | No |

| | | | | |
|---|---|---|---|---|
| Malayalam | SOV | Southern Dravidian | Dravidian | Yes |
| Maleu | SVO | Oceanic | Austronesian | No |
| Malgwa | SVO | Biu-Mandara | Afro-Asiatic | No |
| Mam | VSO | Mayan | Mayan | No |
| Mamanwa | VSO | Greater Central Philippine | Austronesian | No |
| Manadonese | SVO | Malayo-Sumbawan | Austronesian | No |
| Manam | SOV | Oceanic | Austronesian | Yes |
| Manchu | SOV | Tungusic | Altaic | Yes |
| Mandan | SOV | Core Siouan | Siouan | Yes |
| Mandarin | SVO | Chinese | Sino-Tibetan | No |
| Mangarrayi | OVS | Mangarrayi | Mangarrayi-Maran | Yes |
| Mangbetu | SVO | Mangbetu | Central Sudanic | No |
| Manggarai | SVO | Central Malayo-Polynesian | Austronesian | No |
| Mangghuer | SOV | Mongolic | Altaic | Yes |
| Maninka (Western) | SOV | Western Mande | Mande | Yes |
| Mano | SOV | Eastern Mande | Mande | No |
| Manobo (Western Bukidnon) | VSO | Greater Central Philippine | Austronesian | No |
| Mansi | SOV | Ugric | Uralic | Yes |
| Maori | VSO | Oceanic | Austronesian | No |
| Mapudungun | SVO | Araucanian | Araucanian | Yes |
| Mara | No dominant order | Mara | Mangarrayi-Maran | Yes |
| Marathi | SOV | Indic | Indo-European | Yes |
| Margany | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Margi | SVO | Biu-Mandara | Afro-Asiatic | No |
| Maricopa | SOV | Yuman | Hokan | Yes |
| Marind | SOV | Marind Proper | Marind | No |
| Martuthunira | SVO | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Maru | SOV | Burmese-Lolo | Sino-Tibetan | No |
| Masakin | No dominant order | Talodi | Kordofanian | No |
| Masalit | SOV | Maban | Maban | Yes |
| Matis | SOV | Panoan | Panoan | Yes |
| Mauka | SOV | Western Mande | Mande | No |
| Maung | SVO | Iwaidjan | Iwaidjan | No |
| Maybrat | SVO | North-Central Bird's Head | West Papuan | Yes |
| Mba | SVO | Ubangi | Niger-Congo | Yes |
| Mbay | SVO | Bongo-Bagirmi | Central Sudanic | Yes |
| Mbum | SVO | Adamawa | Niger-Congo | No |
| Me'en | SVO | Surmic | Eastern Sudanic | Yes |
| Mehri | No dominant order | Semitic | Afro-Asiatic | No |

| | | | | |
|---|---|---|---|---|
| Mekens | SOV | Tupari | Tupian | No |
| Mende | SOV | Western Mande | Mande | No |
| Menomini | No dominant order | Algonquian | Algic | Yes |
| Menya | SOV | Angan | Trans-New Guinea | Yes |
| Meryam Mir | SOV | Western Fly | Western Fly | Yes |
| Meyah | SVO | East Bird's Head | East Bird's Head | No |
| Mian | SOV | Ok | Trans-New Guinea | Yes |
| Mien | SVO | Hmong-Mien | Hmong-Mien | No |
| Mikasuki | SOV | Muskogean | Muskogean | Yes |
| Mikir | SOV | Kuki-Chin | Sino-Tibetan | Yes |
| Mina | SVO | Biu-Mandara | Afro-Asiatic | No |
| Minangkabau | No dominant order | Malayo-Sumbawan | Austronesian | No |
| Mising | SOV | Tani | Sino-Tibetan | Yes |
| Miwok (Southern Sierra) | No dominant order | Miwok | Penutian | Yes |
| Mixtec (Chalcatongo) | VSO | Mixtecan | Oto-Manguean | No |
| Mixtec (Jicaltepec) | VSO | Mixtecan | Oto-Manguean | No |
| Mixtec (Ocotepec) | VSO | Mixtecan | Oto-Manguean | No |
| Mixtec (Peñoles) | VSO | Mixtecan | Oto-Manguean | No |
| Miya | No dominant order | West Chadic | Afro-Asiatic | No |
| Mizo | SOV | Kuki-Chin | Sino-Tibetan | Yes |
| Mocoví | SVO | South Guaicuruan | Guaicuruan | No |
| Moghol | SOV | Mongolic | Altaic | Yes |
| Mohawk | No dominant order | Northern Iroquoian | Iroquoian | Yes |
| Mokilese | SVO | Oceanic | Austronesian | No |
| Momu | SOV | Fas | Kwomtari-Baibai | Yes |
| Momuna | SOV | Momuna | Trans-New Guinea | Yes |
| Mon | SVO | Monic | Austro-Asiatic | No |
| Mondunga | SVO | Ubangi | Niger-Congo | No |
| Mongol (Khamnigan) | SOV | Mongolic | Altaic | Yes |
| Mono (in United States) | SOV | Numic | Uto-Aztecan | Yes |
| Mooré | SVO | Gur | Niger-Congo | Yes |
| Mordvin (Erzya) | SVO | Mordvin | Uralic | Yes |
| Moro | SVO | Heiban | Kordofanian | Yes |
| Moru | No dominant order | Moru-Ma'di | Central Sudanic | No |
| Motu | SOV | Oceanic | Austronesian | No |
| Movima | No dominant order | Movima | Movima | Yes |
| Mufian | SVO | Kombio-Arapesh | Torricelli | No |
| Muisca | SOV | Chibcha-Duit | Chibchan | Yes |

