Progressives in present and past

Ghazaleh Vafaeian
Stockholm University

This study investigates the relationship between progressive patterns and present and past time reference. First, it looks at the shared distribution of more than 90 progressives in two parallel corpora and discusses the characteristics of these contexts. It is shown that while progressives are used for dramatic and topical events in the present, they are typically used as backgrounding, supportive material in the past. Second, it is shown that progressives generally have more occurrences in contexts with present time reference than past, this is especially true for progressives with many uses, i.e. more grammaticalized progressives. And third, a number of progressives temporally restricted are presented. Two historical explanations for these restrictions are provided, both of which result from the higher frequency of present uses over past.

Keywords: progressive aspect, parallel corpora, present time reference, past time reference, grammaticalization, gram type

1. Introduction

Within typological studies, grammatical patterns are often discussed and compared with regard to their functions.[1] The data is traditionally based on grammatical descriptions of different sorts which may list different functions and give a few examples. Although descriptions often capture the main functions and at times can be very thorough and even include corpus data, it can be difficult for a typologist to really know whether two patterns described by the same name or defined in the same way are used similarly or not. This is problematic when dealing with main and peripheral uses but also when dealing with non-obligatory patterns which are complementary to others. In the latter case, even if two patterns have the same main use, their frequency of use, which is a reflection of their degree of grammaticalization, can differ greatly. Naturally, it is not reasonable to

---

1. The findings in this paper are partly based on the author's dissertation (Vafaeian 2018).

demand the level of detail needed for many typological studies from a general reference grammar. Corpus studies of individual languages can be more advantageous in this regard since they have the possibility to explore the whole range of uses of a certain pattern and investigate how frequencies of uses differ. But again, it can be difficult to compare the results from one language with another, even if both have large corpora, since every corpus has its own limitations and biases. In addition to that, most of the languages of the world lack large corpora.

One recent development within the field which seeks to forego these problems, is the use of parallel texts. This approach is not without its problems either but certainly has several advantages. One of them is that limitations or biases of the corpus are often applicable to all languages included. This means that we can discuss and compare patterns within one confined space, the targeted text. What is more, parallel texts enable us to discuss and analyze patterns with regard to their distributions, meaning that we can actually see whether two patterns referred to by the same name or similar definition also share their distribution and to what extent they differ. For example, we can compare total frequencies of patterns, which are highly relevant for their level of grammaticalization, compare frequencies of different uses and also determine prototypical uses for the whole group.

This study explores the distribution of more than 90 progressive patterns in two parallel corpora. The focus of this study is the relationship between progressive patterns and present and past time reference. This is done by comparing prototypical contexts, i.e. those places in the corpus in which most of the patterns occur, in both present and past, comparing the proportion of present and past uses among these grams as well as presenting and discussing grams with temporal restrictions. The investigation involves two collections of texts, the Parallel Bible Corpus and the partially parallel TED (Technology, Entertainment, Design) corpus, with focus on the former. The discussion will, however, not always be restricted to the languages of these two corpora.

Progressives are patterns that are primarily used for ongoing events at reference time (Dahl 1985; Bybee et al. 1994). Examples of such are the English COP + GER pattern, the Spanish *estar* + GER pattern and the Indonesian (Austronesian) *sedang*. I will often use the term **gram** when referring to these constructions. A **gram** is a grammatical item with a specific form and a specific meaning and/ or function (Bybee & Dahl 1989). It can be represented by a single morpheme or a syntactic construction in a language. Grams with similar functions that are found cross-linguistically constitute **gram types**, this is roughly the same as cross-linguistic categories. The progressives just listed can be said to be part of the **progressive gram type**. The grams subsumed under a gram type have overlapping distributions, usually including a set of contexts shared by all or almost all of

them – referred to as the **prototypical contexts** of the gram type. A gram type may however also include sub-types which can have other central or peripheral uses. Gram types and sub-types do not, however, constitute discrete entities. They may be overlapping or be at different stages of a diachronic development (Dahl 2020). The present study will provide evidence for differences between present progressive and past progressive uses and some explanations for these differences. In doing so, the paper provides new insights into how we conceptualize time. Also, differences between less mature and more mature[2] progressive grams, as well as progressives restricted temporally are discussed, problematizing the concept of the progressive gram type.

The article is organized in the following way: Section 2 presents the two corpora and the methods used in the study as well as discusses some problems relevant for the results; Section 3 presents the shared distributions of the grams in the two corpora, i.e. the lines in the two corpora where most grams occur and discusses what distinguishes these contexts; Section 4 presents temporal asymmetries found among the progressives and offers some explanations for these findings; and Section 5 concludes the paper. Supplementary information regarding the sample used in the Parallel Bible Corpus is listed in Section 6.

2.   Methodology

2.1   The two corpora

The present study have made use of two corpora: the Parallel Bible Corpus (henceforth PBC) which contains a collection of New Testament translations and the partially parallel TED corpus which contains a collection of subtitle texts in a number of languages. These two corpora differ from one another in a number of ways and therefore complement each other: while the PBC is fully parallel and includes more than 1,200 languages, many of which are 'non-standard', i.e. languages with no official status, no written tradition and few speakers, the TED corpus is only partially parallel and includes around 100 languages, most of which are 'standard' languages. There is an overrepresentation of Indo-European languages in the TED corpus but not in the PBC. On the other hand, the TED corpus is bigger than the PBC as it contains several million words for some languages while translations of the New Testament typically contain about 200,000 words. Also, the texts in the TED corpus are based on oral sources, presentations, which can be

---

2. Mature patterns are defined as those presupposing a non-trivial prehistory (Dahl 2004: 2).

assumed to include less conventions typical of written languages. This will affect progressives that are restricted to or more common in spoken language.

Both corpora are developed and maintained at the Department of linguistics, Stockholm University. Due to time and resource limitations it is not possible to thoroughly control the quality of the translations. In some cases, systematic anomalies in the translations have been noted, in which case those texts have been excluded from the corpora. Since conclusions drawn in this paper based on data from the two corpora are only drawn at group level, occasional errors and anomalies will not affect the general points. In what follows, more information about the two corpora is provided.

2.1.1   The Parallel Bible Corpus

The PBC is a massively parallel corpus consisting of translations of the New Testament. It is automatically annotated for part of speech and aligned at word level using algorithms developed by Östling (2015). The version of the PBC used for this study contains 1,267 translations from 1,107 languages, meaning that some languages are represented by more than one translation. The New Testament is divided into verses which often consist of several clauses. Since it may be useful to divide the verses into smaller units, algorithms were developed by Östen Dahl which divide the verses into what will be referred to as **segments**. In the examples in the paper, segments are given in italics. The aim of this division is to single out verb phrases. For English, the algorithm ensures that a segment always contains a main verb or an infinitive. In practice, a segment is often a verb phrase or a clause. When referring to the contexts in which the progressives occur in the PBC, I will use the term **segment**. In order to extract those segments that correspond to imitations of spoken language, segments that are part of quotes have been collected. These segments are referred to as **direct speech** and will be relevant in Section 4.

The annotation and segmentation of verses does not always work in optimal ways and the amount of errors vary between the languages of the corpus. These errors may result in that certain progressive grams are not included in the study (more about the identification of progressives in Section 2.2.1). However, the segmentation helps identify verb phrases which is crucial for the study since we want to know in which verb phrase of the verse the grams have been used. As will become evident, this study is mainly concerned with the shared uses of grams, meaning that smaller losses will not affect the conclusions drawn in a notable way.

Another problem with using the PBC is that the language of some translations of the New Testament is archaic and written sources may differ in their conventions from spoken language, for instance with regard to constraints on the use of progressives. This is however mainly true for 'standard' languages and applies to a lesser extent to smaller, non-official languages without written traditions. There

may also be biases towards the usually European language from which the texts are translated (see e.g. de Vries 2007). That being said, there are also many advantages to using the PBC, one of them being that it contains a large amount of the world's languages, many of which are non-standard. Some languages of the corpus lack grammatical descriptions, which would normally exclude them from typological research.

2.1.2   The TED corpus

The TED corpus is a collection of subtitle texts compiled by Östen Dahl from approximately 1,900 TED and TEDx (local TED-like) talks.[3] The talks are held in front of a live audience and given most often in English. The corpus contains subtitles to these talks from 99 languages, is not annotated and only partially parallel since each talk does not have subtitles in all languages. For example, the English collection of subtitles has 3,639,000 words, while Persian has 2,579,000 words and Indonesian 1,759,000 words. The line numbers in the corpus refer to the original subtitles and correspond to 2–3 seconds of speech. All languages are linked to one another by these lines. When referring to the contexts in the TED corpus in which the progressive grams occur I will use the term **line**.

2.2   The two samples

In this section the processes of identifying progressive grams in the two corpora are presented. These processes have yielded two samples, one of 89 progressive grams from the PBC and one of 14 progressive grams from the TED corpus. Some grams occur in both samples, namely English COP + GER, Indonesian *sedang*, Portuguese *estar* + GER, Spanish *estar* + GER and Vietnamese *đang*.

2.2.1   Identification of progressive grams in the PBC

The method used for the identification of progressive grams in the PBC is adopted from Dahl & Wälchli (2016) and is thoroughly explained in Vafaeian (2018). Here, a more concentrated explanation of the procedure is presented.

   The process of finding progressive patterns can be divided into two parts, one **collection** step and one step involving the **trimming** of the sample. Since the sample will be trimmed down at the second step, the first step can be allowed to be rather inclusive.

   Some grams have been **collected** through automatic methods while others are included since they are described as progressives in the literature. The auto-

---

3. All presentations and subtitles (referred to as "transcripts" on the site) are available at www .ted.com/talks (last access 21 October 2022).

matic method uses **seed grams**, a method adopted from Dahl & Wälchli (2016) and used in studies such as e.g. Wälchli (2019) and Dahl (2022). As a first step, the English progressive was used as a seed gram. This involves the creation of a list of all the segments in which the English progressive occurs in the PBC. This list is then used by an algorithm which searches for any string of letters in the PBC that has a similar distribution as the English progressive, i.e. occurs roughly in the same list of segments as the English progressive. The English progressive is a good candidate for a seed gram since it is close to being obligatory for ongoing events which entails that it has a high frequency in the corpus. At this point, we are interested in capturing as many patterns as possible and we are not yet concerned with erroneously captured patterns (i.e. patterns that we would not want to call progressives). The search using English as a seed gram results in a list of strings which contains words or part of words from different languages. Since we are dealing with automatic methods, this first list contains many errors. For example, the search may only have captured part of a pattern or just one tense form of a marker that also has other tense forms. This list is then corrected manually using grammatical descriptions for those languages that have such. Since the corpus is annotated for part of speech, the manual correction can contain specifications which require that a certain word occurs with a certain form of a verb. If a word is conjugated, all those forms can also be included. This was e.g. already done for the English pattern which contains all forms of the copula in the present and past occurring together with the gerund form of the verb. Once this is done, the new list with corrected patterns are run as seed grams. This means that it is no longer the English progressive that defines the distribution of progressives.

The next step involves the **trimming** of the sample. At this point, we want to ensure that we have captured patterns with similar distributions. This is done through the statistical criteria for relevance **recall** and **precision**. Recall is defined in information retrieval as the fraction of the total amount of relevant instances that were actually retrieved. In our case, it measures the extent to which a pattern occurs in the most typical segments for progressives. This is done by first creating a list of the top 20 segments for all the captured grams and then check for each gram how often they occur in this list. Such a list constitutes the prototypical uses of the progressive gram type. Grams with a low recall value are excluded from the sample.

Precision is generally defined in information retrieval as the fraction of relevant instances among the retrieved instances. In our case, precision is a measure of the extent to which the uses of one gram are included in the shared distribution of all the member grams of the gram type. The precision of one gram was measured against a set of segments which represented 85% of all the occurrences of the members of the sample, excluding a set of segments at the lower end of the

frequency scale which together made up 15% of all occurrences. So, while the measurement for recall involves a top list of the 20 most frequent segments shared by the grams, the measurement for precision involves the whole top list of segments minus the lowest 15%. A low precision value is an indication of many uses that are not typical of the targeted gram type. For example, an imperfective gram could occur in the same segments as a progressive and have a high recall value but will also have many additional uses and will therefore have a low precision value. Grams with a low precision value were also excluded from the sample.

