

Perfects and iamitives: two gram types in one grammatical space

Perfeitos e iamitivos: dois tipos de gram em um espaço gramatical

Östen Dahl

Department of Linguistics, Stockholm University

Bernhard Wälchli

Department of Linguistics, Stockholm University



Abstract: This paper investigates the grammatical space of the two gram types – perfects and iamitives. Iamitives (from Latin *iam* ‘already’) overlap in their use with perfects but differ in that they can combine with stative predicates to express a state that holds at reference time. Iamitives differ from ‘already’ in having a higher frequency and showing a strong tendency to be grammaticalized with natural development predicates. We argue that iamitives can grammaticalize from expressions for ‘already’. In this study, we extract perfect grams and iamitive grams iteratively starting with two groups of seed grams from a parallel text corpus (the New Testament) in 1107 languages. We then construct a grammatical space of the union of 370 extracted grams by means of Multidimensional Scaling. This grammatical space of perfects and iamitives turns out to be a continuum without sharp boundaries anywhere.

Keywords: Perfect; Iamitive; Gram (type); Grammatical space; Parallel texts; Generalized distribution; Multidimensional Scaling; Grammaticalization; Natural development; Transition to new scene; Extended time span

Resumo: Este artigo investiga o espaço gramatical de dois tipos de gram – perfeitos e iamitivos. Iamitivos (do latim *iam* “já”) sobrepõem-se em seu uso com perfeitos, mas diferem naquilo em que podem combinar-se com predicados estativos que expressem um estado detentor de referência temporal. Iamitivos diferenciam-se de “já” por apresentarem maior frequência e forte tendência à gramaticalização com desenvolvimento natural de predicados. Argumenta-se que iamitivos podem gramaticalizar a partir de expressões para “já”. Neste estudo, extraem-se grams perfeitos e grams iamitivos iterativamente, começando com dois grupos de grams disseminados de um texto de corpus paralelo (o Novo Testamento) em 1107 línguas. Em seguida, constrói-se um espaço gramatical da união de 370 grams extraídos por meio de Escalonamento Multidimensional. Tal espaço gramatical de perfeitos e iamitivos acaba mostrando-se um contínuo sem fronteiras nítidas em qualquer lugar.

Palavras-chave: Perfeito; Iamitivo; Gram (tipo); Espaço gramatical; Texto paralelo; Distribuição generalizada; Escalonamento Multidimensional; Gramaticalização; Desenvolvimento natural; Transição para novo cenário; Período de tempo prolongado

1 Introduction

This paper investigates the grammatical space of the two gram types perfects and iamitives, which are introduced in Section 2.¹ We argue that the perfect does not have a single unified meaning. Both Current relevance

theory (“CR theory”) and Extended now theory (“XN theory”), to use McCoard’s (1978) terms, are needed to capture the semantics and pragmatics of perfects, in that different uses of perfects involve elements of both to varying degrees. There is a close affinity between the semantics of the perfect and that of English *already* and its translational equivalents in other languages. Both have a “transition to new scene” profile, along the lines of Current relevance theory of the perfect. Many languages

¹ The research for this paper has been partly funded by the Swedish Research Council (*Vetenskapsrådet*, 421-2011-1444). We thank Tania Kuteva for comments on an earlier version.



have more grammaticalized expressions for ‘already’, which we call iamitives (from Latin *iam* ‘already’) and which overlap in their use with perfects but differ in that they can combine with stative predicates to express a state that holds at reference time, in the same way as English *already*. Iamitives differ from ‘already’ in having a much higher rate of occurrence and showing a strong tendency to be grammaticalized with natural development predicates, that is, predicates that become true sooner or later under normal circumstances.

Both perfects, such as the English Perfect, and iamitives, such as Indonesian *sudah*, are language-particular grammatical categories with specific forms and specific meanings, so called grams. In Section 3 we argue that gram types can be viewed as clusters of grams in grammatical space.

Perfects differ in their use across languages and so do iamitives. The question thus arises as to whether perfects and iamitives form two distinct clusters or one continuum. In order to explore this, we extract perfect grams and iamitive grams iteratively starting with two groups of seed grams from a parallel text corpus (the New Testament) in 1107 languages. Method and data are introduced in Section 4.

We construct a grammatical space of the union of 305 extracted grams by means of Multidimensional Scaling (MDS). The resulting grammatical space is explored in Section 5. In the four corners of the first two dimensions of this space we identify four extreme groups of grams three of which happen to be areal clusters at the same time: the “European” group of perfects, the “Indonesian” group of iamitives, the “Philippine” groups of iamitives, and forms meaning ‘already’. Between these extreme corners of grammatical space, a large number of other grams cannot be identified as clearly belonging to one of these four clusters. We therefore conclude that perfects and iamitives form a continuum in grammatical space without sharp boundaries anywhere. However, the two dimensions seem to represent two different directions of grammaticalization in which expressions for ‘already’ can expand: one which involves uses in natural development contexts with mainly stative predicates, and one which involves an increase mainly with dynamic predicates, causing iamitive grams to be more similar to European-style perfects. While the first one applies to both “Philippine” and “Indonesian” type iamitives, the second one only applies to the “Indonesian” type of iamitives. However, there are also a few functions that are mostly restricted to the “European” group of perfects, such as negated experientials and universal perfects, which are all strongest in the European group and are very weakly if at all represented elsewhere in the grammatical space considered. These are not associated with a “transition

to a new scene”, which means that they do not involve “current relevance”. Rather, they make a statement about the way the event type identified by the predicate occurs in an extended time span ending at reference time.

The major findings of this paper are summarized in Section 6.

2 Background: perfects and iamitives

In the majority of the world’s languages, there is more than one alternative for tense and aspect marking of sentences that refer to the past. Many languages have a distinction like the one found in English between a perfect (*I have sinned*) and a past (*I sinned*). Discussions of the semantics of perfects often quote the list of four meanings of the present perfect in English presented in Comrie (1976), which goes back to a similar list in McCawley (1971) (terms used in this paper are given in *boldface*):

1. **Perfect of result** (or resultative perfects) – “a present state is referred to as being the result of some past situation”
2. **Experiential** perfect (or existential perfects) – “a given situation has held at least once during some time in the past leading up to the present”
3. Perfect of persistent situation (*universal* or continuative perfects) – “a situation ... started in the past but continues (persists) into the present”
4. Perfect of recent past – “the present relevance of the situation is simply one of temporal closeness, i.e. the past situation is very recent”.

Over time, many different proposals have been advanced as to the semantics of perfects. There is no agreement on whether or not the readings listed above can be reduced to a general meaning or on how such a general meaning should be characterized. McCoard (1978) divides theories about the English Perfect into four groups: Current relevance theory, Indefinite past theory, Extended now theory, and Embedded past theory. Of these, only varieties of Current relevance theory and Extended now theory have enjoyed significant popularity in recent work. McCoard defines them as follows:

1. Current relevance theory (“CR theory”): the perfect expresses “the pastness of the event(s) embodied in the lexical verb, together with a certain applicability, pertinence, or relevance” to utterance time
2. Extended now theory (“XN theory”): the perfect marks prior events which are nevertheless included within the overall period of the present, the “extended now”.

We want to argue here that neither of these theories alone is sufficient to capture the semantics and pragmatics of perfects. Rather, different uses of perfects will involve

elements of both of them to varying degrees. Along the lines of earlier proposals by Sandström (1993) and Laca (2010) among others, we suggest that uses of perfects are characterized by one or both of the following characteristics or “profiles”:

1. The “transition to new scene” profile, which updates a representation of the world to a new one via the assertion that an event with non-trivial consequences has taken place.
2. The “extended time span” profile, in which an assertion is made about some time span ending at reference time.

The first profile, which is consonant with CR theory, is primarily characteristic of perfects of result whereas the second is characteristic of experiential perfects and perfects of persistent situation, but in many cases both profiles are relevant.

Perfects are characterized not only by where they are used but also by where they are not used. In Dahl (1985), it was suggested that there is a scale along which the probability of using a perfect gradually diminishes, at the end of which, one finds narrative discourse, which is typically temporally self-contained and maximally detached from the point of speech. Accordingly, Lindstedt (2000) proposes as a criterion on perfects that they are not used as narrative tenses.