| | | | | |
|---|---|---|---|---|
| Mumuye | SVO | Adamawa | Niger-Congo | No |
| Muna | SVO | Celebic | Austronesian | No |
| Mundang | SVO | Adamawa | Niger-Congo | No |
| Mupun | SVO | West Chadic | Afro-Asiatic | No |
| Murle | VSO | Surmic | Eastern Sudanic | Yes |
| Mursi | SVO | Surmic | Eastern Sudanic | Yes |
| Muruwari | No dominant order | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Musgu | SVO | Biu-Mandara | Afro-Asiatic | No |
| Mutsun | SVO | Costanoan | Penutian | Yes |
| Mwotlap | SVO | Oceanic | Austronesian | Yes |
| Nabak | SOV | Finisterre-Huon | Trans-New Guinea | Yes |
| Naga (Tangkhul) | SOV | Kuki-Chin | Sino-Tibetan | Yes |
| Nahuatl (Tetelcingo) | SVO | Aztecan | Uto-Aztecan | No |
| Nakanai | SVO | Oceanic | Austronesian | No |
| Nambikuára (Southern) | SOV | Nambikuaran | Nambikuaran | No |
| Nandi | VSO | Nilotic | Eastern Sudanic | Yes |
| Nara (in Ethiopia) | SOV | Nara | Eastern Sudanic | Yes |
| Nasioi | SOV | East Bougainville | East Bougainville | Yes |
| Natchez | SOV | Natchez | Natchez | Yes |
| Navajo | SOV | Athapaskan | Na-Dene | No |
| Ndebele (in South Africa) | SVO | Bantoid | Niger-Congo | Yes |
| Ndjébbana | No dominant order | Ndjébbana | Mangrida | Yes |
| Ndonga | SVO | Bantoid | Niger-Congo | Yes |
| Ndyuka | SVO | Creoles and Pidgins | other | No |
| Nenets | SOV | Samoyedic | Uralic | Yes |
| Neo-Aramaic (Arbel Jewish) | SOV | Semitic | Afro-Asiatic | Yes |
| Nepali | SOV | Indic | Indo-European | Yes |
| Nevome | SOV | Tepiman | Uto-Aztecan | No |
| Newar (Dolakha) | SOV | Mahakiranti | Sino-Tibetan | No |
| Newari (Kathmandu) | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Nez Perce | No dominant order | Sahaptian | Penutian | Yes |
| Ngaanyatjarra | SOV | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Ngäbere | SOV | Guaymiic | Chibchan | Yes |
| Ngad'a | SVO | Central Malayo-Polynesian | Austronesian | No |
| Ngalakan | No dominant order | Ngalakan | Gunwinyguan | Yes |
| Ngalkbun | SOV | Gunwinygic | Gunwinyguan | Yes |
| Ngambay | SVO | Bongo-Bagirmi | Central Sudanic | Yes |
| Nganasan | SOV | Samoyedic | Uralic | Yes |

| | | | | |
|---|---|---|---|---|
| Ngan'gityemerri | No dominant order | Ngankikurungkurr | Southern Daly | Yes |
| Ngankikurungkurr | No dominant order | Ngankikurungkurr | Southern Daly | Yes |
| Ngawun | SVO | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Ngiti | No dominant order | Lendu | Central Sudanic | No |
| Ngizim | SVO | West Chadic | Afro-Asiatic | No |
| Nhanda | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Nias | VOS | Northwest Sumatra-Barrier Islands | Austronesian | Yes |
| Nicobarese (Car) | VOS | Nicobarese | Austro-Asiatic | No |
| Nisgha | VSO | Tsimshianic | Penutian | No |
| Niuean | VSO | Oceanic | Austronesian | No |
| Nivkh | SOV | Nivkh | Nivkh | Yes |
| Nkore-Kiga | SVO | Bantoid | Niger-Congo | No |
| Nocte | SOV | Northern Naga | Sino-Tibetan | Yes |
| Noghay | SOV | Turkic | Altaic | Yes |
| Noni | SVO | Bantoid | Niger-Congo | Yes |
| Noon | SVO | Northern Atlantic | Niger-Congo | No |
| Norwegian | SVO | Germanic | Indo-European | Yes |
| Nuaulu | SVO | Central Malayo-Polynesian | Austronesian | No |
| Nubian (Dongolese) | SOV | Nubian | Eastern Sudanic | Yes |
| Nuer | No dominant order | Nilotic | Eastern Sudanic | Yes |
| Nung (in Vietnam) | SVO | Kam-Tai | Tai-Kadai | No |
| Nunggubuyu | No dominant order | Nunggubuyu | Gunwinyguan | Yes |
| Nupe | SVO | Nupoid | Niger-Congo | No |
| Nyimang | SOV | Nyimang | Eastern Sudanic | Yes |
| Nzakara | SVO | Ubangi | Niger-Congo | No |
| Obolo | SVO | Cross River | Niger-Congo | No |
| Ocuilteco | SVO | Matlatzincan | Oto-Manguean | Yes |
| Ogbronuagum | SVO | Cross River | Niger-Congo | No |
| Oirat | SOV | Mongolic | Altaic | Yes |
| Ojibwa (Eastern) | No dominant order | Algonquian | Algic | Yes |
| Oksapmin | SOV | Oksapmin | Oksapmin | Yes |
| Olo | SVO | Wapei-Palei | Torricelli | No |
| One | SVO | West Wapei | Torricelli | Yes |
| Onge | SOV | South Andamanese | South Andamanese | Yes |
| O'odham | No dominant order | Tepiman | Uto-Aztecan | No |
| Ordos | SOV | Mongolic | Altaic | Yes |

| Oriya | SOV | Indic | Indo-European | Yes |
|---|---|---|---|---|
| Orokaiva | SOV | Binanderean | Trans-New Guinea | No |
| Oromo (Harar) | SOV | Lowland East Cushitic | Afro-Asiatic | Yes |
| Oromo (Waata) | SOV | Lowland East Cushitic | Afro-Asiatic | Yes |
| Orya | SOV | Orya | Tor-Orya | Yes |
| Osage | SOV | Core Siouan | Siouan | No |
| Ossetic | SOV | Iranian | Indo-European | Yes |
| Otoro | SVO | Heiban | Kordofanian | Yes |
| Pa'a | SVO | West Chadic | Afro-Asiatic | No |
| Paakantyi | SVO | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Paamese | SVO | Oceanic | Austronesian | No |
| Páez | SOV | Páezan | Páezan | Yes |
| Paiwan | No dominant order | Paiwan | Austronesian | No |
| Palauan | SVO | Palauan | Austronesian | No |
| Palikur | SVO | Eastern Arawakan | Arawakan | No |
| Pame | SVO | Pamean | Oto-Manguean | No |
| Panjabi | SOV | Indic | Indo-European | Yes |
| Panyjima | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Päri | OVS | Nilotic | Eastern Sudanic | Yes |
| Pashto | SOV | Iranian | Indo-European | Yes |
| Pattani | SOV | Bodic | Sino-Tibetan | Yes |
| Patwin | No dominant order | Wintuan | Penutian | Yes |
| Paulohi | SVO | Central Malayo-Polynesian | Austronesian | No |
| Paumarí | SVO | Arauan | Arauan | Yes |
| Pawaian | SOV | Pawaian | Teberan-Pawaian | No |
| Pero | SVO | West Chadic | Afro-Asiatic | No |
| Pilagá | SVO | South Guaicuruan | Guaicuruan | No |
| Pima Bajo | SOV | Tepiman | Uto-Aztecan | Yes |
| Pipil | No dominant order | Aztecan | Uto-Aztecan | No |
| Piro | SOV | Purus | Arawakan | Yes |
| Pitta Pitta | No dominant order | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Pohnpeian | SVO | Oceanic | Austronesian | No |
| Poko-Rawo | No dominant order | Serra Hills | Skou | No |
| Pokot | VSO | Nilotic | Eastern Sudanic | Yes |
| Polish | SVO | Slavic | Indo-European | Yes |
| Pomo (Eastern) | SOV | Pomoan | Hokan | Yes |
| Pomo (Southeastern) | SOV | Pomoan | Hokan | Yes |
| Popoluca (Sierra) | SVO | Mixe-Zoque | Mixe-Zoque | Yes |
| Port Sandwich | SVO | Oceanic | Austronesian | No |