In order to refine the sample, the process of collection and trimming was done several times. This means that once the sample was trimmed down, it was used as a seed sample again, collecting new grams for a new sample, which was trimmed down and used as a seed sample again and so on. Also, due to the difference in frequency, the grams were divided into three clusters obtained by the clustering algorithm PAM (Partitioning Around Medoids).[4] Recall and precision were calculated for the whole gram set, as well as for the three subsets (clusters). At the final trimming, only grams that had a recall value higher than 65% and a precision value higher than 50% **or** a precision value higher than 80% at cluster level were included.[5] The clusters roughly correspond to (a) grams with high frequencies (i.e. with expanded uses towards the imperfective) and progressives restricted to the present tense, (b) a group of mainly Austronesian grams with many past uses (discussed in Section 4.2.1) and (c) grams that have no temporal preference. The process of collecting and trimming resulted in a sample of 89 progressive grams, all of which have a precision value above 62% for the whole gram set, meaning that they share at least 62% of their segments in the corpus with the other grams (the gram with the lowest precision value, 62%, is Zemba (Niger-Congo)). Most grams, however, have a much higher precision value. The 89 grams and their recall and precision values are listed in the supplementary material in Section 6. This study will not present these clusters separately but discuss the 89 grams together. For a more detailed information on the clustering process and the characteristics of the clusters, the reader is referred to Vafaeian (2018).

The sample of 89 grams will be referred to as the PBC sample. One language, Vietnamese (Austro-Asiatic), is represented by two grams, *đương* and *đang*, which occur in two different translations, the former from 1934 and the latter in the Easy-to-read Version from 2011. Both markers occur in both translations but the translation from 1934 mainly uses *đương* while the translation from 2011 mainly uses *đang*.

---

4. The algorithm is pam in the R cluster package. (https://cran.r-project.org/web/packages /cluster/cluster.pdf, last access 21 October 2022).

5. I have made two exceptions for Sekpele and Makaa as they instead fulfil the requirements of recall and precision at the gram set level.

Given that the PBC includes written texts, we expect the grams in the PBC sample to be highly grammaticalized. Less grammaticalized progressives, like for example the Swedish (Indo-European) *håller på att/och*, the French (Indo-European) *être + en train de + INF* and the Persian (Indo-European) *dar hāl=e + INF* or *dāštan + IPFV* do not have a sufficient number of occurrences to be included in the sample. They have more occurrences in the TED corpus and meet the criteria of inclusion and are therefore present in that sample. At least for Swedish and the *dāštan + IPFV* in Persian, the low occurrences of progressives in the PBC but not in the TED corpus can be (perhaps partly) explained by these grams occurring more often in spoken language than in written sources (see Blensenius 2015: 33, 296; Windfuhr & Perry 2009). In addition to that, the TED corpus contains more words per language and some grams may therefore be more frequent in that corpus.

2.2.2   Methodological issues regarding automatic extraction

The method of automatic extraction described above has the advantage of using distribution as a criterion for the identification and comparison of progressive grams. The method is also able to identify patterns which have not yet been identified as progressives in the literature as well as progressives belonging to languages for which there exist no, only inadequate, or unobtainable grammatical descriptions or descriptions written in languages not known to the typologist.

That being said, there are certain issues that need to be mentioned regarding the automatic extraction of progressive grams. In its present form, the algorithm which identifies grams is restricted in several ways, it can for example not capture irregular or suppletive paradigms nor periphrastic forms. If a progressive marker is inflected, the captured search string may just be the most frequent form or part of the progressive marker. For example, one of the first searches yielded the form *est* for Spanish (Indo-European). The search string for Spanish was then manually adjusted to include all forms of *estar* followed by the gerund. In similar ways, I have adjusted the search string for all languages for which grammatical descriptions are available. For a marker used both in progressive and in locative constructions, the search conditions have to indicate that the marker is combined with a verb in order to exclude the locative uses. Since the corpus is annotated for part of speech, such criteria can be specified. In some cases, such specifications did not improve the results, a matter which is probably an outcome of errors in the automatic annotation of the translation for that language. Some languages with locative markers used for ongoing events were excluded from the sample due to this.

For languages lacking grammatical descriptions these types of manual adjustments could not be performed. They were simply included if their distribution met the recall and precision criteria. This means that their progressive forms may

contain errors. These grams are in minority and will not affect the general conclusions drawn in this paper.

2.2.3   Identification of progressive grams in the TED corpus

The identification of progressive grams in the TED corpus also involved a **collection** step and a **trimming** step, although the process was much more straightforward and less time consuming since we are dealing with a smaller number of languages with often good and easily obtainable descriptions. The **collection** of progressive grams was done through searches in these grammatical descriptions. For the purposes of this study, only languages with a relatively high number of words were searched, the language with the lowest number of words in the TED corpus which is included in the sample is Finnish with 332,490 words.

The **trimming** process employed here also involved the statistical criteria of recall and precision. The grams in the sample all have a recall value of 100% for the top 20 lines and a precision value above 60%. Patterns with low recall or precision values, or both, were excluded from the sample. For example, the Turkish *-Iyor*, which is included in the PBC, is not included in the TED corpus since its distribution suggests that it has too many uses not shared by the other grams. The gram *-Iyor* may occur with stative predicates and has thus extended its uses towards the imperfective (see e.g. Kornfilt 1997: 357). This is also reflected in its relative frequency which is 21.1 per 1,000 words in the TED corpus, three times the English progressive as seen in Table 1. Thus, in a written and more formal contexts like the PBC, *-Iyor* has a more limited distribution similar to progressive grams while in the TED corpus, a more modern and spoken-like source, it has a distribution more similar to an imperfective. In the PBC then, the original uses of *-Iyor* are preserved, while the marker has extended uses in the TED corpus. This can be explained by the fact that patterns expand to new functions in spoken language first and that the texts in the TED corpus are more close to spoken language than the texts in the PBC. In fact, as already mentioned, grams such as the Persian *dāštan* + IPFV, the Swedish *håller* + *på* + *att*/*och* and the French *être* + *en train de* included in the TED corpus, are less grammaticalized and therefore do not occur, or occur very sparsely, in the PBC.

The process of collection and trimming resulted in a sample of 14 progressive grams from 12 languages. All languages, their total occurrences and their frequencies are displayed in Table 1. As can be seen, the English progressive is the most frequently used progressive in the TED sample. Persian and Portuguese are represented by two grams each. For Persian, both the *dar hāl=e* + INF and the *dāštan* + IPFV progressive patterns are included. The former is more formal and often used in written sources while the latter is most frequently used in colloquial language. Despite being a formal progressive, the *dar hāl=e* + INF pattern did not have many

occurrences in the PBC. In Portuguese, both *estar + a +* INF and *estar +* GER are included. As can be seen, the former is much more frequent than the latter. For these two languages, the two grams are in complementary distribution, i.e. they never occur together in one line.

Table 1. The progressive grams of the TED corpus[*]

LANGUAGE	LANGUAGE FAMILY	PROGRESSIVE GRAM	TOTAL OCC. OF GRAM	FREQUENCY OF GRAM PER 1000 WORDS
English	Indo-European	*be +* GER	24,100	6.6
Finnish	Uralic	*olla +* NMLZ	378	0.8
French	Indo-European	*être + en train de +* INF	898	0.3
Indonesian	Austronesian	*sedang*	1,402	0.8
Italian	Indo-European	*stare +* GER	9,380	2.7
Mandarin Chinese	Sino-Tibetan	*zhèngzài*	2,399	0.7
Persian	Indo-European	*dar hāl=e +* INF	2,208	0.9
		dāštan + IPFV	2,890	1.1
		Total	5,098	2.0
Portuguese	Indo-European	*estar + a +* INF	7,768	3.9
		estar + GER	256	0.1
		Total	8,024	4.0
Spanish	Indo-European	*estar +* GER	10,008	3.0
Swedish	Indo-European	*håller + på + att/och*	156	0.1
Thai	Tai-Kadai	*kamlang*	4,232	2.5
Vietnamese	Austroasiatic	*đang*	13,770	4.8

[*] Table from Vafaeian (2018).

There are several methodological issues regarding the TED corpus as well. First, since the corpus is not completely parallel we do not know if a progressive would have been used in lines lacking sub-texts for a particular language. Also, progressives that would otherwise have been used may have been omitted for reasons of space only, especially if the progressive is marked periphrastically. In addition, due to time restrictions, it has not been possible to check the translation quality of the sub-texts. We also do not know if a sub-text has been translated directly from a TED-talk or from another sub-text.

3.   Shared distribution of progressive grams

In this section, the shared contexts of the progressive grams, i.e. the segments or lines with the highest incidences of progressive grams, are presented, both for the sample of 89 grams in the PBC and for the 14 grams in the TED corpus. This section will mainly be concerned with the top 50 sentences that are shared among the progressives in both corpora. The analysis of the examples made here relies on the English translation or sub-text and are generalizations for the two samples.

In Example (1), the top five segments in the PBC that contain progressives are shown. The New Testament book, chapter, and verse are indicated for each example, followed by the percentage of progressive grams in the PBC that were used in the example. In (1a) then, Acts 9:11 identifies its place as Acts of the Apostles, Chapter 9, verse 11. 84% of the progressives of the sample were used here. The segments in which the progressives occur are given in italics. This means that it is not only in verse 11 that 84% of the grams of the PBC sample occur, but more specifically in the clause *For behold, he is praying*. Sometimes, additional clauses exceeding the Bible verse are included for context, these are given in square brackets (see e.g. Example (3)). All English examples are quoted from the Lexham English Bible.[6]

(1)   Top five segments, PBC
   a.   Acts 9:11, 84%
         And the Lord said to him, "Get up, go to the street called 'Straight' and in the house of Judas look for a man named Saul from Tarsus. *For behold, he is praying*,
   b.   Luke 3:21, 76%
         Now it happened that when all the people were baptized, Jesus also was baptized, and *while he was praying*, heaven was opened,
   c.   Matt. 17:5, 73%
         While he was still speaking, behold, a bright cloud overshadowed them, and behold, a voice from the cloud said, "This is my beloved Son, with whom I am well pleased. Listen to him!"
   d.   Acts 10:30, 72%
         And Cornelius said, "Four days ago at this hour, *the ninth, I was praying in my house*. And behold, a man in shining clothing stood before me
   e.   Mark 14:43, 71%
         And immediately, *while he was still speaking*, Judas one of the twelve arrived, and with him a crowd with swords and clubs, from the chief priests and the scribes and the elders.

---

6. Copyright 2012 Logos Bible Software. Lexham is a registered trademark of Logos Bible Software.

A top list of the lines in which progressives occur the most in the TED corpus has also been created. These examples are preceded by the name of the presenter and the time of the targeted utterance. The targeted utterance refers to the line in the corpus in which the progressives occurred and is given in italics. Additional lines are given for context. The name and time are followed by the number of grams that used their progressive in this line divided by the number of languages that had a line in the corpus. Recall that the corpus is only partially parallel, meaning that not all languages have all the corpus lines. Also, the first 37 top lines have a percentage of 100, meaning that all languages that had that line in the corpus marked that line with a progressive. Below, (2) presents some of these lines. For example, in (2a), 7 of the 12 languages in the sample had a line in the corpus and all of them contained a progressive, indicated by 7/7. There is no line in the data where all 12 languages had a line containing a progressive.

(2)   Top lines where all languages with a line in the corpus used their progressive, TED corpus

   a.   John Underkoffler, 08:24 (7/7)
        So, from there to the world of real work. Here's a logistics system, *a small piece of one that we're currently building*. There're a lot of elements.

   b.   Suzanne Lee, 06:06 (10/10)
        Bruno Giussani: Suzanne, just a curiosity, what you're wearing is not random.

        Suzanne Lee: No.

        Bruno Giussani: This is one of the jackets you grew?

        Suzanne Lee: Yes, it is. It's probably – part of the project's still in process because this one *is actually biodegrading in front of your eyes*. It's absorbing my sweat, and it's feeding on it.

   c.   Matt Killingsworth, 08:33 (7/7)
        Turns out, they wander a lot. In fact, really a lot. Forty-seven percent of the time, people are thinking *about something other than what they're currently doing*. How does that depend on what people are doing?

   d.   Janna Levin, 13:19 (9/9)
        In this Hubble image, we see two galaxies. They look like they're frozen in some embrace. And each one probably harbors a super-massive black hole at its core. But they're not frozen; *they're actually merging*. These two black holes are colliding,

   e.   Don Tapscott, 00:45 (8/8)
        The Internet is becoming a giant global computer, and every time you go on it, you upload a video, you do a Google search, you remix something, you're programming this big global computer that we all share. *Humanity is building a machine* and this enables us to collaborate in new ways.