There is a close affinity between the semantics of the perfect and that of English *already* and its translational equivalents in other languages, as was noted at least as early as by Traugott & Waterhouse (1969). Cf. the following:

- (1) It has become dark
- (2) It is already dark

Both (1) and (2) are typically used in a situation following upon a transition from light to darkness. They differ in that (1) uses a dynamic predicate (‘become dark’) while the predicate in (2) is stative (‘be dark’). But *already* can also be added to a dynamic predicate in the perfect or past, and in fact combinations of *already* and perfects are frequent:

- (3) It has already become/already became dark

Already also normally involves expectations of two kinds: (i) that the transition would occur at some point; (ii) that it would not take place as early as it did. These may also be present with perfects, but are more characteristic of *already*.

Given the similarities in use between *already* and perfects, it is perhaps not so surprising that we find constructions in languages which seem to combine

features of both. Thus, words like Indonesian *sudah* and Tok Pisin *pinis* are often translated as ‘already’ but are also sometimes said to be perfect markers. Some such cases were identified in Dahl (1985) as perfects on the basis of having a distribution similar to that of, for instance, the English Perfect, as in (4):

- (4) Indonesian
Dia sudah membaca buku ini.
he iam read book this
‘He has read this book’

It was also suggested in Dahl (1985) that markers meaning ‘already’ and ‘finish’ could be diachronic sources for perfects. Dahl & Velupillai (2005), referring to a proposal by Ebert (2001), mention the possibility that markers like Indonesian *sudah* should also synchronically be treated separately from “ordinary” perfects, an argument being that they still behave like English *already* in some respects as illustrated by the use of Indonesian *sudah* in (5):

- (5) Indonesian
Maria sudah ada di sini.
Maria IAM COP in here
‘Mary is already here’

Here, *sudah* has been added to a stative predicate ‘be here’. Like the English translation with *already*, the natural interpretation is that Mary is still here. But if we use the Perfect in English, with or without *already*, as in (6), it will be understood as meaning that she has left again.

- (6) Mary has (already) been here.

A number of grams which behave similarly to Indonesian *sudah* are treated in Olsson (2013) under the label “iamitives” (from Latin *iam* ‘already’), including Indonesian/Malay *sudah*, Thai *léw*, Vietnamese *đã* and *rôi* and Mandarin sentence-final *le*.

However, iamitives also tend to have another property illustrated in (7) from an Indonesian NT translation:

- (7) Indonesian
Saya sudah tua, dan istri saya juga sudah tua.
I IAM old and wife my also IAM old
‘(Zacharias said to the angel, “How can I be sure of this?) For I am an old man, and my wife is well advanced in years.”’ (ind-bah 42001018²)

² I.e. Luke 1:18.

Although (7) contains two occurrences of *sudah*, the English translation does not feature any perfects, nor does it contain any word like *already*. The predicate combining with *sudah* is in both cases *tua* ‘old’, which can be characterized as a “natural development predicate”, a predicate which becomes true sooner or later under normal circumstances. With such predicates, iamitives may become used systematically or even obligatorily, which can be interpreted as a case of grammaticalization. Olsson (2013, p. 18) gives the following example with the word *busuk* ‘rotten’:

- (8) Indonesian
 Kamu tidak bisa memakan-nya. Itu sudah busuk.
 You not can eat-it that IAM rotten.
 ‘You cannot eat it. It is rotten.’

Obviously, the natural development is slightly less determined by fate here than in the case of ‘old’, and even less so for a predicate such as *kawin* ‘married’. Yet, marriage may still be considered the normal case in many cultures; hence some Indonesian ID cards indicate the bearer’s civil status as *sudah kawin* ‘IAM married’.

We shall use the label ‘iamitive’ somewhat loosely for forms and constructions that (i) are used both with dynamic and stative predicates with a sense similar to that of English *already* and that (ii) show a tendency to be grammaticalized in natural development contexts. A major aim of our investigation is to see how iamitives, so defined, relate to perfects. We have already seen that their distributions overlap with those of “traditional” perfects but also that they include some areas that are not typical of the latter.

3 Grams and grammatical space

A GRAM (Bybee and Dahl, 1989) is a grammatical item in a particular language with specific form and specific meaning and/or function.³ A gram can have different formal exposures; it can be a functional word, a morphological marking (segmental or suprasegmental) or a complex construction. Examples are the English Perfect and Indonesian *sudah*. Cross-linguistically, grams tend to form clusters whose members are highly similar in meaning and behavior. Such a cluster can be seen as making up a GRAM TYPE. One of the aims of this paper is to explore whether perfects and iamitives can be subsumed under one gram type or not.

³ The notion of gram overlaps with the notion of *construction* in construction grammar. According to Croft (2001, p. 18) constructions are units with form and meaning linked by symbolic correspondence. This fits well with the notion of gram as language-specific grammatical item with specific form and specific meaning. However, not all constructions are grams.

Grams are not to be confounded with entities such as tense, aspect, gender, and case, traditionally referred to as “grammatical categories”, argued by Bybee & Dahl (1989) to be less relevant than grams for grammatical description. Grammatical categories are not neat superclasses for grams, which can already be seen from the fact that there is no agreement in the literature with respect to the classification of perfects as tenses and/or aspects.

Grams differ in how TRANSPARENT their form is (and hence in how easily they can be identified automatically in texts). Indonesian *sudah* is maximally transparent in being a single word with a constant shape. The English Perfect, consisting of the auxiliary *have* with its different forms and the past participle is less transparent. The German and French Perfects with their different auxiliaries *haben/sein* and *avoir/être* according to the lexical verb chosen are even less transparent. The low transparency of European perfects implies that it is not possible to identify their occurrence fully automatically in texts. In less transparent grams it is also often difficult to determine what should be counted as an occurrence of a gram. In European perfects, there is the problem of the pluperfect (formally a combination of the perfect with the past) which can either be considered a tense-aspect form of its own or an instance of perfect. Here we include pluperfects in perfects to make comparisons easier with languages which do not distinguish them.

The equivalence of a gram in the lexicon is the lexeme, which like grams, may involve a single morpheme or a complex construction. A lexeme type would be a cross-linguistic cluster of lexemes with highly similar meanings. Grams and lexemes are not strictly distinct, since the difference between lexical and grammatical meanings is gradual as some lexical items can gradually turn into grammatical items by grammaticalization. From the point of view of form there is a tendency to view transparent items as lexemes rather than as grams – so *sudah* is typically listed in dictionaries whereas the English Perfect is not – whereas at the level of types what matters is whether they express more grammatical or more lexical meanings. Grams of the same gram type can vary considerably in their formal properties (free vs. bound, simple morphemes vs. complex constructions, segmental vs. suprasegmental).

Gram types have more and less PROTOTYPICAL USES. The more likely a use is in a gram type, the more prototypical it is. Hence cross-linguistic comparison on the level of concrete language use is important for determining prototypical uses of a gram type.

Gram types are related to the notion of FUNCTIONAL DOMAIN in the sense that prototypical uses of a gram type often reflect a certain functional domain. According to Miestamo (2007: 293), a functional domain is “any

domain of related (semantic or pragmatic) functions that (one or more) language(s) encode with the formal means they possess”. However, taking into consideration also non-prototypical uses, a gram type can encode more than one functional domain.

Grams can be addressed both extensionally and intensionally. In a text or in a corpus a gram has a set of occurrences – its extension (distribution). But it also has a number of types of uses – approximately its “senses” or “functions” which is a statement at the intensional level. The advantage of using parallel texts is that this allows us to define grams extensionally and hence to establish gram types and prototypical uses of gram types without having to start with a priori intensional definitions. This makes it possible to test whether certain “senses” or “functions” which have been claimed to represent the core meaning of cross-linguistically similar grammatical items actually are what a gram type is dedicated to. Once a gram type is established, its prototypical uses can be interpreted intensionally as its “sense”. Put differently, whereas the functional domain approach goes from meaning to form (onomasiological), the gram type approach goes from form to meaning (semasiological).

Grams can be viewed as points in GRAMMATICAL SPACE. The closer two grams in grammatical space, the more similar their meanings. Grammatical space allows us to investigate the internal diversity of gram types (density of clusters of points) and the relationship between grams of different types (how neatly distinct clusters are). Grammatical space is similar to, but not identical with, semantic maps, which are sometimes also referred to as conceptual space. A semantic map is a graph or a multidimensional space where the basic units (nodes or points) are semantic or pragmatic functions whose similarity is reflected by their distance in the graph or space (Haspelmath, 2003). Semantic maps modelled as multidimensional space can be computed by Multidimensional Scaling (Croft & Poole, 2008); Wälchli & Cysouw, 2012). For any pair of nodes or points it holds that the closer two functions, the more similar their meaning. Since cross-linguistically recurrent identity of form reflects similarity in meaning (Isomorphism Hypothesis; Haiman, (1985: 19) semantic maps can be constructed by arranging its points such that all grammatical categories (grams) considered are represented as compactly as possible (as subgraphs of the graph or as contiguous areas of the space). Grams are thus represented as subgraphs or zones in a space, which is convenient if the extensions of different grams in the same language or a small number of grams in different languages are to be depicted. However, semantic maps where many similar grams of different languages are depicted are difficult to read.