| | | | | |
|---|---|---|---|---|
| Portuguese | SVO | Romance | Indo-European | No |
| Prasuni | SOV | Nuristani | Indo-European | Yes |
| Pumi | SOV | Qiangic | Sino-Tibetan | No |
| Purépecha | SVO | Tarascan | Tarascan | Yes |
| Purki | SOV | Bodic | Sino-Tibetan | Yes |
| Quechua (Huallaga) | SOV | Quechuan | Quechuan | Yes |
| Quechua (Imbabura) | SOV | Quechuan | Quechuan | Yes |
| Quileute | VSO | Chimakuan | Chimakuan | Yes |
| Rapanui | VSO | Oceanic | Austronesian | No |
| Rashad | SOV | Rashad | Kordofanian | No |
| Rawang | SOV | Nungish | Sino-Tibetan | No |
| Rembarnga | SOV | Rembarnga | Gunwinyguan | Yes |
| Resígaro | SOV | Inland Northern Arawakan | Arawakan | Yes |
| Retuarã | SOV | Tucanoan | Tucanoan | Yes |
| Ritharngu | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Romani (Welsh) | No dominant order | Indic | Indo-European | Yes |
| Romanian | SVO | Romance | Indo-European | Yes |
| Rotuman | SVO | Oceanic | Austronesian | No |
| Rukai (Tanan) | No dominant order | Rukai | Austronesian | No |
| Rumu | SOV | Turama-Kikorian | Turama-Kikorian | No |
| Runga | SOV | Maban | Maban | No |
| Russian | SVO | Slavic | Indo-European | Yes |
| Rutul | SOV | Lezgic | Nakh-Daghestanian | Yes |
| Saami (Northern) | SVO | Saami | Uralic | Yes |
| Sahaptin (Northern) | VSO | Sahaptian | Penutian | Yes |
| Sahu | SVO | North Halmaheran | West Papuan | No |
| Saliba (in Papua New Guinea) | SOV | Oceanic | Austronesian | No |
| Salinan | SVO | Salinan | Salinan | Yes |
| Salt-Yui | SOV | Chimbu | Trans-New Guinea | No |
| Samoan | No dominant order | Oceanic | Austronesian | No |
| Sandawe | SOV | Sandawe | Sandawe | Yes |
| Sango | SVO | Ubangi | Niger-Congo | No |
| Santa | SOV | Mongolic | Altaic | Yes |
| Santali | SOV | Munda | Austro-Asiatic | Yes |
| Sanuma | SOV | Yanomam | Yanomam | Yes |
| Sarcee | SOV | Athapaskan | Na-Dene | No |
| Sare | SOV | Sepik Hill | Sepik | Yes |
| Savi | SOV | Indic | Indo-European | Yes |
| Savosavo | SOV | Savosavo | Solomons East Papuan | Yes |

| | | | | |
|---|---|---|---|---|
| Sedang | SVO | Bahnaric | Austro-Asiatic | No |
| Selkup | SOV | Samoyedic | Uralic | Yes |
| Sentani | SOV | Sentani | Sentani | No |
| Serbian-Croatian | SVO | Slavic | Indo-European | Yes |
| Seri | SOV | Seri | Hokan | No |
| Shabo | SOV | Shabo | Shabo | Yes |
| Sharanahua | SOV | Panoan | Panoan | Yes |
| Shatt | SVO | Daju | Eastern Sudanic | No |
| Shilluk | SVO | Nilotic | Eastern Sudanic | Yes |
| Shina | SOV | Indic | Indo-European | Yes |
| Shipibo-Konibo | SOV | Panoan | Panoan | Yes |
| Shiriana | SOV | Yanomam | Yanomam | Yes |
| Shoshone | SOV | Numic | Uto-Aztecan | Yes |
| Shuswap | No dominant order | Interior Salish | Salishan | Yes |
| Siane | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Sidaama | SOV | Highland East Cushitic | Afro-Asiatic | Yes |
| Sila | SOV | Daju | Eastern Sudanic | No |
| Sinaugoro | SOV | Oceanic | Austronesian | Yes |
| Sinhala | SOV | Indic | Indo-European | Yes |
| Sio | SVO | Oceanic | Austronesian | No |
| Sipakapense | VSO | Mayan | Mayan | No |
| Siroi | SOV | Madang | Trans-New Guinea | Yes |
| Sisiqa | SVO | Oceanic | Austronesian | No |
| Siuslaw | No dominant order | Siuslawan | Oregon Coast | Yes |
| Skou | SOV | Western Skou | Skou | Yes |
| Slave | SOV | Athapaskan | Na-Dene | No |
| Slovene | SVO | Slavic | Indo-European | Yes |
| So | VSO | Kuliak | Eastern Sudanic | Yes |
| Sobei | SVO | Oceanic | Austronesian | No |
| Sonsorol-Tobi | SVO | Oceanic | Austronesian | No |
| Spanish | SVO | Romance | Indo-European | No |
| Squamish | VSO | Central Salish | Salishan | Yes |
| Stieng | SVO | Bahnaric | Austro-Asiatic | No |
| Suena | SOV | Binanderean | Trans-New Guinea | Yes |
| Suki | SOV | Suki | Gogodala-Suki | Yes |
| Sulka | SVO | Sulka | Sulka | No |
| Sundanese | SVO | Malayo-Sumbawan | Austronesian | No |
| Taba | SVO | South Halmahera - West New Guinea | Austronesian | No |
| Tacana | No dominant order | Tacanan | Tacanan | No |
| Tagalog | VSO | Greater Central Philippine | Austronesian | No |
| Tahitian | VSO | Oceanic | Austronesian | No |
| Tairora | SOV | Eastern Highlands | Trans-New Guinea | Yes |