These two top lists represent the shared distribution of the progressive grams. This means that cross-linguistically, they are the typical contexts for progressives. Looking at the top lists from these two corpora, we can see that, as expected, the examples all refer to ongoing (dynamic) events. We can also note that when events can be viewed as telic, (2a, b, d and e), the end points of these events are depicted as potential rather than as realized. The study then confirms that the prototypical uses of progressives refer to ongoing events, where ongoing is understood as involving dynamic events, and that any telic end points are potentialized (see e.g. Bertinetto & Delfitto 2000; Johanson 2000).

It is noted here that in Example (1), four of the five lines in the PBC are in the past tense in the English translation. In fact, the majority of segments in the top 50 in the PBC are in the past tense in English, in contrast to the top 50 lines of the TED corpus as shown in Table 2. The assignment of temporal reference for these contexts has been done manually based on the English translation.

Table 2. Division into English present and past of the top 50 contexts

	PRS	PST	TOTAL
PBC	12	38	50
TED corpus	46	4	50

The numbers from the PBC are, however, misleading. They are skewed as a consequence of a group of Austronesian grams which almost exclusively occur in *while*-clauses. Since the grams in the rest of the sample also occur in *while*-clauses these end up in the top of the list. This shows that although these lists may provide us with important data regarding shared uses cross-linguistically, they need to be used with caution. The next sections will look at contexts with present and past time reference separately, demonstrating important differences between them. I will return to the Austronesian languages in Section 4.1.3.

3.1   Contexts with present time reference

In this section, the top examples with present time reference in the PBC and the TED corpus are presented and analyzed. The assignment of temporal reference is done manually both in the PBC and in the TED corpus and is based on tense choice in English. The choice between present and past tenses in English is used as a proxy for the distinction between present and past time reference. Example (3) shows the five top present tense phrases from the PBC in which the languages of the sample used their progressive grams the most. Example (1a) shown above is repeated here as (3a).

Notably, all five segments occur within quotation marks in the English translation. This is a typical feature of uses of progressives in the present tense in the PBC and is reflected in the rest of the top list where 8 out of 12 lines in the English present tense occur in direct speech. This suggests that present progressives are often associated with spoken language even in a written source such as the New Testament.

(3)   Top examples with present time reference in the PBC
   a.   Acts 9:11 84%
         And the Lord said to him, "Get up, go to the street called 'Straight' and in the house of Judas look for a man named Saul from Tarsus. *For behold, he is praying*,"
   b.   John 9:37, 69%
         [Jesus heard that they had thrown him out, and finding him, he said, "Do you believe in the Son of Man?" He answered and said, "And who is he, sir, that I may believe in him?"] Jesus said to him, "You have both seen him, and he is the one *who is speaking with you*."
   c.   Mark 1:37, 64%
         [And getting up early in the morning while it was very dark, he departed and went to a deserted place, and there he was praying. And Simon and those who were with him searched diligently for him.] And they found him and said to him, "*Everyone is looking for you*!"
   d.   Mark 10:33, 62%
         "*Behold, we are going up to Jerusalem*, and the Son of Man will be handed over to the chief priests and the scribes, and they will condemn him to death and will hand him over to the Gentiles."
   e.   Mark 15:35, 60%
         And some of the bystanders, when they heard it, said, "*Behold, he is summoning Elijah*!"

It can also be noted that the word *behold*, or *look* in other translations, is used in these contexts and also many others in the top 50 segments. In this data then, progressive grams are used in contexts that aim at turning the attention of the listener towards an ongoing event present in the participants' shared field of vision. In this way, progressives can be used in marking that an urgent or important event is happening right now which the listener needs to see or be aware of. Examples (3b), and (3c) are not attention seeking but they are similar to such contexts in that they are somewhat dramatic.

Example (4) shows a selection of top lines in the TED corpus which are in the English present tense. In Example (4), all languages that had a line in the corpus used their progressive.

(4)   Segments in the present, TED corpus

    a.   Paul Root Wolpe, 04:13 (9/9)

       But something much, much more powerful *is happening now*. These are normal mammalian cells genetically engineered with a bioluminescent gene taken out of deep-sea jellyfish.

    b.   Jan Chipchase, 14:00 (7/7)

       And I know TED is about big ideas, *but actually, the benchmark for a big idea is changing*. If you want a big idea, you need to embrace everyone on the planet, that's the first thing.

    c.   Janna Levin, 13:19 (9/9)

       In this Hubble image, we see two galaxies. They look like they're frozen in some embrace. And each one probably harbors a super-massive black hole at its core. But they're not frozen; *they're actually merging*.

    d.   Frederic Kaplan, 03:28 (8/8)

       We are setting up a 10-year digitization program which has the objective of transforming this immense archive into a giant information system.

    e.   Don Tapscott, 00:15 (8/8)

       Openness. It's a word that denotes opportunity and possibilities. Open-ended, open hearth, open source, open door policy, open bar. (Laughter). *And everywhere the world is opening up*, and it's a good thing.

    f.   Anne-Marie Slaughter, 16:38 (5/5)

       The revolution for human equality can happen. *It is happening*. It will happen.

Similar to the top contexts in the PBC, the uses in the TED corpus often involve contexts that are somewhat urgent, indicated here e.g. by words like *actually* and *now*, or the rhetorical use of tense in (4f). Already Hatcher (1951) discusses the English progressive in terms of the "involvement of the subject" on the verb and notes that the replacement of the simple form with the progressive adds "exceptional emphasis" to the relevant activity. Similar conclusions having to do with the busyness of the subject in the event have been drawn for Persian (Vafaeian 2018). Comrie (1976:37) also noted that some uses of the English progressive had an **emotive effect**. More recent research mainly concerned with Indo-European languages such as English, French, German, Dutch and Persian have shown similar effects (De Wit et al. 2013; De Wit & Brisard 2014; Anthonissen et al. 2016; Vafaeian 2018). In fact, many of the top lines in the present tense include a word such as *actually, currently, now* or emphatic modifications such as *rapidly* and *really*. These emotively charged contexts then seem to increase the likelihood of the use of a progressive.

The top lines from the TED corpus, as opposed to the top segments in the PBC, do not contain many attention-seeking uses. This is related to the presenta-

tion setting and the type of topics that are being discussed. There are simply not many ongoing events happening around the presenter to which s/he would want the audience to turn. When showing something on the prompter, the presenters instead use utterances such as *And the thing I want you to look at here...*[7] or *Let's look at the squares...*,[8] rather than *look, X is* v-*ing* which would be the correspondence to the *behold* examples of the PBC.

Previous research on progressives often distinguishes between punctual reference points (e.g. *right now, at 5 o clock*) and durative reference points (e.g. *from 2 to 5* or *during the whole time of class*). Bertinetto et al. (2000) show that progressives may be sensitive to this distinction: there are progressives such as the Italian *stare* + GER, the French *être en train de* + INF and the Albanian *po* which cannot occur in contexts with durative reference points, and there are also progressives such as patterns from various Romance languages formed with motion verbs, e.g. the *andare/venire* + GER in Italian, that occur with durative reference points only (Bertinetto 2000). Many progressives such as for example the English progressive, the Persian *dāštan* progressive and the Swedish *håller på att/och* progressive can occur in both types of context (see e.g. Vafaeian 2018). The examples above from the PBC and TED corpora all have a punctual reference point as opposed to a durative one, with the exception of a few uncertain cases. This is also reflected in the rest of the top 50 segments and lines. We can then conclude that typologically, punctual reference points are more typical for progressives with present time reference than durative ones, even though progressives do not need to be restricted in this regard. Also, apart from the motion verb progressives in Romance languages, I have not come across other progressives which prefer or are restricted to durative reference points. The impression is then that such patterns are typologically rare.

We now turn to top lists in the PBC and the TED corpus involving contexts translated to the English past tense.

3.2   Contexts with past time reference

The top contexts which are in the past tense in the English translation and include progressive grams differ from those in the present tense to some extent. These contexts are typically narrations in which the progressive applies to an event that is the background of another, usually telic, event. Past events often involve narrations, which in turn include information that can be classified into foregrounded and backgrounded, where the former constitutes "the language of the actual story

---

7. Anil Ananthaswamy, 12:55.

8. Arthur Benjamin, 02:04.

line" and the latter "the language of supportive material which does not itself narrate the main event" (Hopper 1979: 213).

In Examples (1b–e) we have already seen the top four past segments in the PBC. They are all background contexts and almost all of them are 'while'-clauses. In fact, the majority of past segments in the top 50 are of this sort. It was noted above that the reason why past uses are at top of the list in Example (1) is due to the preference for while-clauses in the Austronesian languages of the sample. However, even if we were to remove the Austronesian languages from the sample, the top list in the past would still be dominated by 'while'-clauses, showing that typologically, backgrounding clauses are prototypical contexts for progressives in the past.

This does not mean that past progressives are restricted to backgrounding uses and there are also sentences that can be analyzed as foregrounded with past time reference, shown in Example (5). It should be noted that the percentage of progressives that occur in these past foregrounded contexts is lower (60, 58 and 57% respectively) than those occurring in past backgrounded contexts shown in Example (1) (76, 73, 72 and 71% respectively). This means that these contexts are not shared by as many progressive grams, which in turn means that they are less typical of progressives in the past.

(5)   Foregrounded contexts with past time reference, PBC
   a.   John 7:28, 60%[9]
        Then Jesus cried out in the temple courts, teaching and saying, "You both
        know me and you know where I am from! And I have not come from
        myself, but the one who sent me is true, whom you do not know.
   b.   Mark 9:4, 58%
        [And he was transfigured before them, and his clothing became radiant –
        extremely white, like no cloth refiner on earth can make so white.] And
        Elijah appeared to them together with Moses, and *they were talking with
        Jesus*. [And Peter answered and said to Jesus, "Rabbi, it is good that we are
        here!]
   c.   Luke 5:2, 57%
        [Now it happened that while the crowd was pressing around him and
        hearing the word of God, he was standing beside the lake of Gennesaret,]
        and he saw two boats there beside the lake, but the fishermen had gotten
        out of them and *were washing their nets*. [And he got into one of the boats,
        which was Simon's, and asked him to put out from the land a little.]

9. Notably, this segment does not include a progressive in the English translation but 60% of the languages of the sample did use their progressive here. It is unclear which verb in the segment the progressive applies to. For example, in the Spanish translation (Palabra de Dios para

The data in the TED corpus is similar to the PBC data in that the top lines in the English past tense are often backgrounding contexts. Example (6) provides some examples. Since past uses are rare among the top 50 examples, some examples further down the list are also given here.

(6) Backgrounding contexts in the past, TED corpus
   a. Carl Safina, 15:40 (5/5)
       At an aquarium in South Africa was a little baby bottle-nosed dolphin named Dolly. *She was nursing, and one day a keeper took a cigarette break* and he was looking into the window into their pool, smoking.
   b. Sean Gourley, 00:34 (8/9)
       So as a naive New Zealander I thought, well I'll go to the Pentagon. Can you get me some information? (Laughter) No. So I had to think a little harder. *And I was watching the news one night in Oxford.* And I looked down at the chattering heads on my channel of choice. And I saw that there was information there.
   c. Jennifer Granholm, 07:27 (7/8)
       …and I was standing in the back of the room during one of the demonstrations and standing next to one of the Chinese officials, *and we were watching, and he says*, "So, Gov, when do you think the U.S. is going to get national energy policy?"
   d. Michael Pollan, 00:55 (6/7)
       Like a lot of my ideas, like a lot of the tools I use, I found it in the garden; I'm a very devoted gardener. *And there was a day about seven years ago: I was planting potatoes*, it was the first week of May – this is New England, when the apple trees are just vibrating with bloom; they're just white clouds above. […] And the question I asked myself that afternoon in the garden
   e. Jehane Noujaim, 06:36 (6/7)
       […] because before the war started, there was kind of this media war that was going on. *And I was watching television in New York*, and there seemed to be just one point of view that was coming across […]

In the upcoming section, the characteristics of the uses presented so far are discussed.

Todos), the segment is *Por eso Jesús, que estaba enseñando en el área del templo, dijo con voz fuerte*, where the progressive is used with *teaching*. In either case, the example can be seen as foregrounded.