If many similar grams from different languages are to be compared it is more convenient to represent each gram as a point in grammatical space. (An alternative and more general name for it would be “categorical space” since any type of category, both grams and lexemes, can be investigated.) Grammatical space is a representation of grams or other linguistic categories of different languages, each visualized by a dot, where the distances between all dots reflects the dissimilarity between categories. Like semantic maps, grammatical space can be visualized by Multidimensional Scaling (MDS).

A major advantage of the semantic map approach is that the multiplicity of uses of a grammatical category can be acknowledged without having to claim that the different uses are entirely different meanings. Put differently, no decision has to be made between a polysemist and a monosemist position (Haspelmath, 2003: 213). Grammatical space has a similar advantage when considering the relationship between grams. Even though grams cluster to gram types the borders between related gram types are not always clear. Grammatical space allows us to acknowledge the diversity of similar grams without having to claim that they belong to entirely different gram types. In grammatical space we can view two related gram types, such as perfects and iamitives, as a continuum.

4 Method and data

The investigation involves the following major steps:

- creating a sample of relevant grams;
- creating a representation of the sample reflecting the distances (degrees of general dissimilarity) between the distributions of the grams in the respective translations;
- identifying and mapping the specific factors behind the similarities and differences in distribution.

4.1 *The Parallel Bible Corpus*

We use parallel texts to identify gram types, which – as pointed out above – has the advantage that we can define grams extensionally and hence establish gram types and prototypical uses of gram types without having to start with a priori intensional definitions. Dahl (1985) used a questionnaire of about 200 sentences in context and the work presented here can be seen as continuing this approach, using a different kind of data – parallel corpora. Like translation questionnaires, these contain sets of translationally equivalent sentences or texts but are less governed by the investigator’s preconceived notions of what distinctions will be relevant.

The corpus we use consists of translations of the New Testament (NT) in digitalized form, in most cases available on the Internet. It contains 1267 translations representing 1107 different ISO 639-3 language codes (about 15 per cent of the world's languages). It should be kept in mind, however, that the choice of languages depends on the availability of Bible translations, leading to a rather patchy coverage of the languages of the world, as shown in Figure 1. A typical NT translation contains around 200,000 words.

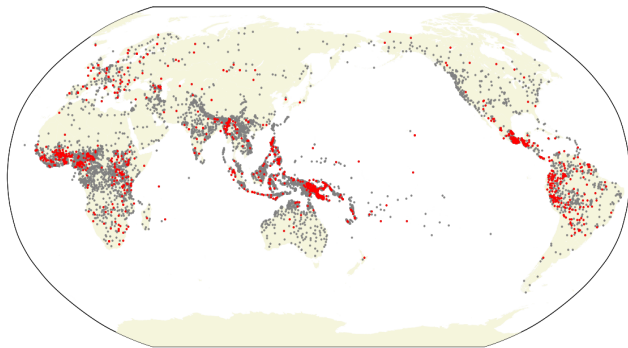


Figure 1. The languages represented in the corpus (red) among the languages of the world (grey).

No method is without its drawbacks. It is easy to see that the use of parallel corpora presupposes that the texts are “parallel” enough semantically and structurally – in other words, the quality and faithfulness of the translations, as well as the approach of the translators, are crucial (see de Vries (2007) for a discussion)⁴. In addition, we usually do not know how well a translation represents the language of the community for which it was made. Still, translations are samples of how humans use language, and at least as long as we see them as representing themselves (i.e. particular doculects of a language rather than the language they are written in in its entirety) they are valid objects of study. Any generalization to other language varieties, however, has to be treated with great caution.

The use of parallel corpora relies on the possibility of ALIGNING the texts. One advantage of Bible texts is that they are usually organized in chapters and verses in a strongly uniform fashion. Using verse-level alignments in comparing the distribution of elements in the corpus texts means that “having the same distribution” will be interpreted as “occurring in the same verses”. When dealing with high frequency grams which can be expected to occur at several different places in the same verse, this will often yield a low PRECISION (many false positives) of

the linking but a good RECALL (few cases missed). Related to the precision problem is the fact that if a gram is frequent enough, it may appear in more than half the verses in an NT translation, which makes it rather difficult to draw any conclusions about any relationship in meaning with other high frequency grams.

We have been fortunate in being able to exploit the results of the work of Robert Östling, reported in his doctoral thesis (Östling, 2015), where he explores Bayesian models for word alignment of parallel corpora. As an outcome of his research, we now have a version of the Bible corpus with word class annotations and links at the word level from all the texts to dependency parsed versions of translations into several European languages. This enables us to make more precise alignments of the text. Instead of aligning at the level of verses, we now divide the verses into what we call “segments”. The idea of a segment is a set of words in the text (not necessarily continuous) which are all syntactic dependents of the same verb. The starting point is a parsed English text (the Lexham English Bible); after marking up its “segments” we transfer the markings to other texts: a segment in a target text will consist of all words that are linked to words in one and the same segment of the source text. This annotation is enhanced in two ways – it is assumed that an unlinked word which stands between two words linked to the same segment or between one linked word and a punctuation mark belongs to the segment the links point to. Furthermore, the links are enhanced by using indirect links via parsed versions in the other languages. The results vary widely, depending on factors such as the similarity between the languages involved and the complexity of the structure of words in the target language. This obviously creates a bias towards the languages where the linking results are better.

4.2 Finding a cluster of grams

As described above, a gram type can be seen as a cluster of grams represented as points in grammatical space where the distance between two grams depends on the degree of similarity between their distributions. The task is thus to identify such clusters.

The first step in the procedure is to choose a set of “seed grams”, that is, a set of grams with known distributions that based on our prior knowledge can be assumed to be members of the same cluster. We calculate their GENERALIZED DISTRIBUTION, that is, the probability for a member of the set to be represented in each location in the corpus. This generalized distribution is the point of departure for the search for other grams which are similar enough to be included in the cluster. To perform such a search, a similarity measure is necessary – more

⁴ Some consolation can be found in the fact that comparing texts that are too different to be called “parallel” will usually not yield the wrong results but rather no results at all.

about the choice of such a measure below. In each text that is included in the corpus, we look for expressions which have a similarity to the generalized distribution that exceeds a certain predefined threshold.

The result of the search will then be a new set of candidate grams to serve as the basis for calculating a new generalized distribution. This distribution, which is best defined on a balanced subset of the candidate grams in order to minimize areal and genetic bias, can in its turn be used for a further search. The procedure can be iterated as many times as desired. In the course of the process, the generalized distribution tends to lose any more idiosyncratic features that characterized the set of seed grams. It should also be noted that there is no guarantee that all the seed grams will survive the iterated search procedure: it may well happen that some of them were not as similar to the rest as assumed.

The above is a somewhat idealized description of the actual procedure. One complication mentioned above is that grams are not always transparent. We will sometimes have to use our knowledge about the languages involved to make more or less educated guesses to find the candidate grams.

How is similarity between grams measured? There are many different statistical measures that could be applied, and we have tried several of them, such as Jaccard and Dice indices, Pearson’s *r*, T-score as defined by Fung and Church (1994), pointwise mutual information and chi-square. These methods are useful especially if starting out with a single gram, but they share the drawback that they tend to be influenced by size in exaggerating the similarities between high-frequency patterns while not being able to pick out patterns with a lower frequency. A related consideration is that a single measure will not be able to distinguish recall and precision. We should therefore look for at least one positive and one negative criterion.

What we do is to exploit sets of locations where the members of a cluster tend to behave in a uniform fashion. “Locations” here being equivalent to segments, this means that we will be looking at the segments with a maximal tendency to be linked across translations – which means that but they represent passages which tend to be translated without “noise”, that is, which tend not to be paraphrased and which do not involve ambiguities in the source text. Within those, we look for locations where the probability that a member will be represented is close either to one or to zero. The segments with maximal linking and maximal recall – where close to 100 per cent of the members of a cluster appear – will then include only prototypical uses of the gram type, making it possible to formulate a positive criterion for inclusion in the cluster. Our experience is that a threshold of 65 per

cent of the 20 top segments is enough to give plausible results. Surprisingly often, we find that grams are used in 95 or 100 per cent of the top segments.