| | | | | |
|---|---|---|---|---|
| Tajik | SOV | Iranian | Indo-European | No |
| Takelma | SOV | Takelma | Takelma | Yes |
| Talinga | SVO | Bantoid | Niger-Congo | Yes |
| Tama | SOV | Taman | Eastern Sudanic | Yes |
| Tamagario | SOV | Kayagar | Kayagar | No |
| Tamang (Eastern) | SOV | Bodic | Sino-Tibetan | Yes |
| Tamashek | VSO | Berber | Afro-Asiatic | Yes |
| Tamil | SOV | Southern Dravidian | Dravidian | Yes |
| Tapieté | SOV | Tupi-Guaraní | Tupian | Yes |
| Tarahumara (Central) | SOV | Tarahumaran | Uto-Aztecan | Yes |
| Tarahumara (Western) | SOV | Tarahumaran | Uto-Aztecan | Yes |
| Tarao | SOV | Kuki-Chin | Sino-Tibetan | Yes |
| Tariana | SOV | Inland Northern Arawakan | Arawakan | Yes |
| Tashlhiyt | No dominant order | Berber | Afro-Asiatic | Yes |
| Tatar | SOV | Turkic | Altaic | Yes |
| Tauya | SOV | Madang | Trans-New Guinea | Yes |
| Tawala | SOV | Oceanic | Austronesian | No |
| Tboli | VSO | Bilic | Austronesian | No |
| Telugu | SOV | South-Central Dravidian | Dravidian | Yes |
| Temein | SVO | Temein | Eastern Sudanic | Yes |
| Temiar | SVO | Aslian | Austro-Asiatic | Yes |
| Tennet | VSO | Surmic | Eastern Sudanic | Yes |
| Teop | SVO | Oceanic | Austronesian | No |
| Tepehua (Huehuetla) | No dominant order | Totonacan | Totonacan | Yes |
| Tepehua (Tlachichilco) | SVO | Totonacan | Totonacan | Yes |
| Tepehuan (Northern) | VSO | Tepiman | Uto-Aztecan | Yes |
| Tera | SVO | Biu-Mandara | Afro-Asiatic | No |
| Teribe | SOV | Talamanca | Chibchan | No |
| Teso | VSO | Nilotic | Eastern Sudanic | Yes |
| Tetun | SVO | Central Malayo-Polynesian | Austronesian | No |
| Thai | SVO | Kam-Tai | Tai-Kadai | No |
| Thangmi | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Thulung | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Tibetan (Modern Literary) | SOV | Bodic | Sino-Tibetan | Yes |
| Ticuna | No dominant order | Ticuna | Ticuna | Yes |
| Tidore | SVO | North Halmaheran | West Papuan | No |
| Tigak | SVO | Oceanic | Austronesian | No |
| Tigré | SOV | Semitic | Afro-Asiatic | No |

| | | | | |
|---|---|---|---|---|
| Tigrinya | SOV | Semitic | Afro-Asiatic | Yes |
| Tiipay (Jamul) | SOV | Yuman | Hokan | Yes |
| Tima | SVO | Katla-Tima | Kordofanian | No |
| Timugon | VSO | North Borneo | Austronesian | No |
| Tinrin | No dominant order | Oceanic | Austronesian | No |
| Tiriyo | OVS | Cariban | Cariban | No |
| Tirmaga | SVO | Surmic | Eastern Sudanic | Yes |
| Tiwa (Northern) | No dominant order | Kiowa-Tanoan | Kiowa-Tanoan | Yes |
| Tiwi | SVO | Tiwian | Tiwian | No |
| Tlapanec | VSO | Subtiaba-Tlapanec | Oto-Manguean | Yes |
| Tlingit | SOV | Tlingit | Na-Dene | Yes |
| Toba | No dominant order | South Guaicuruan | Guaicuruan | No |
| Tobati | OSV | Oceanic | Austronesian | Yes |
| Tobelo | SOV | North Halmaheran | West Papuan | Yes |
| Tolai | SVO | Oceanic | Austronesian | No |
| Tonga (in Zambia) | SVO | Bantoid | Niger-Congo | Yes |
| Tongan | No dominant order | Oceanic | Austronesian | No |
| Tonkawa | No dominant order | Tonkawa | Tonkawa | Yes |
| Trique (Copala) | VSO | Mixtecan | Oto-Manguean | No |
| Trumai | No dominant order | Trumai | Trumai | Yes |
| Tsafiki | SOV | Barbacoan | Barbacoan | Yes |
| Tsez | SOV | Avar-Andic-Tsezic | Nakh-Daghestanian | Yes |
| Tshangla | SOV | Bodic | Sino-Tibetan | Yes |
| Tsimshian (Coast) | VSO | Tsimshianic | Penutian | No |
| Tsova-Tush | SOV | Nakh | Nakh-Daghestanian | Yes |
| Tubu | SOV | Western Saharan | Saharan | No |
| Tukang Besi | VOS | Celebic | Austronesian | No |
| Tulu | SOV | Southern Dravidian | Dravidian | Yes |
| Tümpisa Shoshone | SOV | Numic | Uto-Aztecan | Yes |
| Tunen | SOV | Bantoid | Niger-Congo | No |
| Turkana | VSO | Nilotic | Eastern Sudanic | Yes |
| Turkish | SOV | Turkic | Altaic | Yes |
| Tuscarora | No dominant order | Northern Iroquoian | Iroquoian | Yes |
| Tutelo | SOV | Core Siouan | Siouan | No |
| Tuvaluan | OVS | Oceanic | Austronesian | No |
| Tuvan | SOV | Turkic | Altaic | Yes |
| Tuyuca | SOV | Tucanoan | Tucanoan | Yes |
| Tzutujil | No dominant order | Mayan | Mayan | No |
| Ubykh | SOV | Northwest Caucasian | Northwest Caucasian | Yes |