3.3   Features increasing the likelihood of the use of a progressive

Section 3 has shown that the uses of progressive grams differ in contexts with present and past time reference. In the present, progressives are sometimes used in contexts with more emotive components and for utterances in which the speaker wishes to turn the attention of the listener towards an ongoing event. In the past, progressives are used in backgrounding contexts as part of narrations. This is a direct consequence of how we talk of events with present versus past time reference: events in the present are more often topical and urgent as they are ongoing at the speech moment while events in the past are less often so since they have already taken place. In addition, in the past, utterances are often part of narrations which includes foregrounded and backgrounded events. It is the foregrounded events that push the story line forward while the backgrounding events provide supportive material. The reason why progressives are more often found in the latter has to do with their meaning of incompleteness. Previous research has noted that imperfective patterns are non-propulsive and incapable of advancing a narration plot (Hopper 1979: 239; Johanson 2000: 76; Dahl 2013: 65). Thus, the incomplete, atelic, event of e.g. *She was working* is pausing the narration, bringing it to a plateau as it were, needing a telic event such as e.g. *when the lights went off* to advance the plot. *She is working* on the other hand does not have this sense of pausing the narration since an event taking place at the speech moment is not expected to be completed. It is this temporal asymmetry that is reflected in the results presented in this section. The difference in uses with present and past time reference lead to the somewhat peculiar conclusion that while progressives are used for dramatic and topical events in the present, they are typically used as backgrounding, supportive material in the past.

Based on the shared distribution of the progressives in the PBC and the TED corpus and on previous research presented above, I suggest that there are several features that increase the likelihood of using a progressive gram in non-backgrounding contexts, especially those with present time reference. I hypothesize that, the more features that are present, the more likely it is that a progressive gram will be used.

(7)   Features increasing the likelihood of the use of progressives in non-backgrounding contexts
   a.   a punctual reference point
   b.   an emotive component
   c.   the involvement or busyness of the subject in the event
   d.   the desire to turn the attention of the addressee towards, or make the addressee aware of, an ongoing event

These features increase the likelihood of the use of a progressive but do not have to be present in all utterances with a progressive. In fact, I hypothesize that these features are related to the level of grammaticalization of a progressive gram in that they are especially relevant for non-obligatory uses, including new contexts that progressives take on in their maturation process. For less grammaticalized progressive grams the requirement of these features is stronger as almost all contexts are non-obligatory while for more mature progressives the requirement is only relevant for newer contexts and not those in which the progressive is the only or most natural choice. In what follows, I will discuss the feature of **emotive component** further as it has attracted some attention in the literature by comparing the two grams with the highest and lowest frequencies in the TED sample, from English and Swedish, respectively.

In the literature, emotive nuances related to progressive uses have been discussed using terms such as 'emotive effect', 'affective', 'subjective', 'modal', 'atypical', 'epistemic contingency', 'extravagance' etc. in mainly Indo-European languages such as English, French, German, Dutch, Swedish, Persian, Icelandic, Western Armenian, Albanian, but also the Niger-Congo language Igbo and a number of Bantu languages. (Einarsson 1949; Comrie 1976; De Wit et al. 2013; De Wit & Patard 2013; De Wit & Brisard 2014; Anthonissen et al. 2016; Vafaeian 2018; De Wit et al. 2020). Although these terms may be defined in different ways, it is my impression that they aim at capturing the same feature or characteristics of certain contexts in which progressive occur. In my opinion these characteristics are related to non-obligatory uses of progressives, which in turn are related to the process of grammaticalization for this gram type. The level of grammaticalization of a progressive gram is directly reflected in the frequency of the gram, e.g. the English progressive has a frequency of 6.6 occurrences per 1,000 words in the TED corpus, while the frequency of Swedish *håller på att/och* is 0.1 per 1,000 words. This difference is due to the level of obligatoriness for ongoing event of the two markers. That is, while both are used for marking ongoing events, the former is often a (close to) obligatory marker for these events while the latter is an option to the simple tenses. In English, the question *What are you doing?* can only be answered by the use of the progressive *I'm working*, not the simple present *I work*, which instead is used for generic meaning. Notably, *I'm working* does not express any emotive nuance but simply marks that the event of working is ongoing. In Swedish, both the progressive *Jag håller på att jobba* 'I am working' and the simple present *Jag jobbar* 'lit. I work' are available as an answer to *Vad gör du?* 'What are you doing?'. The use of the progressive here adds a sense of 'busyness' to the event similar to Hatcher's (1951) "involvement of the subject"; it emphasizes the event of working and/or the occupation of the subject in this event to a greater extent than the simple present, which is a rather neutral statement.

For ongoing events then, the English progressive is often the most natural option and has started to push the present tense towards generic meaning. This will be reflected in the frequency of the progressive and affects the level of pragmatic markedness of the contexts in which the progressive occurs: something that is obligatorily used necessarily loses any emotive nuance. (Such nuances can, however, be expressed by other elements of the clause.) This is a crucial characteristic of the process of grammatical maturation which has been referred to as **rhetorical devaluation** (Dahl 2004: 121). In Swedish, on the other hand, the use of the progressive signals something since there is a choice between the progressive and the simple tenses. In addition, there is a greater cost in the use of the Swedish progressive in terms of production efficiency in that the Swedish example containing the progressive will add four syllables to the utterance while the English progressive example adds one (in spoken language). This means that the use of the Swedish progressive needs to be more motivated than the use of the English progressive. This is also expected since grammatical maturation is accompanied by phonological reduction. The process of rhetorical devaluation should also be expected to happen in the past, as past progressive constructions increase in use. The present study cannot, however, verify such a process and future studies should investigate the grammaticalization process of past uses of progressives in more detail.

Interestingly, using the English progressive in contexts in which it is not obligatory or the default choice will give rise to similar emotive nuances. Habitual contexts and uses with stative verbs are usually not seen as typical for progressive grams and can be viewed as uses that the progressive takes on in its expansion toward the imperfective. The English progressive can have these uses under certain conditions. The pattern is however far from reaching a full blown imperfective stage since it has 'ongoing uses at reference time' as its main use and since these new uses are rather infrequent. Comrie (1976: 37) discusses the example *She's always buying far more vegetables than they can possibly eat* as adding more emphasis to the utterance in comparison to the simple present *She always buys far more vegetables than they can possibly eat*. Combinations of progressives with stative predicates have also been analyzed as having emotive nuances, in e.g. sentences such as *John is being silly* in comparison to the simple *John is silly*. It is noted that while *John is silly* means 'John is a silly person', an utterance such as *John is being silly* means 'John is behaving in a silly way', where the latter has "emotional overtones of irritation" (Anthonissen et al. 2016: 19). With stative predicates, the use of the progressive seems to give rise to a more dynamic reading. Ljung (1980: 40) refers to this property of the English progressive when used with statives as the possibility to "'bring out' the 'acting' interpretation" and distinguishes between those statives that have a feature of 'dynamicness' and those that don't.

In the example then, 'is being silly' is interpreted as 'acting silly'. I believe that this is linked to the restriction to or strong association of the English progressive with dynamic verbs. When this pattern is used with a stative predicate, the stative predicate will be re-interpreted as dynamic or dynamic-like. In both habitual contexts and with stative predicates, the simple present is the neutral and default choice, whereas the use of the progressive signals something, it is a choice which also entails a small cost. Therefore, it will necessarily have a pragmatically marked sense. In my opinion, this is applicable to all instances where the progressive has not become obligatory or the default choice, i.e. where the simple present is also available, for example in the difference between the neutral *I tell you* and the more 'extravagant use' *I am telling you* as discussed in De Wit et al. (2020: 498–499). The more infrequent the use of the progressive is, the stronger will the emotive effect be. For a young progressive like the Swedish one, all contexts are non-obligatory and therefore pragmatically marked while the use of a more mature progressive like the English one has pragmatic nuances in contexts in which it is not yet the default choice. This can be compared to e.g. a simple present or a past imperfective which are obligatorily marked and therefore do not have pragmatic nuances as such. In this sense, the progressive is a somewhat volatile gram type in that its members are at different stages of maturation towards an obligatory and more frequent imperfective.

Already Einarsson (1949: 144–145) noted that verbs that typically would not combine with the progressive *vera að* in Icelandic (Indo-European) can do so in affective contexts such as those expressing surprise, disgust or dissatisfaction. According to him, it is the progressive that adds such nuances to these utterances. In a number of papers, De Wit and her colleagues somewhat similarly argue that uses referred to as e.g. 'extravagant', 'subjective' or 'atypical' arise as a consequence of the semantics or core meaning of the present progressive (De Wit et al. 2013; De Wit & Brisard 2014; Anthonissen et al. 2016; De Wit et al. 2020). The examples given often include habitual contexts, uses with stative predicates, performative contexts and other non-obligatory uses. Although acknowledging that the prototypical uses of progressives are those that mark ongoingness, and that extravagant uses are linked to contexts in which the progressive is interchangeable with the simple tense and not those in which it is obligatory, they still conclude that there is something inherent about the semantics of the progressive gram type that makes it more suitable for usage in extravagance utterances (De Wit et al. 2020). However, if adding 'extravagance' or any other type of emotive component to the core meaning of the progressive, one would need to explain the absence of these nuances in all neutral cases: why and how is the 'extravagance' sense annulled in an example like *I am working* as an answer to *What are you doing?* This is especially problematic since these pragmatically neutral cases may constitute a great

deal of a progressives total uses, as they have been shown to do for English (see e.g. De Wit & Brisard 2014:70). The question of what constitutes the progressives core meaning is obviously complicated as the pattern meaning interplays with other aspect of the utterance, such as the verbal meaning, meanings associated with typical uses of the pattern (like in the example of *John is being silly* above) and pragmatic nuances present in other elements. It may therefore not be possible to entirely disentangle the meaning that the progressive adds to an utterance from meanings added by other elements of the utterance in order to find one comprehensive core meaning activated in all instances of use. In my opinion, what we can say at this point is that the emotive components discussed in the literature are linked to non-default uses, meaning that they are not present in instances where the progressive has become the default choice.

4.   Temporal asymmetries

We have seen that present and past uses of progressives differ in character. But some progressives also occur more often with one time reference than with the other and can even be restricted to present or past uses. This section will present data on these temporal asymmetries and examine the proportion of present and past segments of the 89 progressive grams in the PBC. It is not possible to calculate the proportion of present and past segment for the grams in the TED-corpus since the corpus is only partially parallel. Below, Section 4.1 presents data on the temporal preferences of the grams in the PBC and Section 4.2 discusses grams that are temporally restricted.

    In order to calculate the proportion of present and past uses, the segments of the PBC were divided into past and present depending on the tense marker that is used in that segment in the Esperanto translation. Esperanto was chosen for convenience as it has a regular way of marking tense: all present tense verb forms take an *as* suffix and all past tense verb forms take an *is* suffix. It would not have been possible to calculate the present and past proportions automatically for the whole PBC using a natural language since tense is usually irregularly marked or unmarked. In Section 3, the division of tense was based on English and done manually for the top 50 segments. This would not have been possible to do for all the segments in which progressives occur in the PBC. What is more, natural languages often have idiosyncratic uses of tenses where a tense form may not always be used for the corresponding time reference.

    The occurrences of each of the 89 progressive grams were checked against Esperanto present and past tense, i.e. the sum of the occurrences that matched Esperanto present tense were calculated, on the one hand, and the sum of the

occurrences that matched Esperanto past tense on the other. I will refer to these uses as **present uses** and **past uses** respectively. For each gram, there were also a number of occurrences that did not match present or past. To give an example, in Achinese (Austronesian) *teungoh* is a progressive gram which occurs 59 times in segments that correspond to Esperanto present tense and 168 times in segments that correspond to Esperanto past tense. This gives the proportion of 26% (59/227) present versus 74% (168/227) past. *Teungoh* then occurs most often in contexts with past time reference.

The proportion of present and past verb forms in the whole Esperanto translation is 43% and 57%, meaning that past verb forms are slightly more frequent. But if we look at the total distribution of all progressive grams, 58% of all occurrences of progressive grams in the sample are found in contexts where Esperanto uses present tense, in spite of those contexts being in minority. This means that progressives generally have more uses with present time reference than past in a text that has a majority of past forms. As will be shown in this section, there are grams that have a majority of past uses too, those grams are, however, restricted areally.

In segments identified as direct speech, the proportions of Esperanto present and past tenses are perhaps not surprisingly higher for the present, 70% and 30% respectively. The total proportions of present and past in direct speech segments among the progressive grams are 86% and 14%. Thus, the direct speech segments also show an overall preference for present time reference among the progressive grams in the PBC.

Table 3 shows the proportion of present segments for the 89 grams of the PBC sample. The grams have been classified into five groups depending on their proportion of present segments. As can be seen, the biggest group consists of progressive grams that do not show temporal asymmetry, followed by a group of grams that occur most often in contexts with present time reference and finally grams that occur most often in contexts with past time reference. A few grams have a large majority for either present or past. Looking at direct speech segments only, there is a very strong preference for present time reference, meaning that in segments that imitate spoken language, progressive grams are used with present time reference almost exclusively.