Consider now the negative criterion. The gram that we are testing should not have too many occurrences that are outside the generalized distribution of the cluster. One problem is that the generalized distribution has a “long tail” of segments where a non-zero but very low number of members of the gram type appear. We therefore interpret “outside the generalized distribution” as “outside the set *S* of segments where at least two members of the original set of grams appear”. The crucial question is now how many occurrences a gram has in the complement set of *S*. We define the threshold as 3 per cent of the size of the complement set, which in practice means a maximum of around 1200 occurrences. This may not be the ideal solution, but for our purposes it has given reasonable results so far.

4.3 Creating a working sample

In order to be able to study the variation among relevant grams using Multi-dimensional Scaling and other techniques, we need to create what can be called a working sample of grams. Since one of our aims is to get a better understanding of the relationship between perfects and iamitives, it seems motivated to start out with seed samples of maximally uncontroversial examples of each. For traditional perfects, we choose constructions from two translations each of five European languages: English, Estonian, Finnish, Swedish, and Spanish, representing the Germanic and Romance branches of Indo-European and the Finnic branch of Uralic. They were all identified in Dahl (1985) as being clear examples of the proposed cross-linguistic category (gram type) PERFECT. In addition, they are all alike in using a single auxiliary – ‘have’ in the Indo-European languages and the copula in the Uralic ones. For the iamitive seed sample, we choose ten markers that are discussed as examples of iamitives in Olsson (2013):

Table 1. Iamitive seed grams (following Olsson, 2013)

Language	ISO	Family	Iamitive marker
Central Khmer	khm	Austro-Asiatic	<i>ban</i>
Indonesian	ind	Austronesian	<i>sudah</i>
Lao	lao	Tai-Kadai	<i>laev</i>
Malay	zlm	Austronesian	<i>sudah</i>
Mandarin Chinese	cmn	Sino-Tibetan	sentence-final <i>le</i>
Paraguayan Guaraní	gug	Tupian	<i>-ma</i>
Thai	tha	Tai-Kadai	<i>lǎew</i>
Vietnamese	vie	Austro-Asiatic	<i>rôi</i>
Vietnamese	vie	Austro-Asiatic	<i>đã</i>
Yoruba	yor	Niger-Congo	<i>ti</i>

We now run a number of iterations of the searches based on the generalized distributions of these two gram sets – henceforth referred to as the “E gram set” and the “I gram set”. Each new generalized distribution is constructed on a balanced sample of the resulting grams of the previous iteration the “best” example from each genus (defined as in WALS). In the end, we construct a joint sample, with no more than one representative of each language with a small number of exceptions where two non-cognate grams from one language are included.

5 Results

As the searches are iterated, the number of found grams increases, but we can also observe that the E and the I gram sets converge. Table 2 shows what happens:

Table 2. Convergence of perfect and iamitive gram samples in iteration

Iteration	E gram set	I gram set	Intersection	Union
0	10	10	0	14
1	25	161	11	178
2	59	295	40	314
3	139	308	98	349
4	214	307	153	368

Already in the first round, 11 grams fulfill the criteria for both purported gram types. Although the expansion of the E gram set is slower than that of the I gram set, the 25 grams obtained in the first iteration of the E gram

set include Vietnamese *dã* which was also a member of original I gram set. The set of grams that are included in both sets grows with each iteration. After four iterations, the two gram sets share about 40 per cent of their members. The relationship is asymmetric: only about 30 per cent of the E gram set is not included in the I gram set, whereas 50 per cent of the latter set are not included in the former.

We find convergence also in the generalized distributions of the gram sets. The top 20 lists of the original gram sets share only two segments. This figure rises to 11 in the fourth iteration. Since the later iterations mainly result in an increased overlap between the gram sets, we base our final sample on the second iteration. After removing some redundant items, the number of grams is 305. They represent 302 ISO-codes and 48 families.

Figure 2 shows an MDS plotting of the two original gram sets.

While the members of the E sample all cluster on the right edge of the diagram, the members of the I sample are more spread out, forming three more or less well-defined clusters. The distance between the leftmost and the rightmost members of the I sample is considerably greater than that between the latter and the members of the E sample. In fact grams from as many as four of the languages represented in the original E sample – Lao, Mandarin Chinese⁵, Thai, and Yoruba – do not survive the iteration process and are not found in the final sample in spite fulfilling the positive criteria for inclusion. The reason is that these grams also have uses that are not found in other perfects and iamitives and therefore have too many occurrences outside the generalized distribution.

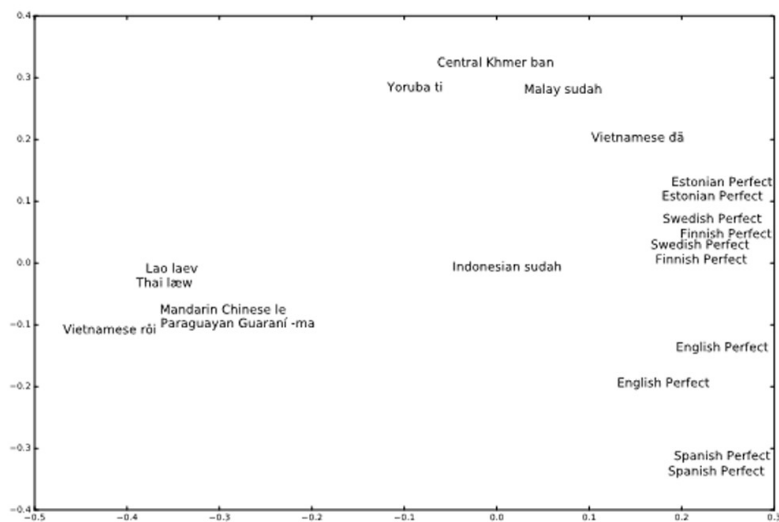


Figure 2. MDS plotting of the united perfect and iamitive seed gram sets

⁵ Instead of sentence-final *le*, which would appear to be the prime candidate for an iamitive, Mandarin *yijing*, usually translated as ‘already’, meets the criteria for inclusion in the sample. The relationship between these two words deserves a special study.

The final sample is too large for it to be possible to show the individual language codes in a diagram (but see supplementary materials). Instead we show below each gram by a symbol with its size reflecting the frequency of that gram in the NT text. The frequency of the grams, tends to grow as we move away from the upper left corner in the diagram. A closer look reveals that the cluster of low-frequency grams in that corner mainly represents markers such as Portuguese *já* or Russian *уже*, that is words corresponding to English *already* but with a higher frequency, suggesting an initial grammaticalization in the direction towards iamitives. Forms similar to English *already* do not make it to the iamitive sample. The more grammaticalized iamitives like the ones in the initial seed sample are usually found in the lower half of the MDS diagram. Perfects from other historical sources than ‘already’ and ‘finish’ – including an outlier group of European perfects but also quite a few others – tend to be located in the upper right quadrant.

Some genealogical and/or areal groupings come out as relatively well delineated here. Figure 3 shows three such groupings, which happen to be located at three different corners of the diagram: (i) the “Philippine” group in red, comprising grams from six Greater Central Philippine languages within Austronesian⁶; (ii) the “Indonesian” group in green, comprising grams from nine Austronesian

languages spoken in Indonesia⁷; (iii) the “European” group in blue: grams from eleven languages spoken in Europe from the Germanic and Romance branches of Indo-European and the Finnic branch of Uralic spoken in Europe⁸. These three groups will provide us with a suitable point of departure for discussing the variation within the final sample. Our main approach in studying this variation will be to try and identify sets of segments whose propensity of occurrence with grams in the sample correlates with the dimensions of the MDS analysis.

Both dimensions of the diagram are correlated with frequency. In particular, the horizontal dimension (the *x*-axis), representing MDS Dimension 1, has a high correlation with the frequency of the grams (Pearson’s *r*: 0.81): The (negative) correlation between frequency and Dimension 2 (the *y*-axis) is much lower – -0.22. This low figure is deceptive, however, since the grams in the upper half of the diagram are of two radically different types: to the left, we find “pre-iamitives” in an early stage of grammaticalization, to the right, European-style perfects. We obtain a much higher figure – -0.72 – if we remove the grams in the upper right quadrant from the calculation.

The primary question is now how to account for the differences in frequency that we find in the sample. The answer may be a key to understanding the diachronic processes behind the evolution of perfects and iamitives.