| | | | | |
|---|---|---|---|---|
| Udi | SOV | Lezgic | Nakh-Daghestanian | Yes |
| Udihe | SOV | Tungusic | Altaic | Yes |
| Udmurt | SOV | Permic | Uralic | Yes |
| Ukrainian | SVO | Slavic | Indo-European | Yes |
| Uldeme | SVO | Biu-Mandara | Afro-Asiatic | No |
| Ulithian | SVO | Oceanic | Austronesian | No |
| Una | SOV | Mek | Trans-New Guinea | No |
| Ungarinjin | OVS | Worrorran | Worrorran | Yes |
| Uradhi | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Urak Lawoi' | SVO | Malayo-Sumbawan | Austronesian | No |
| Urarina | OVS | Urarina | Urarina | No |
| Urat | SVO | Wapei-Palei | Torricelli | No |
| Urubú-Kaapor | SOV | Tupi-Guaraní | Tupian | No |
| Usarufa | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Ute | No dominant order | Numic | Uto-Aztecan | Yes |
| Uyghur | SOV | Turkic | Altaic | Yes |
| Uzbek | SOV | Turkic | Altaic | Yes |
| Vai | SOV | Western Mande | Mande | No |
| Vietnamese | SVO | Viet-Muong | Austro-Asiatic | No |
| Wagiman | SOV | Wagiman | Wagiman | Yes |
| Wahgi | SOV | Chimbu | Trans-New Guinea | Yes |
| Wakhi | SOV | Iranian | Indo-European | Yes |
| Walman | SVO | Wapei-Palei | Torricelli | No |
| Wambaya | No dominant order | Wambayan | Mirndi | Yes |
| Wambon | SOV | Awju-Dumut | Trans-New Guinea | Yes |
| Wambule | SOV | Mahakiranti | Sino-Tibetan | Yes |
| Wangkumara | No dominant order | Central Pama-Nyungan | Pama-Nyungan | Yes |
| Wappo | SOV | Wappo | Wappo-Yukian | Yes |
| Waray (in Australia) | SVO | Warayic | Gunwinyguan | Yes |
| Wardaman | No dominant order | Yangmanic | Yangmanic | Yes |
| Warekena | SVO | Inland Northern Arawakan | Arawakan | Yes |
| Wari' | VOS | Chapacura-Wanham | Chapacura-Wanham | No |
| Warlpiri | No dominant order | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Warndarang | SVO | Warndarang | Mangarrayi-Maran | Yes |
| Warrgamay | No dominant order | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Warrwa | No dominant order | Nyulnyulan | Nyulnyulan | Yes |
| Waskia | SOV | Madang | Trans-New Guinea | No |

| | | | | |
|---|---|---|---|---|
| Wathawurrung | VOS | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Wedau | SOV | Oceanic | Austronesian | Yes |
| Welsh | VSO | Celtic | Indo-European | No |
| Wembawemba | VOS | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| West Makian | SVO | North Halmaheran | West Papuan | No |
| Wichí | SVO | Matacoan | Matacoan | No |
| Wik Munkan | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Wik Ngathana | OSV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Wikchamni | No dominant order | Yokuts | Penutian | Yes |
| Wintu | No dominant order | Wintuan | Penutian | Yes |
| Wiyot | No dominant order | Wiyot | Algic | Yes |
| Wolaytta | SOV | North Omotic | Afro-Asiatic | Yes |
| Woleaian | SVO | Oceanic | Austronesian | No |
| Wolio | VSO | Celebic | Austronesian | No |
| Wolof | SVO | Northern Atlantic | Niger-Congo | No |
| Womo | SOV | Serra Hills | Skou | Yes |
| Xârâcùù | SVO | Oceanic | Austronesian | No |
| Xasonga | SOV | Western Mande | Mande | No |
| Xhosa | SVO | Bantoid | Niger-Congo | Yes |
| Yagaria | SOV | Eastern Highlands | Trans-New Guinea | Yes |
| Yahgan | SOV | Yámana | Yámana | Yes |
| Yakut | SOV | Turkic | Altaic | Yes |
| Yale (Kosarek) | SOV | Mek | Trans-New Guinea | Yes |
| Yaminahua | SOV | Panoan | Panoan | Yes |
| Yapese | VSO | Yapese | Austronesian | No |
| Yaqui | SOV | Cahita | Uto-Aztecan | Yes |
| Yareba | SOV | Yareban | Yareban | Yes |
| Yawelmani | No dominant order | Yokuts | Penutian | Yes |
| Yawuru | No dominant order | Nyulnyulan | Nyulnyulan | Yes |
| Yidiny | SOV | Northern Pama-Nyungan | Pama-Nyungan | Yes |
| Yindjibarndi | SVO | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Yingkarta | SVO | Western Pama-Nyungan | Pama-Nyungan | Yes |
| Yokuts (Yaudanchi) | No dominant order | Yokuts | Penutian | Yes |
| Yoruba | SVO | Defoid | Niger-Congo | No |
| Yukaghir (Kolyma) | SOV | Yukaghir | Yukaghir | Yes |
| Yukaghir (Tundra) | SOV | Yukaghir | Yukaghir | Yes |
| Yukulta | SVO | Tangkic | Tangkic | Yes |

| | | | | |
|---|---|---|---|---|
| Yulu | SVO | Bongo-Bagirmi | Central Sudanic | No |
| Yup'ik (Central) | No dominant order | Eskimo | Eskimo-Aleut | Yes |
| Yuwaalaraay | SOV | Southeastern Pama-Nyungan | Pama-Nyungan | Yes |
| Zande | SVO | Ubangi | Niger-Congo | Yes |
| Zaparo | SVO | Zaparoan | Zaparoan | Yes |
| Zapotec (Isthmus) | VSO | Zapotecan | Oto-Manguean | No |
| Zapotec (Mitla) | VSO | Zapotecan | Oto-Manguean | Yes |
| Zapotec (Zoogocho) | VSO | Zapotecan | Oto-Manguean | No |
| Zayse | SOV | North Omotic | Afro-Asiatic | Yes |
| Zhuang (Northern) | SVO | Kam-Tai | Tai-Kadai | No |
| Zoque (Copainalá) | VOS | Mixe-Zoque | Mixe-Zoque | Yes |
| Zoque (Ostuacan) | No dominant order | Mixe-Zoque | Mixe-Zoque | Yes |