The five groups in Table 3 are presented in different colors in Map 1 where the areal distribution of these grams and their grouping is illustrated. On the map, the darker red color indicates majority of present uses, while the lighter red indicates majority of past uses. As can be seen, the lighter red colors are almost exclusively found in Southeast Asia which indicates that this is an areal feature. As will be discussed in Section 4.1.3, the many past uses of these grams seem to have to do with subordination rather than past time reference.
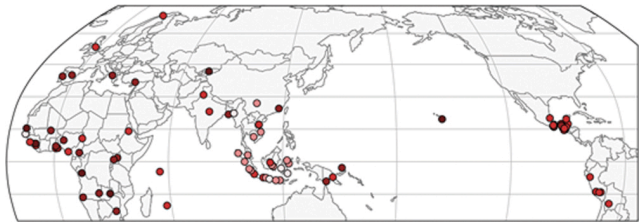
Table 3. Proportion of present segments for the 89 progressive grams[*]

		NO. OF PROGRESSIVE GRAMS	
PRS/(PRS+PST)	GROUPING	WHOLE CORPUS	DIRECT SPEECH
0% ≤ x < 20%	large majority of past uses	5	1
20% ≤ x < 40%	majority of past uses	16	1
40% ≤ x < 60%	no temporal preference	39	0
60% ≤ x < 80%	majority of present uses	22	21
80% ≤ x < 100%	large majority of present uses	7	66

* Table from Vafaeian (2018).

Since no temporal preference and majority of present is more spread out, and since the uses with present time reference are in majority when putting all the progressive segments together, I conclude that, typologically, it is more common for progressives to either have no temporal preference or a majority of present uses.

Map 1. The proportion of present segments among the languages in the PBC. Darker red indicates more present uses.
Map from Vafaeian (2018).

4.1   Temporal preference

In this section, grams that show no temporal preference, grams that show a preference for present tense and grams that show a preference for past tense in the PBC sample will be presented. This includes those grams with a present percentage between 20–79 in Table 3.

4.1.1   Grams with no temporal preference

A total of 39 of the 89 grams of the sample have no temporal preference. This group includes all Quechuan and Creole languages in the sample and most of the Zapotecan and Mayan languages. Also grams such as for example the English progressive, the Ot Danum (Austronesian) *rahat*, the Indonesian (Austronesian)

sedang, Hindi (Indo-European) *rah* + COP, Panjabi (Indo-European) *rah* + COP and Ewe (Niger-Congo) v-*m* fall into this group.

4.1.2   Grams with a majority of present uses

All Bantoid languages in the sample have a majority of present uses, some are even restricted to present and will not be discussed here. The Indo-European languages of the sample have either no temporal preference or a majority of present uses. Examples of the latter are Tosk Albanian *po*, Spanish and Portuguese *estar* + GER and Bengali v-*(c)ch(il)*. Other examples of grams that fall into this group are Min Nan Chinese (Sino-Tibetan) *the*, Bine (Western Fly) v-*eni* and Makaa (Niger-Congo) *ŋgə́*.

In Turkish (Turkic), -*Iyor* is used to express ongoing meaning. The data suggests that it has more present uses than past uses in the PBC. The other two Turkic languages of the sample, Kirghiz and Uzbec, have grams with high preference for present time reference and will be discussed in Section 4.2. These grams also occur much less often in the PBC, 516 and 520 times while -*Iyor* occurs 2,100 times. This is partly explained by -*Iyor* being less restricted temporally, but also by the fact that -*Iyor* has expanded its uses towards the imperfective and may for example occur with stative verbs (Kornfilt 1997: 357).

I believe, that the explanation for progressive grams generally having a majority of present uses has to do with incomplete contexts in general being more frequent in the present than in the past. In order to compare frequencies of use for incomplete versus completed events, we can look at languages that have an imperfective/perfective distinction. Russian is such a language and Janda & Lyashevskaya (2011: 723) show that past imperfective verb forms are much less common than both past perfective verb forms and non-past imperfective forms. Since almost all utterances in the present tense are incomplete in nature, performative utterances (e.g. "I now pronounce you husband and wife") being the only exception, incomplete utterances can be said to be much more frequent in the present than in the past. A language with a progressive will apply to a subset of these incomplete events, and close to none of the completed ones, which explains why progressives occur more often in contexts with present time reference than past: there are simply more incomplete events with present time reference than past. The asymmetry between incompleteness and time reference is probably even stronger in spoken language.

There is a strong correlation between the total frequency of the progressive grams of the PBC and their present and past proportions: grams with more occurrences in the corpus always have a high (but not the highest) proportion of present segments. This is illustrated in Figure 1 where the grams with high total frequency are found among those with a high present proportion.
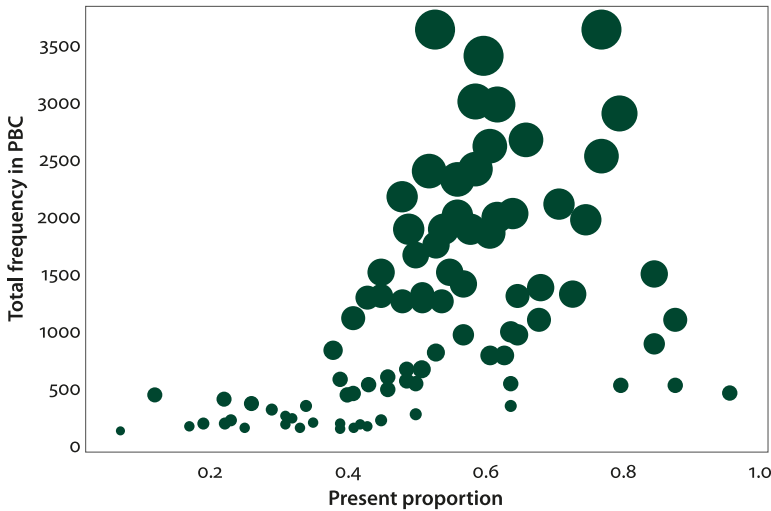
Figure 1. Total frequency and proportion of present segments in the PBC

In the figure, the total frequency of the grams is shown both on the y-axes and on the size of the dot. The present percentages are taken from Table 3. As can be seen, the largest grams have percentages spanning from approximately 50–80. Note also the drop in frequency when the percentage is higher than 80 which shows grams restricted to the present. The table also shows that while grams with fewer occurrences can be restricted to either present, past, or show no temporal asymmetry, grams with many occurrences must have many uses in the present. The fact that incomplete contexts are more frequent in the present than in the past has as a consequence that the takeover of new uses, which is part of the grammaticalization process of progressives, will have to involve more expansion in the present than in the past. It is as if there is more space in the present than in the past in which progressives can swell in their maturation process. Typically then, highly grammaticalized progressives are not restricted temporally but have a majority of present uses.

4.1.3   Grams with a majority of past uses

As already noted, most of the grams with a majority of past uses belong to the Austronesian language family. Apart from the two oceanic languages Hawaiian (Austronesian) and Kara (Austronesian), the Austronesian as well as the Austro-Asiatic languages of the sample only include languages that have either no temporal preference or a majority of past uses. Interestingly, almost all other non-Austronesian grams with past preference are found in this area: Central Khmer (Austro-Asiatic), both grams in Vietnamese (Austro-Asiatic), Hmong Daw (Hmong-

Mien), and Bawm Chin (Sino-Tibetan). This suggests that past preference for progressives is an areal pattern. Only two grams with majority of past uses are found that are not spoken in this area, Ozolotepec Zapotec (Zapotecan) and Jola-Fonyi (Niger-Congo). The latter is a past progressive and will be discussed in Section 4.2. Most, but not all of the Austronesian grams with high or majority of past uses are spoken in Indonesia and are therefore subject to influence from the majority language Indonesian (Austronesian). It may also be that their Bible translation has been influenced by the Indonesian translation. However, the Indonesian *sedang* has no temporal preference which suggests that the many Austronesian past uses are not merely direct calques of *sedang*.

In most cases, there is not sufficient data explaining this preference as some of these languages do not have (available) language descriptions. The grams with a majority of past uses are not past progressives, since they also have uses with present time reference. As shown in Table 3, such uses are even in the majority in 'direct speech' for all but two grams. The two grams with past preferences in 'direct speech' in the table are *lako* in Jola-Fonyi and *liau(ah)* in Bawm Chin. Although tense is not marked grammatically in Bawm Chin the gram seems to have very strong past preference.

A closer look at those languages for which more data is available shows that some grams have additional meanings somehow related to the meaning of 'while'. This could explain their preference for past time reference, which then is rather a closeness to the meaning 'while' or, as will be discussed below, a preference for subordination.

To start with, some grams have original meanings of 'middle'. The progressive *pintanga'* in Balantak (Austronesian), for example, is glossed as 'in.the.middle' in the grammatical description and translated as 'while' in Example (8). The grammar translates *tanga* as 'middle' (van den Berg & Busenitz 2012: 72).

(8) BALANTAK, AUTRONESIAN
    Pintanga'   *bisara koi-ya'a,*   *uar-kon-on-mo*   *a*   *panganon.*
    in.the.middle speak like-DEM3 go.out-CAUS-PV-1-PRF ART proposal.gift
    'While they are speaking thus, the marriage proposal gifts are taken out.'
                                                    (van den Berg & Busenitz 2012: 32)

Similarly, the Achinese (Austronesian) *teungoh* functions both as a marker of progressive and as meaning 'middle/in the middle of, during' (Asyik 1987: 115, 167–168).

(9) ACHINESE, AUSTRONESIAN
    a. *Ayah teungöh geu-peugah haba ngön jamèe.*
        father PROG   3-tell   story with guest
        'Father is/was talking with a guest.'

    b.   *Beuklam jaga-lōn teungöh malam.*
        last.night awake-1 middle night
        'Last night I awoke in the middle of the night'   (Asyik 1987: 167–168)

In Muna (Austronesian), the gram *tangasano* occurs almost exclusively in clauses with past time reference even though this language does not mark tense grammatically. In fact, René van den Berg (p.c.) comments that *tangasano* is not restricted temporally and may occur with both present and future time reference as seen in Example (10a) and (10b) respectively, although it typically occurs in subordinate clauses. In addition, René van den Berg notes that the Muna *tanga* is probably related to the Indonesian *tengah* 'middle' and could be a loan from the more prestigious language Wolio (Austronesian) (a language not included in the sample). Thus, the Balantak, Achinese and Muna grams seem to have originated from the meaning 'middle'. These progressive grams can then perhaps be assumed to have a meaning of '*being in the middle of* v-*ing*', which in turn is very similar to the meaning 'while', especially if this meaning is used in subordinations.

(10)   Present and future time reference with *tangasano* in Muna (Austronesian)
    a.   Acts 9:11
        Nokowambamo dua Ompu, Kalamo we kaangka'a konea'ano Sala Melaa.
        Ondofi we lambuno Yudas seemie mai'aono we Tarsus, neano Saulus.
        Ampa aitu tangasano nosambahea.
        'And the Lord said to him, "Get up, go to the street called 'Straight' and in the house of Judas look for a man named Saul from Tarsus. *For behold, he is praying*,'
    b.   Luke 17:35
        Ane dahodua hobhine tangasano megilino kahitela, seemie dawowo'oe, seemieno dahumunsae.
        '*There will be two women* grinding at the same place; one will be taken and the other will be left.'

Also, the Indonesian *sedang* (Austronesian) has, in addition to being a progressive marker, a meaning of 'medium, moderate, average' (Intan Fuji, p.c., Google translate). This gram is probably related to both Vietnamese (Austro-Asiatic) *đang* as well as Central Malay (Austronesian) *dang*.[10]

    Other grams have 'while' or 'as' as an additional meaning to the ongoing use: in Sasak (Austronesian), *kenyeke* is reported to mark both 'in the process of' as well as being translated to 'while' and 'as' (Austin 2012: 241–242; Eades 1998: 122),

---

10. Muna *se-tanga* has the meaning 'a half', despite the resemblance René van den Berg (p.c.) does not think this is related to Indonesian *sedang*.

and Vietnamese *đương* is translated as 'in the act of, during, while' (Thompson 1987: 270).

It is worth noting that there are also other links between progressives and 'while'. In Indonesian (Austronesian), the expression for 'while, whereas' is *sedangkan* (Sneddon et al. 2010: 346) which contains the progressive *sedang*.[11] Another way of expressing 'while' in Indonesian is through *lagi pas*, constructed from *lagi* 'action in progress' which can also mean 'again, still, more, other' and *pas* 'when' (Sneddon 2006: 54; Sneddon et al. 2010: 205).[12] Interestingly, it is commented that *pas lagi* is used for backgrounding contexts, or "to specify that the event occurred when something else was in progress" (Sneddon 2006: 54, 93).