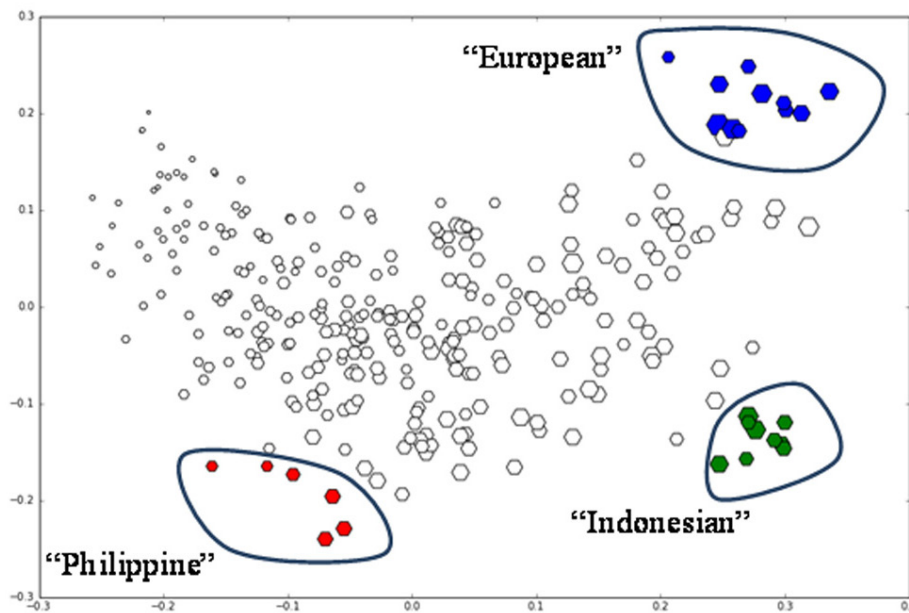


Figure 3. “Philippine” iamitives (red), “Indonesian” iamitives (green) and European perfects (blue) as extreme clusters in the grammatical space of perfects and iamitives (Dimension 1 – horizontal, Dimension 2 – vertical)

⁶ Cebuano [ceb] *na*, Hiligaynon [hil] *na*, Kinaray-A [krj] *run*, Paranan [prf] *dán*, Tagabawa [bgs] *dán*, Waray [war] *na*.

⁷ Achinese [ace] *ka*, Indonesian [ind] *sudah*, Lampung Api [ljp] *radu*, Ma’anyan [mby] *haut*, Minangkabau [min] *lah*, Ngaju [nij] *jari*, Ot Danum [otd] *jari*, Sangir [sxn] *séng*, Sasak [sas] *sampun*.

⁸ Perfects in Catalan [cat], Danish [dan], English [eng], Estonian [est], Faroese [fao], Finnish [fin], German [deu], Icelandic [isl], Norwegian Bokmål [nob], Spanish [spa], Swedish [swe].

We will look first at Dimension 2. For any segment s in the NT text, it is possible to compute the correlation between the distribution of grams in the sample that appear in s and each of the dimensions in the MDS analysis. We have identified the ten segments with the highest (negative) correlation with Dimension 2. For reasons of space, we quote only three representative examples here (see supplementary materials for a full list):

- (9) Examples of top segments on Dimension 2
- Now the Passover, the feast of the Jews, was near.* (43006004)
 - And not being weak in faith, he considered his own body as good as dead, *because he was approximately a hundred years old*, and the deadness of Sarah's womb. (45004019)
 - You also be patient. Strengthen your hearts, *because the coming of the Lord is near.* (59005008)

The correlations range from -0.51 to -0.62. Taken together, the segments have a correlation of -0.81 with Dimension 2. The distribution in the sample is displayed in Figure 4. As is predicted by the negative correlation with Dimension 2, the highest frequencies are found in the lower half of the diagram.

Thus, the grams of both the “Philippine” and the “Indonesian” groups are used nearly across the board in these contexts, while the grams in the European group are hardly used at all.

The majority of the segments chosen contain the predicate ‘be near’, mainly in the temporal sense. Another concerns the age of a person. These are typical cases of stative predicates denoting what was above called “natural developments”. Their distribution in the sample is consistent with the hypothesis that the use of iamitive grams in these contexts becomes close to universal as these grams are grammaticalized. We see an almost categorical divide between the two Austronesian groups on the one hand and the European group on the other. We illustrate here with an example from Javanese, another Austronesian language:⁹

- (10) Javanese
 ...Putra-ning Manungsa wis mèh rawuh.
 child-poss human.being IAM almost come/arrive
 ‘... the Son of Man is near’ (40024033)

Furthermore, the less frequent grams, concentrated in the upper left part of the diagram, also have relatively low incidences here, suggesting that they have not yet

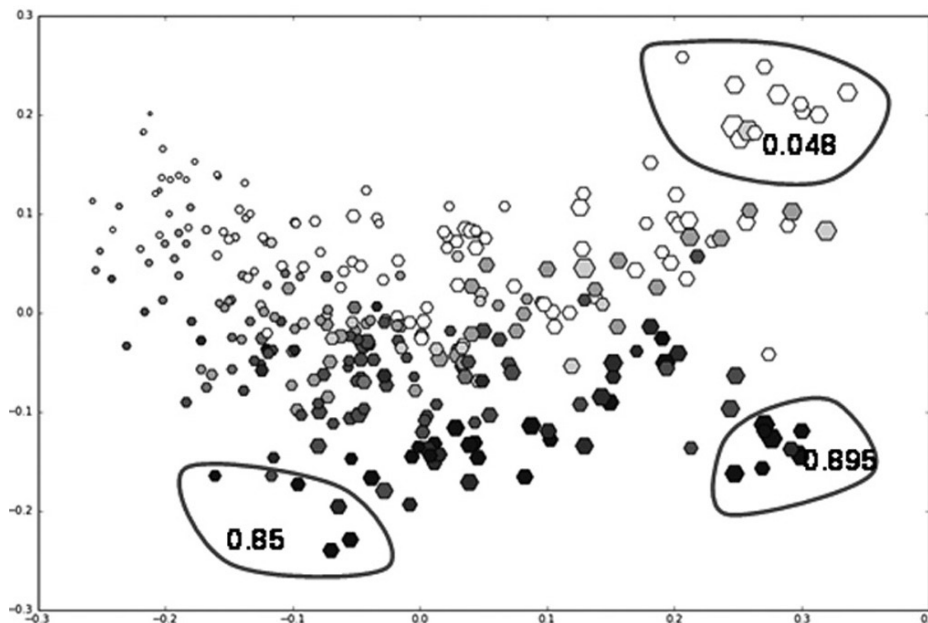


Figure 4. Correlation with the 10 top segments of dimension two (darkness)

⁹ Vander Klok & Matthewson (2015) discuss the particle *wis* in Javanese, which has earlier been treated variously as a “already, a (present) perfect, a past tense, or a perfective.” They note correctly that the meanings of *already* and perfects overlap, arguing that *wis* “is best analyzed as a marker expressing already”. Referring to Olsson (2013), they also suggest that “a new category of iamitives is not

warranted, as many of the properties Olsson discusses can be naturally explained under a focus-sensitive semantic analysis of *already*”. However, what this does not explain is that Javanese *wis* is about 20 times as frequent as English *already* in all NT translations, and that it, unlike *already*, is highly frequent in natural development contexts such as (10).

undergone the relevant grammaticalization. However, we can also observe a fairly smooth increase between the more extreme groups as we go down the vertical dimension. In other words, there is a continuum between two poles rather than a sharp division into iamitives and non-iamitives. While it seems clear that the grammaticalization in natural development contexts is of primary importance, we have been able to identify at least another factor that is behind the high frequency of some of the grams in the lower left half of the MDS diagram, including but not restricted to the “Philippine” group. Some iamitive markers combine with negatives to denote ‘no longer’ or ‘not yet’ and are used also for future time reference, as in the following Cebuano example, where the verb form ‘Unreal Potential’ is combined with the iamitive *na*:

- (11) Cebuano
 ...dili na kamo maka-kita kanako
 NEG IAM you IRR.POT-see I
 ‘...you will see me no more,’ (ceb-rpv 43016010)

Such combinations should probably be seen as constructions in their own right in a similar way as pluperfects.