## 9.2 Sample B

| Language | Word order | Genus | Family | Number of cases |
|---|---|---|---|---|
| Abipón | SVO | South Guaicuruan | Guaicuruan | No morphological case-marking |
| Abkhaz | SOV | Northwest Caucasian | Northwest Caucasian | 2 cases |
| Acoma | No dominant order | Keresan | Keresan | No morphological case-marking |
| Alamblak | SOV | Sepik Hill | Sepik | 8-9 cases |
| Albanian | SVO | Albanian | Indo-European | 4 cases |
| Aleut | SOV | Aleut | Eskimo-Aleut | 2 cases |
| Amele | SOV | Madang | Trans-New Guinea | No morphological case-marking |
| Amharic | SOV | Semitic | Afro-Asiatic | 2 cases |
| Arabic (Egyptian) | SVO | Semitic | Afro-Asiatic | No morphological case-marking |
| Arapesh (Mountain) | SVO | Kombio-Arapesh | Torricelli | No morphological case-marking |
| Armenian (Eastern) | No dominant order | Armenian | Indo-European | 5 cases |
| Asmat | SOV | Asmat-Kamoro | Trans-New Guinea | No morphological case-marking |
| Awa Pit | SOV | Barbacoan | Barbacoan | 10 or more cases |
| Aymara (Central) | SOV | Aymaran | Aymaran | 6-7 cases |
| Bagirmi | SVO | Bongo-Bagirmi | Central Sudanic | No morphological case-marking |
| Bambara | SOV | Western Mande | Mande | No morphological case-marking |
| Barasano | No dominant order | Tucanoan | Tucanoan | 2 cases |
| Basque | SOV | Basque | Basque | 10 or more cases |
| Batak (Karo) | No dominant order | Northwest Sumatra-Barrier Islands | Austronesian | No morphological case-marking |
| Bawm | SOV | Kuki-Chin | Sino-Tibetan | 4 cases |
| Beja | SOV | Beja | Afro-Asiatic | 2 cases |
| Berber (Middle Atlas) | VSO | Berber | Afro-Asiatic | 2 cases |
| Bulgarian | SVO | Slavic | Indo-European | No morphological case-marking |
| Burmese | SOV | Burmese-Lolo | Sino-Tibetan | 8-9 cases |
| Burushaski | SOV | Burushaski | Burushaski | 8-9 cases |
| Cahuilla | SOV | California Uto-Aztecan | Uto-Aztecan | 5 cases |
| Canela-Krahô | SOV | Ge-Kaingang | Macro-Ge | No morphological case-marking |
| Carib | SOV | Cariban | Cariban | No morphological case-marking |

| | | | | |
|---|---|---|---|---|
| Catalan | SVO | Romance | Indo-European | No morphological case-marking |
| Chamorro | VSO | Chamorro | Austronesian | No morphological case-marking |
| Chinantec (Lealao) | VOS | Chinantecan | Oto-Manguean | No morphological case-marking |
| Chukchi | No dominant order | Northern Chukotko-Kamchatkan | Chukotko-Kamchatkan | 10 or more cases |
| Chumash (Barbareño) | VOS | Chumash | Chumash | No morphological case-marking |
| Chuvash | SOV | Turkic | Altaic | 6-7 cases |
| Comanche | SOV | Numic | Uto-Aztecan | 3 cases |
| Coos (Hanis) | No dominant order | Coosan | Oregon Coast | 8-9 cases |
| Dani (Lower Grand Valley) | SOV | Dani | Trans-New Guinea | 6-7 cases |
| Diola-Fogny | SVO | Northern Atlantic | Niger-Congo | No morphological case-marking |
| Dong (Southern) | SVO | Kam-Tai | Tai-Kadai | No morphological case-marking |
| Drehu | No dominant order | Oceanic | Austronesian | No morphological case-marking |
| Dutch | No dominant order | Germanic | Indo-European | No morphological case-marking |
| Dyirbal | No dominant order | Northern Pama-Nyungan | Pama-Nyungan | 6-7 cases |
| English | SVO | Germanic | Indo-European | 2 cases |
| Epena Pedee | SOV | Choco | Choco | 10 or more cases |
| Estonian | SVO | Finnic | Uralic | 10 or more cases |
| Evenki | SOV | Tungusic | Altaic | 10 or more cases |
| Ewe | SVO | Kwa | Niger-Congo | No morphological case-marking |
| Fijian | No dominant order | Oceanic | Austronesian | No morphological case-marking |
| Finnish | SVO | Finnic | Uralic | 10 or more cases |
| French | SVO | Romance | Indo-European | No morphological case-marking |
| Fula (Cameroonian) | SVO | Northern Atlantic | Niger-Congo | No morphological case-marking |
| Fur | SOV | Fur | Fur | 4 cases |
| Garo | SOV | Bodo-Garo | Sino-Tibetan | 8-9 cases |
| Georgian | SOV | Kartvelian | Kartvelian | 6-7 cases |
| German | No dominant order | Germanic | Indo-European | 4 cases |
| Gimira | SOV | North Omotic | Afro-Asiatic | 6-7 cases |
| Gooniyandi | No dominant order | Bunuban | Bunuban | 10 or more cases |

| | | | | |
|---|---|---|---|---|
| Grebo | SVO | Kru | Niger-Congo | No morphological case-marking |
| Greek (Modern) | No dominant order | Greek | Indo-European | 3 cases |
| Greenlandic (West) | SOV | Eskimo | Eskimo-Aleut | 8-9 cases |
| Guaraní | SVO | Tupi-Guaraní | Tupian | No morphological case-marking |
| Haida | SOV | Haida | Haida | No morphological case-marking |
| Hamtai | SOV | Angan | Trans-New Guinea | 10 or more cases |
| Hausa | SVO | West Chadic | Afro-Asiatic | No morphological case-marking |
| Hebrew (Modern) | SVO | Semitic | Afro-Asiatic | No morphological case-marking |
| Hixkaryana | OVS | Cariban | Cariban | No morphological case-marking |
| Hmong Njua | SVO | Hmong-Mien | Hmong-Mien | No morphological case-marking |
| Hua | SOV | Eastern Highlands | Trans-New Guinea | 8-9 cases |
| Huave (San Mateo del Mar) | SVO | Huavean | Huavean | No morphological case-marking |
| Huitoto (Minica) | SOV | Huitoto | Huitotoan | 6-7 cases |
| Hungarian | No dominant order | Ugric | Uralic | 10 or more cases |
| Hunzib | SOV | Avar-Andic-Tsezic | Nakh-Daghestanian | 10 or more cases |
| Iaai | SVO | Oceanic | Austronesian | No morphological case-marking |
| Icelandic | SVO | Germanic | Indo-European | 4 cases |
| Igbo | SVO | Igboid | Niger-Congo | No morphological case-marking |
| Ika | SOV | Arhuacic | Chibchan | 6-7 cases |
| Ilocano | VSO | Northern Luzon | Austronesian | No morphological case-marking |
| Indonesian | SVO | Malayo-Sumbawan | Austronesian | No morphological case-marking |
| Irish | VSO | Celtic | Indo-European | 2 cases |
| Italian | SVO | Romance | Indo-European | No morphological case-marking |
| Jakaltek | VSO | Mayan | Mayan | No morphological case-marking |
| Japanese | SOV | Japanese | Japanese | 8-9 cases |
| Ju\|'hoan | SVO | Ju-Kung | Kxa | No morphological case-marking |
| Kalispel | VSO | Interior Salish | Salishan | 6-7 cases |
| Kannada | SOV | Southern Dravidian | Dravidian | 6-7 cases |
| Kanuri | SOV | Western Saharan | Saharan | 6-7 cases |
| Karok | No dominant order | Karok | Karok | 3 cases |