(11)   *Lagi* and *pas lagi* in Indonesian
    a.   *Ada pengarang lagi   nyetir mobil.*
          be   writer   present drive car
          'There was a writer driving a car.'
    b.   *Pas   lagi kita ngobrol kan dia lewat diem aja.*
          when still we talk   DP he pass quiet just
          'While we were talking he walked past, very quietly.'

(Sneddon 2006: 54, 93)

Also, in Persian, a language not included in the PBC sample, the formal progressive *dar hāl=e* and the 'while' phrase *dar hāl=i ke* are based on the same construction, examples of these are shown below.

(12)   *Dar hāl=e* and *dar hāl=i ke* in Persian
    a.   *dar hāl=e   kār   kardan hast-am.*
          in   state=EZ work do.INF COP-1SG
          'I am working.'
    b.   *dar hāl=i   ke kār mi-kard-am omad   tu.*
          in   state=INDF that work IPFV-do-1SG come.PST.3SG in
          'While I was working, s/he came in.'   (Own knowledge)

The relationship between progressives in the Austronesian languages and 'while' is not completely clear at this point. Given the concrete meaning of 'middle' found in Balantak, Achinese and Muna, it is probable that some of these progressives have arisen from words for 'middle'. It also seems as if such developments have spread through language contact in this area, perhaps through calques from some

---

11.   In the PBC sample, the free standing *sedang* has been searched for, which does not include uses of *sedangkan*.

12.   The difference between *sedang* and *lagi* is that the former is used in formal settings while the latter is used in informal settings. *Lagi* is not included in the sample.

dominant languages. We can conclude then, that some grams have a preference for subordination which is linked to these grams' closeness in meaning to 'middle' and 'while' in different ways. Additionally, we can note that the meaning of '*being in the middle of* v-*ing*' is semantically close to the feature 'the involvement or busyness of the subject in the event' discussed in Section 3.3.

4.2   Temporal restrictions

In this section, those grams in Table 3 that showed a high preference for either present (80–100%) or past time reference (0–19%), most of which are restricted temporally, are presented and discussed.

4.2.1   Restriction to present time reference

Among the 89 grams examined in the PBC, several progressives restricted to present time reference were encountered. Among these is Hawaiian (Austronesian) *ke + nei* which was already mentioned as a present progressive gram in Dahl (1985: 94). All three Turkic languages of the sample have a majority of present uses, but while Turkish -*Iyor* is not restricted temporally, the Kirghiz (Turkic) *žata* 'lie' is referred to as the 'complex present tense' (Abylkasymova & Jumabaeva 1997: 306). In the PBC, it has a large majority of present uses, which confirms that this is a present progressive gram. Similarly, the Uzbek (Turkic) progressive gram has a large majority of present uses. This gram includes two forms: *jap* which is referred to as the 'focal present' and *yotgan edi* which is referred to as the 'focal past' (Hervé, nd). As can be seen, the forms are different phonetically as well as structurally and may therefore be considered to belong to two different paradigms. Unfortunately, I have not found any historical explanation for this discrepancy. In the PBC, the present form occurs 473 times while the past form occurs 47 times. If *jap* is analyzed as having no past form it would constitute a present progressive gram. If one were to analyze *jap* and *yotgan edi* as belonging to the same paradigm, we can conclude that the gram has a high preference for present time reference. In either case, there is a clear favoring of present uses among all three Turkic languages of the sample.

Several grams with a high preference for present are found among the Niger-Congo language family. In Southern Kisi (Niger-Congo), *chō* is presented as both a present progressive and as a future tense marker (Childs 1995). This is confirmed in this data as it has a large majority of present uses. Example (13) presents ongoing as well as future use. As can be seen, (13b) is not a future progressive use ('We will be seeing…') but a typical future example.

(13)   *SOUTHERN KISI, NIGER-CONGO*

    a.   *sàà cò ndú tàmbá lòòlùlló.*
         Saa AUX her Tamba beat
         'Saa is beating Tamba for her.'

    b.   *ŋ̀ cò ciiikiaŋ lɔ́ɔ́ ŋ̀   cò hùnɔ̀ɔ-ó.*
         we AUX meet   time you AUX come-SUF
         'We will see you when you come.'   (Childs 1995:117)

Also the Niger-Congo languages Lenje and Koongo[13] have grams that occur almost exclusively in the present. The Lenje *too* is referred to as a "present progressive and progressive where the action continuous from present to immediate future" (Kagaya 1987:24–25) and may then, similar to Southern Kisi *chō* have both progressive and (immediate) future uses. Unfortunately, no description is found where the Koongo gram *eti* is mentioned.

    Present progressives are also mentioned elsewhere in the literature. Apart from Hawaiian (Austronesian), Dahl (1985:94)[14] mentions Kikuyu (Atlantic-Congo) which has a marker *ra* used for the present progressive, Karaboro (Atlantic-Congo) which marks the present progressive with *mê* + the verb in the imperfective (referred to as a present continuous by Dombrowsky-Hahn (2015:398)[15] and Oromo (Afro-Asiatic)[16] which uses the imperfective form of the verb + a marker *gira* for the present progressive (see also Moreno 1939:152–153).

    Another example of a present progressives is *kʷahat* in Huastec (Mayan) provided by Edmonson (1988:592). In Paraguayan Guraraní (Tupian), *ína/hína* is presented as a progressive which indicates that something occurred in the present unless there is a modifier indicating that it happened in the past (Gregores & Suárez 1967:155). In Chichewa[17] (Niger-Congo), *li ku/ku* is restricted to the present tense and is a marker which can be regarded as a present progressive with extended uses toward the imperfective since it combines with dynamic as well as stative verbs (Kiso 2012:93–95).

    Given the discussion in Section 4.1.2, the explanation for the existence of present progressive grams is assumed to be a direct consequence of the higher frequency of incomplete events in the present and the higher frequency of completed

---

13. Referred to as South-Central Kikongo on glottolog.org.

14. Dahl (1985:94) also mentions Tigrinya which did not have any occurrences in past contexts in his questionnaire. In the PBC, the Tigrinya progressive has both present and past uses.

15. Dombrowsky-Hahn (2015:400) also mentions a past continuous formed with *puì* but gives no example.

16. Probably a variety in the grouping referred to as Borana-Arsi-Guji Oromo (p.c. Östen Dahl).

17. Referred to as Nyanja on glottolog.org.

events in the past: there is a higher functional need for progressives in the present than in the past which explains why we find languages with progressives restricted to present time reference. Also, this study has shown that the contexts in which progressives occur differ in present and past meaning that there are different functional needs for the progressive in the two time references. As will be shown in the next section, the sample only included one gram restricted to the past, although other past progressives are found elsewhere in the literature.

4.2.2   Restriction to past time reference

Among the 89 grams of the PBC, only one past progressive gram was encountered, namely *lako* in Jola-Fonyi (Niger-Congo). The gram *lako* originates from 'stay, sit' and has previously been referred to as a progressive, an imperfect and a past progressive (Sapir 1965:104; Diatta 1998:199; Hopkins 1995:149). Its occurrences in the PBC strongly indicate that it is restricted to the past. However, there is mentioning of other progressives in the language that have present uses as well (see e.g. Blansitt 1975:17; Hopkins 1995:156–160). These were not included in the sample either because they were not possible to capture automatically or did not meet the requirements of having the same distribution as the other progressives. In addition to marking the progressive, *lako* is still used in its original meaning and also has locative uses (see Hopkins 1995). It is shown in Figure 1 as the gram with present uses of around 20% and a total frequency of 1084. The figure shows that the total frequency is rather high in comparison to other progressives with a majority of past uses. The high frequency is probably due to the erroneous inclusion of some locative uses and original meaning of *lako*.

In the literature, other past progressives have been mentioned. In Lithuanian (Indo-European), which is not included in the sample, the copula *būti* combines with the present active participle to form past and future progressives, while the simple present tense includes present ongoing uses (Blansitt 1975:20). A present counterpart to the past and future progressive constructions exists but is not used "because it would have the same meaning as the simple present" (Dambriunas et al. 1980:341).

Hungarian (Uralic), also not included in the sample, is similarly noted to have a past progressive which lacks a present counterpart. This progressive lacks segmental marking, instead "word order together with a specific intonation contour of the clause allows for a progressive interpretation", therefore it is not regarded as a "genuine progressive" (Bertinetto et al. 2000:525). The Hungarian past progressive also has many restrictions, e.g. with regard to durative adverbials, negation, semelfactive verbs and the thematic role of the agent (see Kiefer 1994).

It is unclear how the temporal asymmetries have come to exist in Jola-Fonyi, Lithuanian and Hungarian. A language with a past but not present progressive

for which we have such information is Taleshi (Indo-European), an Iranian language spoken in Iran and Azerbaijan. This language is not included in the PBC. The constructional pattern differs somewhat depending on the variety of Taleshi but generally speaking, the past progressive in this language is formed through the use of the locative, realized as *dæ* or a similar cognate, which attaches to the infinitive form of the verb followed by a past auxiliary: V=INF=LOC=AUX (Stilo 2008: 373–374). In the present, the same pattern without the auxiliary, V=INF=LOC, is used as a general present tense. This asymmetry in Taleshi can be explained through the process of further grammaticalization of the progressive, which has happened in the present tense but not (yet) in the past. Example (14) shows the present tense pattern and the past progressive pattern, respectively. The language also has a past imperfective and a past perfective form (Stilo 2008).

(14) *NORTHERN TALESHI, INDO-EUROPEAN*
    a.   *vašt-é=dæ=m.*
        jump.PST=INF=TAM=SET1
        'I jump.'
    b.   *vašt-é=dæ=b-im.*
        jump.pst-INF=TAM=AUX=SET1
        'I was jumping.'   (Stilo 2008: 376)

Some Northern Taleshi varieties spoken in Azerbaijan are showing tendencies of a grammaticalization towards imperfective also in the past, where the pattern can be used both with ongoing events (15a) and in habitual contexts (15b).

(15) *TALESHI, LERIKI VARIETY, INDO-EUROPEAN*
    a.   *ayïl vít-dæ=b-e   bæ di mašin-í.*
        child run-LOC=AUX.PST-3SG to after car-OBJ
        'The child was running after the car.'
    b.   *penj sor vaxt doy-dæ=b-in.*
        five year time give-LOC=AUX.PST-3PL
        'They used to give 5 year leeway.'   (Stilo in press)

In the neighboring Iranian languages Tati, Gilaki and Mazandarani, all spoken in Iran, similar constructional patterns form progressive grams. These grams do not have uses which indicate grammaticalization towards the imperfective. In Gilaki and Mazandarani, patterns formed with the locative *dər/dar*, V-INF + LOC and LOC + V, respectively, are used for ongoing events in both tenses. These tenses exist alongside separate present and past tense construction. In Tati, a present progressive pattern, V INF-LOC-COP, is used for ongoing events, here the locative element is realized as *u* instead. No information about a past progressive is available for Tati.

The development of the Taleshi present progressive into a general present has most likely happened under the influence of the dominant Turkic languages Turkish or Azerbaijani. In both language, the progressive patterns *mAktA* and *mAKdA*, respectively, are formed through the infinite and the locative, i.e. v-INF-LOC (see e.g. Schönig 1998). Miller (1953:146), noting the similarity, suggests that Taleshi has borrowed the pattern from Turkish or Azerbaijani. Recall also that all Turkic languages in the sample have a majority of present uses, and that the progressive in at least Kirghiz, is a present progressive. The dominant language of Iran, Persian, on the other hand, has a progressive pattern in spoken language formed with the *dāštan* 'have', which does not show expansions towards the imperfective. This is also reflected in the progressive patterns in Gilaki, Mazandarani and Tati.

In Taleshi then, the present progressive has, under the influence of Turkish and/or Azerbaijani, grammaticalized faster than its past counterpart, replacing the older present form. In the past, the progressive remains a marker of ongoing events although some dialects are showing tendencies towards further grammaticalization in the past as well. This example provides us with one historical explanation of how progressives restricted to past tense can come to exist. Future research is needed in order to establish whether there also exist other paths for this type of temporal asymmetry.

5.   Conclusion

This study has been concerned with the relationship between progressive grams and present and past time reference in several ways: the prototypical uses of progressives in the present and past, the proportion of present and past uses among progressives and progressives with temporal restrictions. In addition to that, progressives at different stages of grammaticalization and uses that arise as progressives mature further have been discussed.