For Dimension 1, one thing is clear: since European-style perfects cluster at the right edge, a gram that moves from left to right will become more “perfect-like”. It is harder to specify exactly what that change consists in. Since perfects and words meaning ‘already’ share uses with dynamic predicates, but differ in how they are used with stative predicates, a convergence could be expected to involve an increase of the former. So far, we have not

been able to find anything that contradicts the idea of a general increase of the use with dynamic predicates, with one notable exception, which looks syntactic rather than semantic, viz. the frequency in relative clauses, shown in Figure 5. A calculation based on the frequency of grams in the sample with clauses marked as relative in the parsed Lexham translation gives a figure as high as 0.91. The high correlation remains (the value is 0.89) if we restrict the comparison to the lower half of the diagram (essentially the highly grammaticalized iamitives). It may be noted that there is no significant correlation between frequency in relative clauses and Dimension 2.

It is important to verify that this increase is not just an effect of the general increase of frequency across Dimension 1. To that end, we have also computed the correlation between that dimension and the ‘weight’ of relative clauses in the distribution of grams (that is, the ratio of occurrences in relative clauses and the total frequency of a gram). It is also high – 0.65 – which shows that the frequency of grams in the sample grows faster in relative clauses than in the translations in general. The value is even higher in the lower half of the MDS diagram (0.85). The correlations are also reflected in the differences between the three areal groups, as seen in the diagram. Another question is whether the findings about relative clauses reflect a tendency in subordinate clauses in general. A test on clauses marked as adverbial in the parsed English version also shows a high incidence correlation with Dimension 1, but the weight correlation is even slightly negative (-0.06), meaning that there is no connection that is independent of the total frequency of grams. In other words, what we have found for relative

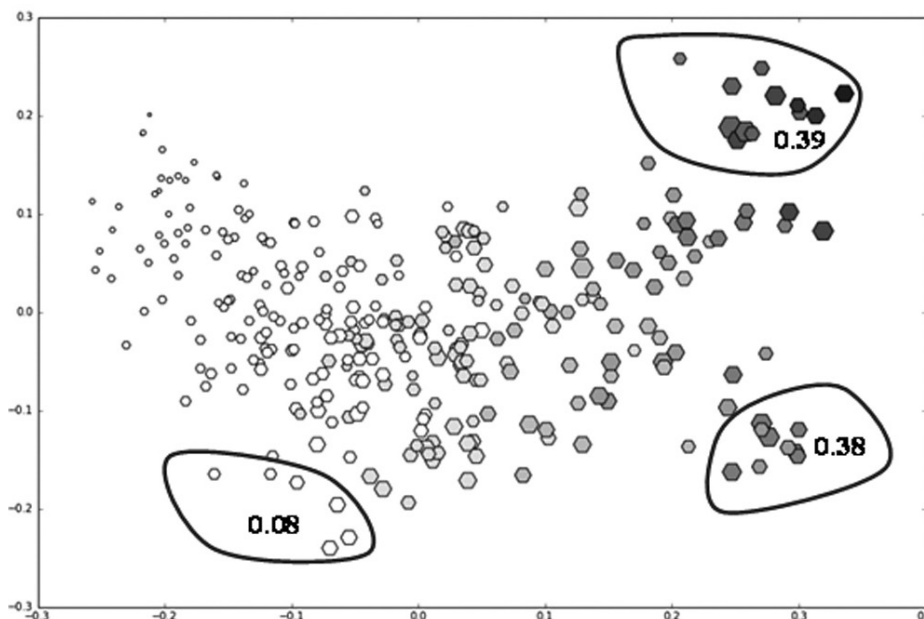


Figure 5. Correlation with relative clauses (Lexham English Bible) (darkness)

clauses cannot be generalized to other types of subordinate clauses. Summing up what we have seen so far, the two dimensions in the diagram seem to represent two different ways in which iamitives can expand: one which involves uses in natural development contexts with mainly stative predicates, and one which involves an increase mainly with dynamic predicates, causing the gram to be more similar to European-style perfects. While the first one applies to all the grams in the lower half of the MDS diagram, the second applies to them to varying degrees, with the “Philippine” and “Indonesian” groups each at one end of the scale.

A number of uses seem to be mainly restricted to the upper right quadrant of the MDS diagram. One of them is related to the experiential use of perfects. It does not seem possible to generalize over all the cases that are usually classified as experientials in the literature. Combinations of perfects with verbs such as ‘see’, ‘sin’, and ‘work’, whose meanings do not entail a specific result state, are often thought of as experiential, but although it is true that their frequency is relatively highly correlated with Dimension 1, they do not differ significantly from other dynamic predicates in this regard and the expansion cannot be seen as independent of the general increase in frequency. But there is a subset of experiential uses that stand out here, viz. the ones appearing under the scope of negation. One way of identifying those is by looking at the distribution of dedicated experiential grams, many of which can be seen as negative polarity items. Figure 6 is based on the incidence of grams in the sample in a set of 20 segments where the dedicated Indonesian experiential marker *pernah* (Dahl, 1985, p. 160) is used in at least 9 out of 10 Indonesian translations. It can be seen that it is

mainly in the European group that members of the sample show up in the negated experiential contexts.

In Section 1, we noted that a major difference between traditional perfects and iamitives is that the latter can be used with stative predicates with present time reference. In order to see what happens in cases where English uses the perfect with a stative predicate, we looked at a set of 42 segments where English translations use the verb *love* with a past or perfect, as in:

- (12) Just as the Father has loved me, I also have loved you. (43015009)

The distribution, shown in Figure 7, is fairly similar to what we saw for the negated experientials.

Due to a certain fuzziness of the border between experiential and universal perfects, examples of the latter are not so easy to identify. We looked at 11 relatively unequivocal cases, such as those in (13) (see supplementary materials for a full list).

- (13) Examples of universal perfects

- a) *Therefore I have experienced help from God until this day*, and I stand here testifying to both small and great, saying nothing except what both the prophets and Moses have said were going to happen, (44026022)
- b) By teaching these things to the brothers, you will be a good servant of Christ Jesus, trained in the words of the faith and of the good teaching *that you have followed faithfully*. (54004006)
- c) And looking intently at the Sanhedrin, Paul said, *“Men and brothers, I have lived my life in all good conscience before God to this day.”* (44023001)

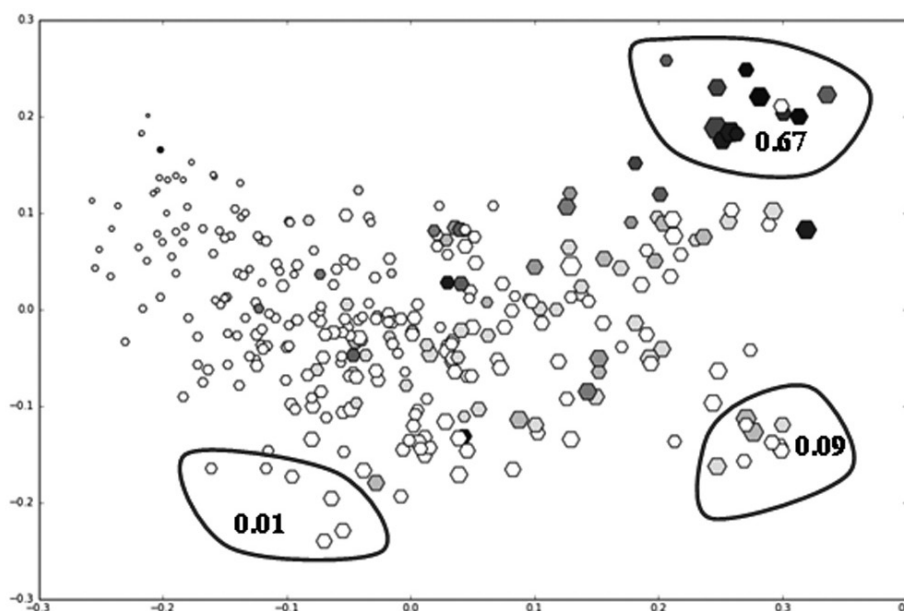


Figure 6. Correlation with negative experiential (Indonesian *pernah*)

Figure 8 shows the distribution of the grams in the sample in the segments in (13) – once more with a concentration to the upper right quadrant.

The three last context types we have looked at, negated experientials, ‘love’ with past time reference, and universal perfects, are all strongest in the European outlier group and are very weakly if at all represented in the left half of the diagram. They may be argued to have something in common: they are not associated with a “transition to a new scene”, which means that they do not involve “current relevance” in the sense of saying something about the state of the world at reference time

(which usually coincides with utterance time). Rather, they make a statement about the way the event type identified by the predicate occurs in an “extended time span” ending at reference time. This can be seen as extensions from the core uses of perfects – extensions which seems to come less easily to what we have called ‘iamitives’, that is, grammaticalizations of expressions originally meaning ‘already’ and ‘finish’. It should be noted that “transition to a new scene” can be seen as an essential part of the meaning of words like *already* (cf. also Ebert’s (2001) label “NEWSIT” – short for “new situation” – for iamitives).