| | | | | |
|---|---|---|---|---|
| Kashmiri | SVO | Indic | Indo-European | 4 cases |
| Kayah Li (Eastern) | SVO | Karen | Sino-Tibetan | No morphological case-marking |
| Kayardild | No dominant order | Tangkic | Tangkic | 10 or more cases |
| Ket | SOV | Yeniseian | Yeniseian | 10 or more cases |
| Kewa | SOV | Engan | Trans-New Guinea | 6-7 cases |
| Khalkha | SOV | Mongolic | Altaic | 8-9 cases |
| Khanty | SOV | Ugric | Uralic | 3 cases |
| Khasi | SVO | Khasian | Austro-Asiatic | No morphological case-marking |
| Khmer | SVO | Khmer | Austro-Asiatic | No morphological case-marking |
| Khmu' | SVO | Palaung-Khmuic | Austro-Asiatic | No morphological case-marking |
| Khoekhoe | SOV | Khoe-Kwadi | Khoe-Kwadi | 2 cases |
| Kilivila | No dominant order | Oceanic | Austronesian | No morphological case-marking |
| Kinyarwanda | SVO | Bantoid | Niger-Congo | No morphological case-marking |
| Kiribati | VOS | Oceanic | Austronesian | No morphological case-marking |
| Koasati | SOV | Muskogean | Muskogean | 6-7 cases |
| Kobon | SOV | Madang | Trans-New Guinea | No morphological case-marking |
| Kombai | SOV | Awju-Dumut | Trans-New Guinea | No morphological case-marking |
| Kongo | SVO | Bantoid | Niger-Congo | No morphological case-marking |
| Korean | SOV | Korean | Korean | 6-7 cases |
| Koromfe | SVO | Gur | Niger-Congo | No morphological case-marking |
| Kosraean | SVO | Oceanic | Austronesian | No morphological case-marking |
| Koyraboro Senni | SOV | Songhay | Songhay | No morphological case-marking |
| Krongo | VSO | Kadugli | Kadu | 6-7 cases |
| Kunama | SOV | Kunama | Kunama | 6-7 cases |
| Kutenai | No dominant order | Kutenai | Kutenai | No morphological case-marking |
| Ladakhi | SOV | Bodic | Sino-Tibetan | 5 cases |
| Lakhota | SOV | Core Siouan | Siouan | No morphological case-marking |
| Lango | SVO | Nilotic | Eastern Sudanic | No morphological case-marking |
| Latvian | SVO | Baltic | Indo-European | 5 cases |
| Lavukaleve | SOV | Lavukaleve | Solomons East Papuan | No morphological case-marking |
| Lepcha | SOV | Lepcha | Sino-Tibetan | 2 cases |
| Lezgian | SOV | Lezgic | Nakh-Daghestanian | 10 or more cases |

| | | | | |
|---|---|---|---|---|
| Lithuanian | SVO | Baltic | Indo-European | 6-7 cases |
| Luvale | SVO | Bantoid | Niger-Congo | No morphological case-marking |
| Maba | SOV | Maban | Maban | 3 cases |
| Makah | VSO | Southern Wakashan | Wakashan | No morphological case-marking |
| Malagasy | VOS | Barito | Austronesian | No morphological case-marking |
| Malayalam | SOV | Southern Dravidian | Dravidian | 6-7 cases |
| Mandarin | SVO | Chinese | Sino-Tibetan | No morphological case-marking |
| Mangarrayi | OVS | Mangarrayi | Mangarrayi-Maran | 8-9 cases |
| Maori | VSO | Oceanic | Austronesian | No morphological case-marking |
| Mapudungun | SVO | Araucanian | Araucanian | 2 cases |
| Marathi | SOV | Indic | Indo-European | 5 cases |
| Maricopa | SOV | Yuman | Hokan | 6-7 cases |
| Marind | SOV | Marind Proper | Marind | No morphological case-marking |
| Martuthunira | SVO | Western Pama-Nyungan | Pama-Nyungan | 10 or more cases |
| Maung | SVO | Iwaidjan | Iwaidjan | No morphological case-marking |
| Maybrat | SVO | North-Central Bird's Head | West Papuan | 2 cases |
| Miwok (Southern Sierra) | No dominant order | Miwok | Penutian | 6-7 cases |
| Mixtec (Chalcatongo) | VSO | Mixtecan | Oto-Manguean | No morphological case-marking |
| Mokilese | SVO | Oceanic | Austronesian | No morphological case-marking |
| Mordvin (Erzya) | SVO | Mordvin | Uralic | 10 or more cases |
| Mundari | SOV | Munda | Austro-Asiatic | 8-9 cases |
| Murle | VSO | Surmic | Eastern Sudanic | 4 cases |
| Nahuatl (Tetelcingo) | SVO | Aztecan | Uto-Aztecan | No morphological case-marking |
| Navajo | SOV | Athapaskan | Na-Dene | No morphological case-marking |
| Ndyuka | SVO | Creoles and Pidgins | other | No morphological case-marking |
| Nenets | SOV | Samoyedic | Uralic | 6-7 cases |
| Nez Perce | No dominant order | Sahaptian | Penutian | 10 or more cases |
| Ngiti | No dominant order | Lendu | Central Sudanic | No morphological case-marking |
| Nivkh | SOV | Nivkh | Nivkh | 8-9 cases |
| Nkore-Kiga | SVO | Bantoid | Niger-Congo | No morphological case-marking |
| Nubian (Dongolese) | SOV | Nubian | Eastern Sudanic | 6-7 cases |