It was shown that the most typical uses of progressives differ in contexts with present and past time reference. While progressives are used for topical and acute events in the present, they are mainly used as backgrounding contexts in the past. The shared contexts among the grams of the two samples resulted in the presentation of four features that increase the likelihood of the use of a progressive: (a) a punctual reference point, (b) an emotive component, (c) the involvement or busyness of the subject in the event and, (d) the desire to turn the attention of the speaker towards an ongoing event.

It was also shown that progressive grams occur most often in contexts with present time reference. This was shown for progressives in the PBC, a written source which in total contains more past verb forms than present ones. This result

was explained by the way we talk of events in the present and past: events occurring at the speech moment are not yet ended while events in the past are most often ended, meaning that almost all utterances with present time reference are incomplete – and therefore available for the progressive – while the majority of events in the past are completed – and therefore not available for the progressive. The relationship between the complete/incomplete asymmetry and temporal reference was shown to be of great importance throughout the paper. It was shown that this temporal asymmetry is a consequence of how we conceptualize present and past events and how this interacts with an aspectual marker such as the progressive. The consequence of the difference in the conceptualization of present versus past time discussed in this paper is not restricted to the progressive but should have implications for other gram types as well.

Progressives with a majority of past uses were shown to be restricted areally and mainly found among Austronesian languages. These uses are linked to subordination and the similarity in meaning between ongoingness and 'being in the middle of V-ing'. It was concluded that such uses constitutes an areal feature.

In the sample of the PBC, several grams restricted to present tense were found but only one past progressive. In the literature, however, other present as well as past progressives have been noted. Two historical processes giving rise to these two temporal asymmetries were suggested: present progressive grams were assumed to be a result of the greater functional need for progressives in the present; past progressives can be residues of patterns where the present has grammaticalized into a general present tense while the past counterpart remains a past progressive. Interestingly, both of these processes are a result of the higher frequency of present uses over past for progressive grams. There are, however, probably also other paths leading to temporal restrictions among progressives. For example, several languages have been mentioned in the article which have separate patterns for present and past progressive, suppletive paradigms as it were, e.g. Uzbek and possibly also Jola-Fonyi. Although not temporally restricted, these paradigms suggest that the progressives in the present and past have come to exist through different paths, an issue that merits further investigation.

In spite of these examples from the literature, the number of attested past progressives remains lower than that of present progressives. Present progressives are also spread out areally as well as genealogically and found in Atlantic-Congo, Niger-Congo, Turkic, Austronesian, Mayan and Tupian languages although the concentration among languages of the African continent as well as Turkic is noted. Past progressives were possibly found in Niger-Congo, Indo-European and Uralic languages. Due to this, I will conclude that present progressive grams are more common typologically than past progressives.

It was also shown that highly grammaticalized progressives will typically have a majority of present uses while progressives with fewer uses can be restricted to the present, the past or have no temporal restrictions. Again, this is linked to how we talk of events in the present and past and the dominance of incomplete events in the present but not in the past: as progressives expand their functions, they will take over more incomplete uses which are more frequent in the present than in the past.

In addition to that, the study has shown that the progressive gram type includes members that differ from each with regard to their level of grammaticalization which ultimately means that these grams may have very different distributions. It was shown that the requirement of emotive component is linked to non-obligatory contexts (which are by definition pragmatically marked) and absent in obligatory ones. This means that emotive components are present in almost all instances of less grammaticalized progressives but also in more mature progressives such as the English one in uses in which it is not the default choice. Such uses are part of the English progressives' expansion to new contexts and part of its further grammaticalization towards the imperfective.

In several ways then, the members of the progressive gram type do not constitute a homogeneous group: progressive grams differ within one and the same gram in that their present and past uses are different, as well as among grams both with regard to their temporal preferences or restrictions and with regard to their level of grammaticalization. In this light, what are the reasons for calling all these patterns by one name? Or put differently: how homogeneous must members of a gram type be? In the case of the progressive, one could argue that all members need to share the prototypical uses of 'ongoing events at reference time' as their main type of use or at least as a prominent type of use. However, it may be warranted to postulate sub-types to the progressive gram type as some grams share their main uses with the majority of grams but differ from them in their peripheral uses either as a result of them being less grammaticalized (in which case they lack peripheral uses) or in the sense that they have uses that can be viewed as part of their further grammaticalization toward the imperfective. Also, temporally restricted progressives partly differ from temporally non-restricted progressives in their shared uses and may therefore also constitute sub-types. A different analysis would be to view these grams as part of different gram types in which case one would need to account for the many shared uses of ongoingness among these patterns.

Many grams discussed in this study are then on a path of grammatizalization either 'into' or 'out of' the progressive gram type. The problem of identifying a homogeneous cluster for the progressive gram type is in different ways a consequence of the volatile characteristics of these grams which is linked to them being

typically non-obligatory in utterances. In this sense, the progressive is somewhat different from other more stable gram types such as e.g. the imperfective.

Abbreviations

–	inflectional boundary	IPFV	imperfective
=	clitic boundary	LOC	locative
1	first person	NMLZ	nominalizer
2	second person	OBJ	object
3	third person	PL	plural
AUX	auxiliary	PRF	perfect
ART	article	PROG	progressive
CAUS	causative	PRS	present
COP	copula	PST	past
DEM3	demonstrative root 3	PV	patient voice
DP	discourse particle	SG	singular
EZ	ezafe	SET1	set of person agreement markers
GER	gerund	SUF	suffix
INDF	indefinite	TAM	tense, aspect, mood
INF	infinitive		

References

Abylkasymova, Mairam & Gulaim Jumabaeva. 1997. *Kyrgyz language manual*. Bishkek, Kyrgyzstan: Peace Corps.

Anthonissen, Lynn, Tanja Mortelmans & Astrid De Wit. 2016. Aspect meets modality: A semantic analysis of the German *am*-progressive. *Journal of Germanic Linguistics* 28(1). 1–30. https://doi.org/10.1017/S1470542715000185

Asyik, Abdul Gani. 1987. A Contextual Grammar of Acehnese Sentences (Complementation). Ann Arbor: UMI PhD dissertation.

Austin, Peter K. 2012. Tense, aspect, mood and evidentiality in Sasak, eastern Indonesia. In Peter K. Austin & Stuart McGill (eds.), *Language documentation and description*, vol. 11, 41–56. London: SOAS.

Berg, René van den & Robert L. Busenitz. 2012. *A grammar of Balantak: A language of Eastern Sulawesi*. Dallas, Texas: SIL International.

Bertinetto, Pier Marco. 2000. The progressive in Romance, as compared with English. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe* (Empirical Approaches to Language Typology 20), 559–664. Berlin: De Gruyter Mouton.

Bertinetto, Pier Marco & Denis Delfitto. 2000. Aspect vs. actionality: Why they should be kept apart. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe* (Empirical Approaches to Language Typology 20), 203–226. Berlin: De Gruyter Mouton.

Bertinetto, Pier Marco, Karen Ebert & Casper de Groot. 2000. The progressive in Europe. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe* (Empirical Approaches to Language Typology 20), 517–558. Berlin: De Gruyter Mouton.

Blansitt, Edward L. Jr. 1975. Progressive aspect. *Working papers on Language Universals* 18. 1–34.

Blensenius, Kristian. 2015. Progressive constructions in Swedish. Gothenburg: Gothenburg University PhD dissertation.

Bybee, Joan & Östen Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13. 51–103. https://doi.org/10.1075/sl.13.1.03byb

Bybee, Joan, Revere Perkins & William Pagliuca. 1994. *The evolution of grammar. Tense aspect, and modality in the languages of the world*. Chicago: The University of Chicago Press.

Childs, G. Tucker. 1995. *A grammar of Kisi: A Southern Atlantic language* (Mouton Grammar Library 16). Berlin, New York: De Gruyter Mouton. https://doi.org/10.1515/9783110810882

Comrie, Bernard. 1976. *Aspect. An introduction to the study of verbal aspect and related problems*. Cambridge: Cambridge University Press.

Dahl, Östen. 1985. *Tense and aspect systems*. Oxford: Blackwell.

Dahl, Östen. 2004. *The growth and maintenance of linguistic complexity*. Amsterdam: John Benjamins. https://doi.org/10.1075/slcs.71

Dahl, Östen. 2013. How telicity creates time. *Journal of Slavic Linguistics* 21(1). 45–76. https://doi.org/10.1353/jsl.2013.0004

Dahl, Östen. 2020. Perfects in typological perspective. In Robert Samuel David Crellin & Thomas Jügel (eds.), *Perfects in Indo-European languages and beyond*, 635–667. Amsterdam; John Benjamins. https://doi.org/10.1075/cilt.352.19dah

Dahl, Östen. 2022. Perfects across languages. *Annual Review of Linguistics* 8(1). 279–297. https://doi.org/10.1146/annurev-linguistics-031120-123428

Dahl, Östen & Bernhard Wälchli. 2016. Perfects and iamitives: Two gram types in one grammatical space. *Letras de Hoje* 51(3). 325–348. https://doi.org/10.15448/1984-7726.2016.3.25454

Dambriunas, Leonardas, Antanas Klimas & William R. Schmalstieg. 1980. *Introduction to modern Lithuanian*. 3rd edn. New York: Franciscan Fathers.

De Wit, Astrid & Frank Brisard. 2014. A Cognitive Grammar account of the semantics of the English present progressive. *Journal of Linguistics* 50(1). 49–90. https://doi.org/10.1017/S0022226713000169

De Wit, Astrid & Adeline Patard. 2013. Modality, Aspect and the Progressive. The semantics of the present progressive in French, in comparison with English. *Languages in Contrast* 13(1). 113–132. https://doi.org/10.1075/lic.13.1.06wit

De Wit, Astrid, Adeline Patard & Frank Brisard. 2013. A contrastive analysis of the present progressive in French and English. *Studies in Language* 37(4). 846–879. https://doi.org/10.1075/sl.37.4.05wit

De Wit, Astrid, P. Petré & F. Brisard. 2020. Standing out with the progressive. *Journal of Linguistics* 56. 479–512. https://doi.org/10.1017/S0022226719000501

Diatta, Christian Sina. 1998. *Parlons jola: Langue et culture* (Collection Parlons). Paris: L'Harmattan.

Dombrowsky-Hahn, Klaudia. 2015. *A grammar of Syer (Western Karaboro, Senufo)* (Gur Monographs 12). Cologne: Köppe.

Eades, Yusuf. 1998. Relativization. In Peter K. Austin (ed.), *Sasak* (Working Papers in Sasak 1), 119–129. Melbourne: University of Melbourne, Lombok and Sumbawa Research Project.

Edmonson, Barbara Wedemeyer. 1988. A descriptive grammar of Huastec (Potosino Dialect). Ann Arbor: UMI PhD dissertation.

Einarsson, Stefán. 1949. *Icelandic: Grammar, texts, glossary*. Baltimore: Johns Hopkins University Press.

Gregores, Emma & Jorge A. Suárez. 1967. *A description of colloquial Guaraní* (Janua Linguarum: Series Practica 27). The Hague: Walter de Gruyter. https://doi.org/10.1515/9783111349633

Hatcher, Anna. 1951. The use of the progressive form in English: A new approach. *Language* 27(3). 254–280. https://doi.org/10.2307/409755

Hervé, Guérin. nd. The Uzbek tense/aspect/modality system. http://humanities-web.s3.us-east-2.amazonaws.com/ceeres/prod/2021-06/the_uzbek_tenseaspectmodality_system.pdf

Hopkins, Bradley L. 1995. *Contribution a une etude de la syntaxe Diola-Fogny* (Cahiers de Recherche Linguistique 4). Dakar: Société Internationale de Linguistique.

Hopper, Paul. 1979. Aspect and foregrounding in discourse. In Talmy Givón (ed.), *Discourse and syntax* (Syntax and Semantics 12), 231–241. New York: Academic Press.

Janda, Laura A. & Olga Lyashevskaya. 2011. Grammatical profiles and the interaction of the lexicon with aspect, tense, and mood in Russian. *Cognitive Linguistics* 22(4). 719–762. https://doi.org/10.1515/cogl.2011.027

Johanson, L. 2000. Viewpoint operators in European languages. In Östen Dahl (ed.), *Tense and aspect in the languages of Europe* (Empirical Approaches to Language Typology 20), 27–188. Berlin: De Gruyter Mouton.

Kagaya, Ryohei. 1987. *A classified vocabulary of the Lenje language* (Bantu Vocabulary Series 4). Tokyo: Institute for the Study of Languages and Cultures of Asia and Africa.

Kiefer, Ferenc. 1994. Aspect and syntactic structure. In Ferenc Kiefer & Katalin É. Kiss (eds.), *Syntax and semantics* (The Syntactic Structure of Hungarian 27), 415–464. San Diego: Academic Press.