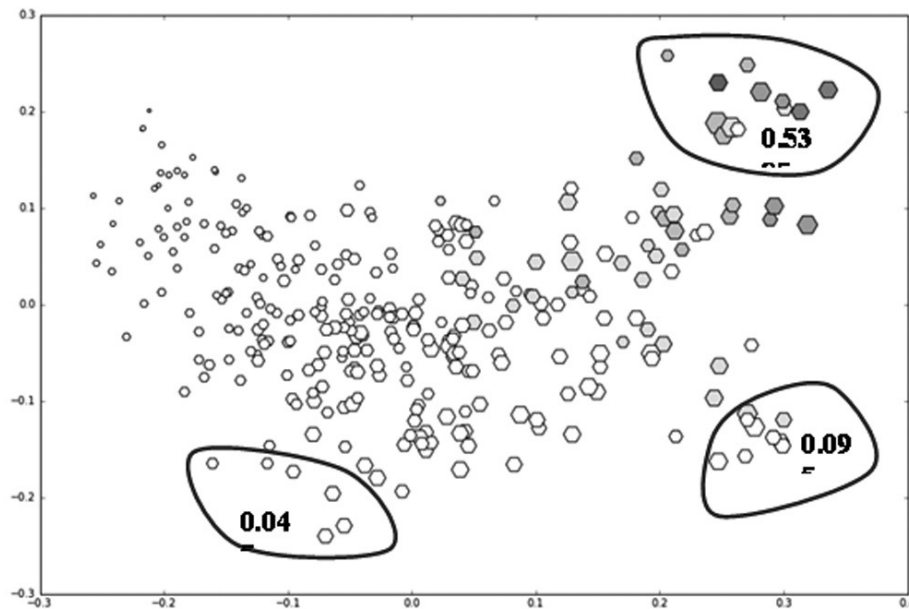


Figure 7. Correlation with English ‘love’ in past or perfect

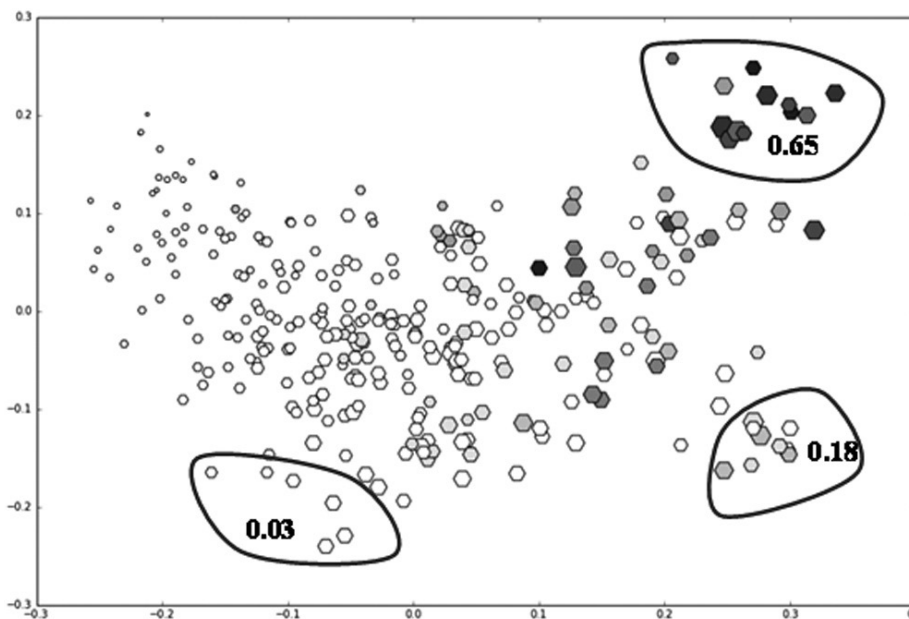


Figure 8. Correlation with universal perfect

6 Conclusions

Perfects are well-known elements of the tense and aspect systems of many European languages. Somewhat less well-known are the “iamitives” particularly common to languages in South East Asia. In the literature, these have often been seen as varieties of perfects, but there have also been proposals to treat them as a separate category – in the terminology used here, as a separate gram type. In this paper, we have tried to clarify the relationship between iamitives and traditional perfects by using quantitative methods on a large parallel corpus.

In our terminology, a gram type is a cluster of language-specific grams whose closeness in meanings and functions is reflected in similar distributions in a parallel corpus. We have assumed that a gram type comes with a set of prototypical uses shared by all members of the gram type modulo the unavoidable noise due to inexactitudes in usage and in translations.

We identified gram types starting from one or more seed grams, trying to find other grams with similar distributions and defining a generalized distribution for them. We then explored what happens if we try to construct gram types departing from two different sets of seed grams: on the one hand, a set of constructions from European languages, traditionally identified as perfects, on the other, the set of grams subsumed under the heading of iamitives in Olsson (2013). What we found was that already in the first round of the search process the set of grams meeting the criteria for each of the purported gram types overlapped quite considerably, and the overlap increased for each iteration of the process, although the sets of grams never coincided totally. The sets of prototypical uses also converged, although at a much slower rate, remaining largely separate. The conclusion is that while perfects and iamitives can be argued to be separate at the gram type level, a significant part of their members cannot be identified as belonging only to one of them. In terms of grammatical space, the two types occupy overlapping regions without sharp boundaries anywhere.

A more general conclusion is that although grammatical items clearly form clusters in grammatical space, these clusters are neither homogeneous internally nor discrete externally, continua between them being a normal phenomenon.

References

- BYBEE, Joan L.; DAHL, Östen. The Creation of Tense and Aspect Systems in the Languages of the World. *Studies in Language*, v. 13, n. 1, p. 51-103, 1989.
- COMRIE, Bernard. *Aspect: an introduction to the study of verbal aspect and related problems*. New York: Cambridge University Press, 1976.
- CROFT, William. *Radical construction grammar: syntactic theory in typological perspective*. New York: Oxford University Press, 2001.
- CROFT, William; POOLE, Keith T. Inferring universals from grammatical variation: Multidimensional Scaling for typological analysis. *Theoretical Linguistics*, v. 34, n. 1, p. 1-37, 2008. Available at: <[doi:10.1515/THLI.2008.001](https://doi.org/10.1515/THLI.2008.001)> Access on: 25 abr. 2016.
- DAHL, Östen. *Tense and aspect systems*. Oxford: Blackwell, 1985.
- DAHL, Östen; VELUPILLAI, Viveka. The Perfect. In: COMRIE, Bernard; DRYER, Matthew; GIL, David; HASPELMATH, Martin (ed.). *World Atlas of Language Structures*. Oxford: Oxford University Press, 2005. p. 271-272, p. 280-281. Available at: <<http://www.livingreviews.org/wals/feature/68>>. Access on: 25 abr. 2016.
- EBERT, Karen. Tense-aspect flip-flop and a somewhat elusive gram type. In: EBERT, Karen H; ZÚÑIGA, Fernando (ed.). *Aktionsart and Aspectotemporality in Non-European Languages*. Arbeiten des Seminars für Allgemeine Sprachwissenschaft 16. Zürich: Universität Zürich, Seminar für Allgemeine Sprachwissenschaft, 2001. p. 141-158.
- FUNG, Pascale; CHURCH, Kenneth Ward. K-vec: A new approach for aligning parallel texts. *Proceedings of the 15th conference on Computational linguistics*, v. 2, p. 1096-1102. Association for Computational Linguistics, 1994. Available at: <<http://dl.acm.org/citation.cfm?id=991328>>. Access on: 26 fev. 2016.
- HAIMAN, John. *Natural syntax: iconicity and erosion*. New York: Cambridge University Press, 1985.
- HASPELMATH, Martin. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. *The new psychology of language*, v. 2, p. 211-242, 2003.
- LACA, Brenda. Perfect semantics. How universal are Ibero-American Present Perfects? *Selected Proceedings of the 12th Hispanic Linguistics Symposium*, 2010. p. 1-16. Available at: <<http://www.lingref.com/cpp/hls/12/paper2401.pdf>>. Access on: 23 set. 2015.
- LINDSTEDT, Jouko. The perfect – aspectual, temporal and evidential. In: DAHL, Östen (ed.). *Tense and aspect in the languages of Europe*. Berlin: de Gruyter, 2000. p. 365-384.
- MCCAWLEY, James D. Tense and time reference in English. In: FILLMORE, Charles J.; LANGENDOEN, D. Terence (ed.). *Studies in Linguistic Semantics*, Irvington, 1971. p. 96-113.
- MCCOARD, Robert W. *The English perfect: tense-choice and pragmatic inferences*. Amsterdam: North-Holland, 1978.
- MIESTAMO, Matti. 2007. Symmetric and asymmetric encoding of functional domains, with remarks on typological markedness. In: MIESTAMO, Matti; WÄLCHLI, Bernhard (eds.). *New Challenges in Typology: Broadening the Horizons and Redefining the Foundations*. Berlin: de Gruyter, 2007. p. 293-314.
- OLSSON, Bruno. *Iamitives: Perfects in Southeast Asia and beyond*. Stockholm: Department of Linguistics, Stockholm University Masters thesis, 2013. Available at: <<http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-91392>>. Access on: 13 fev. 2015.