| | | | | |
|---|---|---|---|---|
| Nunggubuyu | No dominant order | Nunggubuyu | Gunwinyguan | 10 or more cases |
| Oneida | No dominant order | Northern Iroquoian | Iroquoian | No morphological case-marking |
| Oromo (Harar) | SOV | Lowland East Cushitic | Afro-Asiatic | 6-7 cases |
| Paamese | SVO | Oceanic | Austronesian | No morphological case-marking |
| Páez | SOV | Páezan | Páezan | 6-7 cases |
| Paiwan | No dominant order | Paiwan | Austronesian | No morphological case-marking |
| Palauan | SVO | Palauan | Austronesian | No morphological case-marking |
| Panjabi | SOV | Indic | Indo-European | 2 cases |
| Pashto | SOV | Iranian | Indo-European | 3 cases |
| Paumarí | SVO | Arauan | Arauan | 3 cases |
| Persian | SOV | Iranian | Indo-European | 2 cases |
| Polish | SVO | Slavic | Indo-European | 6-7 cases |
| Pomo (Southeastern) | SOV | Pomoan | Hokan | 6-7 cases |
| Quechua (Imbabura) | SOV | Quechuan | Quechuan | 8-9 cases |
| Rama | SOV | Rama | Chibchan | 8-9 cases |
| Rapanui | VSO | Oceanic | Austronesian | No morphological case-marking |
| Romanian | SVO | Romance | Indo-European | 2 cases |
| Russian | SVO | Slavic | Indo-European | 6-7 cases |
| Saami (Northern) | SVO | Saami | Uralic | 6-7 cases |
| Sango | SVO | Ubangi | Niger-Congo | No morphological case-marking |
| Sanuma | SOV | Yanomam | Yanomam | 2 cases |
| Semelai | No dominant order | Aslian | Austro-Asiatic | 3 cases |
| Sentani | SOV | Sentani | Sentani | No morphological case-marking |
| Serbian-Croatian | SVO | Slavic | Indo-European | 5 cases |
| Shipibo-Konibo | SOV | Panoan | Panoan | 6-7 cases |
| Sinhala | SOV | Indic | Indo-European | 5 cases |
| Slave | SOV | Athapaskan | Na-Dene | No morphological case-marking |
| Somali | SOV | Lowland East Cushitic | Afro-Asiatic | 3 cases |
| Spanish | SVO | Romance | Indo-European | No morphological case-marking |
| Squamish | VSO | Central Salish | Salishan | 2 cases |
| Suena | SOV | Binanderean | Trans-New Guinea | 4 cases |
| Swahili | SVO | Bantoid | Niger-Congo | No morphological case-marking |
| Swedish | SVO | Germanic | Indo-European | 2 cases |

| | | | | |
|---|---|---|---|---|
| Taba | SVO | South Halmahera - West New Guinea | Austronesian | No morphological case-marking |
| Tagalog | VSO | Greater Central Philippine | Austronesian | No morphological case-marking |
| Thai | SVO | Kam-Tai | Tai-Kadai | No morphological case-marking |
| Tinrin | No dominant order | Oceanic | Austronesian | No morphological case-marking |
| Tiwi | SVO | Tiwian | Tiwian | No morphological case-marking |
| Tlingit | SOV | Tlingit | Na-Dene | 8-9 cases |
| Trumai | No dominant order | Trumai | Trumai | 5 cases |
| Tsimshian (Coast) | VSO | Tsimshianic | Penutian | No morphological case-marking |
| Tukang Besi | VOS | Celebic | Austronesian | No morphological case-marking |
| Turkana | VSO | Nilotic | Eastern Sudanic | 6-7 cases |
| Turkish | SOV | Turkic | Altaic | 6-7 cases |
| Tuvaluan | OVS | Oceanic | Austronesian | No morphological case-marking |
| Udihe | SOV | Tungusic | Altaic | 8-9 cases |
| Udmurt | SOV | Permic | Uralic | 10 or more cases |
| Una | SOV | Mek | Trans-New Guinea | No morphological case-marking |
| Ungarinjin | OVS | Worrorran | Worrorran | 8-9 cases |
| Urdu | SOV | Indic | Indo-European | 2 cases |
| Urubú-Kaapor | SOV | Tupi-Guaraní | Tupian | No morphological case-marking |
| Vietnamese | SVO | Viet-Muong | Austro-Asiatic | No morphological case-marking |
| Wambaya | No dominant order | Wambayan | Mirndi | 8-9 cases |
| Wardaman | No dominant order | Yangmanic | Yangmanic | 8-9 cases |
| Wari' | VOS | Chapacura-Wanham | Chapacura-Wanham | No morphological case-marking |
| Welsh | VSO | Celtic | Indo-European | No morphological case-marking |
| Wichí | SVO | Matacoan | Matacoan | No morphological case-marking |
| Wintu | No dominant order | Wintuan | Penutian | 4 cases |
| Yaqui | SOV | Cahita | Uto-Aztecan | 2 cases |
| Yawelmani | No dominant order | Yokuts | Penutian | 6-7 cases |
| Yidiny | SOV | Northern Pama-Nyungan | Pama-Nyungan | 8-9 cases |
| Yoruba | SVO | Defoid | Niger-Congo | No morphological case-marking |

| | | | | |
|---|---|---|---|---|
| Yukaghir (Kolyma) | SOV | Yukaghir | Yukaghir | 8-9 cases |
| Yup'ik (Central) | No dominant order | Eskimo | Eskimo-Aleut | 6-7 cases |
| Zoque (Copainalá) | VOS | Mixe-Zoque | Mixe-Zoque | 2 cases |
| Zulu | SVO | Bantoid | Niger-Congo | No morphological case-marking |
| Zuni | SOV | Zuni | Zuni | No morphological case-marking |

## 9.3 Declaration of Authorship

EBERHARD KARLS
UNIVERSITÄT
TÜBINGEN

Philosophische
Fakultät

Name: Haion

Vorname: Fabio

Matrikel-Nu [redacted]

Adresse: [redacted]

**Hiermit versichere ich, die Arbeit mit dem Titel:**

Do certain word orders attract case marking?

im Rahmen der Lehrveranstaltung Bachelorarbeit

im Sommer-~~/Winter~~semester 2018 bei Dr. Christian Bentz

**selbständig und nur mit den in der Arbeit angegebenen Hilfsmitteln verfasst zu haben.**

Mir ist bekannt, dass ich alle schriftlichen Arbeiten, die ich im Verlauf meines Studiums als Studien- oder Prüfungsleistung einreiche, selbständig verfassen muss. Zitate sowie der Gebrauch von fremden Quellen und Hilfsmitteln müssen nach den Regeln wissenschaftlicher Dokumentation von mir eindeutig gekennzeichnet werden. Ich darf fremde Texte oder Textpassagen (auch aus dem Internet) nicht als meine eigenen ausgeben.

Ein Verstoß gegen diese Grundregeln wissenschaftlichen Arbeitens gilt als Täuschungs- bzw. Betrugsversuch und zieht entsprechende Konsequenzen nach sich. In jedem Fall wird die Leistung mit **„nicht ausreichend" (5,0)** bewertet. In schwerwiegenden Fällen kann der Prüfungsausschuss den Kandidaten/die Kandidatin von der Erbringung weiterer Prüfungsleistungen ausschließen; vgl. hierzu die Prüfungsordnungen für die Bachelor-, Master-, Lehramts- bzw. Magisterstudiengänge.

Datum: 08.07.2018 Unterschrift: Fabio Haion