Kiso, Andrea. 2012. Tense and aspect in Chichewa, Citumbuka and Cisena: A description and comparison of the tense-aspect systems in three southeastern Bantu languages. Stockholm: Stockholm University PhD dissertation.

Kornfilt, Jaklin. 1997. *Turkish* (Descriptive Grammars Series 1). London: Routledge.

Ljung, Magnus. 1980. *Reflections on the English progressive*. Göteborg: Acta Universitatis Gothoburgensis.

Miller, Boris V. 1953. *Talyšskij jazyk*. Moskva: Izdat. Akad. Nauk SSSR.

Moreno, Martino Mario. 1939. *Grammatica teorico-pratica della lingua galla, con esercizi*. Milano: Casa Editrice A. Mondadori Collezione Diretta da Carlo Conti Rossini. Accademio D'Italia, Roma.

Östling, Robert. 2015. Bayesian models for multilingual word alignment. Stockholm: Stockholm University PhD dissertation.

Sapir, David J. 1965. *A grammar of Diola-Fogny: A language spoken in the Basse-Casamance region of Senegal* (West African Language Monograph Series 2). Cambridge: Cambridge University Press.

Schönig, Claus. 1998. Azerbaijanian. In Lars Johanson & Éva Ágnes Csató (eds.), *The Turkic languages*, 248–260. London: Routledge.

Sneddon, James Neil. 2006. Colloquial Jakartan Indonesian (Pacific Linguistics 581). Canberra: Research School of Pacific and Asian Studies, Australian National University.

Sneddon, James Neil, Alexander Adelaar, Dwi Noverini Djenar & Michael C. Ewing. 2010. *Indonesian reference Grammar*. 2nd edn. Sydney: Allen & Unwin.

Stilo, Donald L. 2008. Two sets of mobile verbal person agreement markers in the Northern Talyshi language. In Simin. Karimi, Donald L. Stilo & Vida. Samiian (eds.), *Aspects of Iranian linguistics*, 363–390. Newcastle: Cambridge Scholars Publishing.

Thompson, Laurence C. 1987. *A Vietnamese reference grammar*. (Mon Khmer Studies 13–14). Honolulu: University of Hawai'i Press.

Vafaeian, Ghazaleh. 2018. Progressives in use and contact: A descriptive, areal and typological study with special focus on selected Iranian languages. Stockholm: Stockholm University PhD dissertation.

Vries, Lourens de. 2007. Some remarks on the use of Bible translations as parallel texts in linguistic research. *Language Typology and Universals* 60(2). 148–157. https://doi.org/10.1524/stuf.2007.60.2.148

Wälchli, Bernhard. 2019. The feminine anaphoric gender gram, incipient gender marking, maturity, and extracting anaphoric gender markers from parallel texts. In Francesca Di Garbo, Bruno Olsson & Bernhard Wälchli (eds.), *Grammatical gender and linguistic complexity*. Vol. II: *World-wide comparative studies* (Studies in Diversity Linguistics II), 61–131. Berlin: Language Science Press.

Windfuhr, Gernot & John R. Perry. 2009. Persian and Tajik. In Gernot Windfuhr (ed.), *The Iranian languages*, 416–545. London, New York: Routledge.

Appendix

Below, the grams in the PBC are displayed and grouped after their classification. The name of the language is followed by the ISO 639–3 code, the search string used in the PBC, the total number of occurrences of the gram in the corpus and the recall and precision values for the whole gram set. In the search string, the hyphen indicates an optional inflectional boundary (e.g. *sha* searches for *sha, sha, sha* and *sha*) and *=V* means that the search string is an element annotated as a verb in the corpus. For more complex patterns, a structural schema is given. The group numbers refer to the percentage of present according to Table 3:

group 1:   large majority of past uses
group 2:   majority of past uses
group 3:   no temporal preference
group 4:   majority of present uses
group 5:   large majority of present uses

Afro-Asiatic
Biu-Mandara
Merey [meq] *faya* 3648 (0.9, 0.649) group 3
Semitic
Tigrinya [tir] *ʾālo-, ʾālā-, zolo-, zolā-, naye-, nére-, naba-* 1896 (0.2, 0.683) group 3
Austro-Asiatic
Khmer
Central Khmer [khm] *kampoung-* 174 (0.9, 0.958) group 2
Viet-Muong
Vietnamese [vie] *đường* 200 (0.8, 0.946)   group 2
Vietnamese [vie] *đang*   485 (0.85, 0.927) group 3
Austronesian
Barito
Ma–anyan [mhy] *rahat* 178 (0.6, 0.994)   group 2
Ngaju [nij] *metoh(-)*   400 (0.55, 0.789) group 2
Ot Danum [otd] *rahat*   180 (0.35, 0.994) group 3
Celebic
Balantak [blz] *pintangaʹ* 149 (0.75, 0.915) group 2
Muna [mnb] *tangasano* 124 (0.9, 0.975)   group 1
Central Malayo-Polynesian
Bima [bhp] *wunga* 302 (0.85, 0.912) group 2
Greater Central Philippine
Gorontalo [gor] *donggo* 435 (0.95, 0.908) group 3
Javanese
Javanese [jav] *lagi* 223 (0.5, 0.856) group 2
Lampungic
Lampung Api [ljp] *sedang* 171 (0.4, 0.967) group 3
Malayo-Sumbawan
Achinese [ace] *teungoh*   357 (0.95, 0.965) group 2
Central Malay [pse] *dang*   190 (0.55, 0.953) group 2
Indonesian [ind] *sedang*   157 (0.5, 0.972)   group 3
Jarai [jra] *hlak*   566 (0.9, 0.903)   group 2
Madurese [mad] *teppana*   152 (0.65, 0.963) group 1
Minangkabau [min] *sadang*   341 (1.0, 0.964)   group 2
Sasak [sas] *kenyeke*   232 (0.55, 1.0)   group 2
Standard Malay [zsm] *sedang* 155 (0.45, 1.0)   group 2
Northern Luzon
Eastern Bontok [ebk] *cha* 1111 (0.85, 0.77) group 3
Oceanic
Hawaiian [haw] *ke + v + nei*   453   (0.05, 0.918) group 5
Kara (Papua New Guinea) [leu] *taxa* 2009 (0.85, 0.884) group 4
Rejang
Rejang [rej] *gidong* 135 (0.5, 0.984) group 2

South Sulawesi
Bambam [ptu] *mahassa-* 181 (0.8, 0.911) group 1
Creoles and Pidgins
Belize Kriol English [bzj] *di* + v 1516 (0.8, 0.84)   group 3
Morisyen [mfe] *pe*   1754 (0.95, 0.921) group 3
Seselwa Creole French [crs] *pe* 1410 (1.0, 0.936)   group 3
Hmong-Mien
Hmong Daw [mww] *tabtom* 221 (0.95, 0.986) group 2
Indo-European
Albanian
Tosk Albanian [als] *po* 1104 (0.8, 0.803) group 4
Germanic
English [eng] *be* + GER 1252 (0.95, 0.917) group 3
Indic
Bengali [ben] v-*(c)ch(il)-* 1323 (0.65, 0.93)   group 4
Hindi [hin] *rah-* + COP   210   (0.9, 0.733)   group 3
Panjabi [pan] *rah-* + COP 1327 (0.85, 0.886) group 3
Romance
Portuguese [por] *estar* + GER 538 (0.8, 0.969) group 4
Spanish [spa] *estar* + GER   339 (0.5, 0.969) group 4
Mayan
Achi [acr] *tijin*   652   (0.85, 0.892) group 3
Aguacateco [agu] *-tzan-*   994   (0.65, 0.832) group 4
Chol [ctu] *woli-*   1306 (0.9, 0.8)   group 4
Chuj [cac] *van*   581   (0.75, 0.929) group 3
K'iche' [quc] *tajin*   801   (0.95, 0.939) group 3
Kekchí [kek] *yọ-* + *chi-*   1315 (0.95, 0.887) group 3
Popti' [jac] *laňan*   264   (0.7, 0.892)   group 3
Q'anjob'al [kjb] *lanan-*   530   (0.8, 0.94)   group 3
Tz'utujil [tzj] *-emjon*   520   (1.0, 0.916)   group 3
Tzeltal [tzh] *yac-*   2628 (0.95, 0.677) group 4
Uspanteco [usp] *-tijin-*   954   (0.95, 0.923) group 3
Western Kanjobal [knj] *lalan* 551   (0.9, 0.971)   group 3
Yucateco [yua] *táan*   1898 (0.75, 0.777) group 3
Niger-Congo
Bantoid
Koongo [kng] *-eti*   1493 (0.15, 0.793) group 5
Lenje [leh] *-too-*   1097 (0.1, 0.855)   group 5
Makaa [mcp] *ŋgɔ́*   2022 (0.7, 0.806)   group 4
Mbunda [mck] COP + *naku-* 1379 (1.0, 0.887)   group 4
Nyoro [nyo] *n(i)-...-a*   2985 (0.9, 0.725)   group 4
Venda [ven] *khou*   786   (0.5, 0.924)   group 4

Zemba [dhm] *ma-*   2535 (0.0, 0.624)   group 4
Cross River
Gokana [gkn] *gé* 3409 (0.75, 0.679) group 4
Dogon
Toro So Dogon [dts] v-*wɔ-* 3662 (0.5, 0.714) group 4
Eastern Mande
Busa [bqp] *-tɛn* 1846 (0.95, 0.88) group 4
Kru
Kuwaa [blh] v-*nù* 2422 (0.8, 0.824) group 3
Kwa
Ewe [ewe] v-*m*   2022 (0.95, 0.803) group 3
Gen [gej] *le =V*   2676 (0.8, 0.851)   group 4
Sekpele [lip] *-lɛ* + v 972   (0.65, 0.814) group 4
Mel
Southern Kisi [kss] *chō* 2904 (0.05, 0.7) group 5
Northern Atlantic
Jola-Fonyi [dyo] *-lako-* 1084 (0.8, 0.711) group 1
Wolof [wol] *ngi*   885   (0.05, 0.9)   group 4
Western Mande
Susu [sus] v-*fe* 1254 (0.7, 0.728) group 3
Nilo-Saharan
Nilotic
Kumam [kdi] *-tye* + v 777 (0.5, 0.956) group 4
Oto-Manguean
Zapotecan
Amatlán Zapotec [zpo] *ka-*   1295 (1.0, 0.688)   group 3
Chichicapan Zapotec [zpv] *ca-*v   1265 (0.7, 0.807)   group 3
Coatecas Altas Zapotec [zca] *ki-*v   1665 (0.85, 0.74)   group 3
Ozolotepec Zapotec [zao] *nge-*   824   (1.0, 0.776)   group 2
Santo Domingo Albarradas Zapotec [zas] *ca-* 1889 (0.95, 0.778) group 3
Quechuan
Ayacucho Quechua [quy] *-chka-*   2328 (0.85, 0.8)   group 3
Cajamarca Quechua [qvc] *-yka-*   1500 (0.6, 0.808)   group 3
Cusco Quechua [quz] *-sha-*   2398 (0.65, 0.83)   group 3
Eastern Apurímac Quechua [qve] *-sha-* 3016 (0.6, 0.803)   group 3
South Bolivian Quechua [quh] *-sha-* 2176 (0.85, 0.866) group 3
Sino-Tibetan
Chinese
Min Nan Chinese [nan] *teh* 794 (1.0, 0.926) group 4
Kuki-Chin
Bawm Chin [bgr] *liau(ah)* 435 (0.7, 0.701) group 1
Trans-New Guinea
Angan

Angaataha [agm] *taat*ɨ*, taam*ɨ*, taaw*ɨ 659 (0.65, 0.749) group 3
Turkic
Kirghiz [kir] *žata-*   516   (0.1, 0.909)   group 5
Turkish [tur] *-yor-*   2100 (0.15, 0.764) group 4
Uzbek [uzb] *-jap-, -yotgan edi-* 520   (0.1, 0.93)   group 5
Uralic
Saami
Northern Sami [sme] COP + V.AKTIO.ESSIVE 445 (0.9, 0.847) group 3
Uto-Aztecan
Aztecan
Zacatlán-Ahuacatlán-Tepetzintla Nahuatl [nhi] v-*to-* 1259 (0.8, 0.823) group 3
Western Fly
Bine [bon] v-*eni* 1980 (0.15, 0.753) group 4

Address for correspondence

Ghazaleh Vafaeian
Department of Linguistics
Stockholm University
106 91 Stockholm
Sweden
ghazaleh@ling.su.se
https://orcid.org/0000-0002-5167-1424

Publication history