ÖSTLING, Robert. 2015. Bayesian Models for Multilingual Word Alignment, 2015. Available at: <<http://www.diva-portal.org/smash/record.jsf?pid=diva2:798117>>. Access on: 23 jan. 2016.

SANDSTRÖM, Görel. *When-clauses and the temporal interpretation of narrative discourse*. Umeå: Univ., Dept. of General Linguistics, 1993.

TRAUGOTT, Elizabeth Closs; WATERHOUSE, John. “Already” and “yet”: a suppletive set of aspect-markers? *Journal of Linguistics*, v. 5, n. 02, p. 287-304, 1969. Available at: <[doi:10.1017/S0022226700002279](https://doi.org/10.1017/S0022226700002279)>. Access on: 25 abr. 2016.

VANDER KLOK, Jozina; MATTHEWSON, Lisa. Distinguishing already from Perfect Aspect: A Case Study of Javanese wis. *Oceanic Linguistics*, v. 54, n. 1, p. 172-205, 2015. Available at: <[doi:10.1353/ol.2015.0007](https://doi.org/10.1353/ol.2015.0007)>. Access on: 25 abr. 2016.

VRIES, Lourens de. Some remarks on the use of Bible translations as parallel texts in linguistic research. *Language Typology and Universals*, v. 60, n. 2, p. 148-157, 2007.

WÄLCHLI, Bernhard; CYSOUW, Michael. Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics*, v. 50, n. 3. p. 671-710, 2012. Available at: <[doi:10.1515/ling-2012-0021](https://doi.org/10.1515/ling-2012-0021)>. Access on: 26 feb. 2016.

Quotations

If not marked otherwise, Bible quotations are from the *Lexham English Bible (LEB)*. Copyright 2012 Logos Bible Software. Lexham is a registered trademark of Logos Bible Software.

Recebido: 03 de fevereiro de 2016.

Aprovado: 02 de junho de 2016.

Contato: oesten@ling.su.se
bernhard@ling.su.se

SUPPLEMENTARY MATERIALS

The sample

	Freq	Dim1	Dim2
Upper left quadrant			
Afro-Asiatic			
East Chadic			
Kera [ker] <i>ne</i>	934	-0.09	0.05
Masa			
Marba [mpg] <i>wa</i>	429	-0.13	0.04
Algic			
Algonquian			
Severn Ojibwa [ojs] <i>aasha</i>	456	-0.22	0.0
Araucanian			
Mapudungun [arn] <i>dew</i>	629	-0.15	0.01
Arawakan			
Central Arawakan			
Parecís [pab] <i>koxaka</i>	305	-0.19	0.14
Austro-Asiatic			
Viet-Muong			
Vietnamese [vie] <i>rôi</i>	629	-0.12	0.08
Austronesian			
Barito			
Malagasy [mlg] <i>efa</i>	755	-0.03	0.01
Malayo-Sumbawan			
Balinese [ban] <i>suba</i>	630	-0.02	0.04
Northwest Sumatra-Barrier Islands			
Batak Simalungun [bts] <i>domma</i>	425	-0.12	0.03
Oceanic			
Bwanabwana [tte] <i>-ko</i>	930	-0.03	0.03
Gapapaiwa [pwg] <i>namada</i>	384	-0.18	0.11
Hote [hot] <i>yóv</i>	622	-0.1	0.09
Iduna [viv] <i>ahe</i>	437	-0.13	0.1
Keapara [khz] <i>-wara</i>	513	-0.15	0.08
Mangga Buang [mmo] <i>lung</i>	818	-0.06	0.03
Marik [dad] <i>mogo</i>	280	-0.24	0.11
Maskelynes [klv] <i>tia</i>	454	-0.14	0.1
Mengen [mee] <i>tapu</i>	691	-0.13	0.03
North Tanna [tnn] <i>rəkis</i>	861	-0.05	0.05
Owa [stn] <i>noga</i>	735	-0.12	0.0
Sinaugoro [snc] <i>varau</i>	245	-0.21	0.12
Southwest Tanna [nwi] <i>ta</i>	913	-0.06	0.04
Takia [tbc] <i>ja</i>	373	-0.18	0.09
Wuvulu-Aua [wuv] <i>-digi-</i>	1087	-0.05	0.05

...

	Freq	Dim1	Dim2
Barbacoan			
Chachi [cbi] <i>yumaa</i>	348	-0.25	0.04
Chibchan			
Kuna			
Border Kuna [kvn] <i>pato</i>	504	-0.2	0.01
San Blas Kuna [cuk] <i>bato</i>	401	-0.2	0.07
Cofán			
Cofán			
Cofán [con] <i>tayo</i>	509	-0.15	0.01
Dagan			
Dagan			
Maiwa (Papua New Guinea) [mti] <i>wait</i>	223	-0.22	0.18
Indo-European			
Romance			
Portuguese [por] <i>já</i>	358	-0.22	0.06
Slavic			
Russian [rus] <i>yxe</i>	291	-0.25	0.06
Lower Sepik-Ramu			
Mikarew			
Aruamu [msy] <i>gifa</i>	252	-0.18	0.13
Matacoan			
Wichí Lhamtés Nocten [mtp] <i>pajkye</i>	371	-0.14	0.13
Mayan			
Chuj [cac] <i>toxo</i>	245	-0.2	0.1
Kekchí [kek] <i>ac</i>	710	-0.13	0.01
Misumalpan			
Mayangna [yan] <i>uman</i>	214	-0.26	0.11
Mískito [miq] <i>pat</i>	386	-0.24	0.03
Niger-Congo			
Bantoid			
Denya [anv] <i>mɛ-</i>	531	-0.14	0.1
Gogo [gog] <i>-akondya</i>	239	-0.16	0.14
Lenje [leh] <i>-kale</i>	308	-0.2	0.17
Mbunda [mck] <i>-ja</i>	1295	-0.05	0.1
Saamia [ism] <i>-mal oh-</i>	778	-0.03	0.09
Swahili [swh] <i>-mekwisha-</i>	98	-0.21	0.2
Venda [ven] <i>no</i>	1269	-0.05	0.0
Cross River			
Bekwarra [bkv] <i>ma</i>	650	-0.15	0.01
Gbaya-Manza-Ngbaka			
Southwest Gbaya [gso] <i>kari</i>	428	-0.19	0.04
Gur			
Bimoba [bim] <i>-po</i>	825	-0.08	0.09
Eastern Karaboro [xrb] <i>wa</i>	639	-0.15	0.07
Mampruli [maw] <i>pun</i>	647	-0.04	0.12
Kwa			
Fon [fon] <i>ko</i>	853	-0.07	0.04
Western Mande			
Southern Bobo Madaré [bwq] <i>-we</i>	766	-0.12	0.07
Nilo-Saharan			
Bongo-Bagirmi			
Gulay [gvl] <i>lá</i>	561	-0.16	0.06
Mbay [myb] <i>woi</i>	307	-0.2	0.14
Ngambay [sba] <i>mba</i>	503	-0.17	0.08
Nilotic			
Kakwa [keo] <i>azo</i>	104	-0.22	0.18
Oto-Manguean			
Zapotecan			
Yareni Zapotec [zae] <i>chiba</i>	432	-0.16	0.01
Quechuan			
Quechuan			
Ayacucho Quechua [quy] <i>-ñam</i>	169	-0.2	0.12
Eastern Apurímac Quechua [qve] <i>ña</i>	536	-0.14	0.08
Tena Lowland Quichua [quw] <i>ña</i>	235	-0.24	0.08
Sino-Tibetan			
Chinese			
Hakka Chinese [hak] <i>yi-kín</i>	478	-0.09	0.04
Mandarin Chinese [cmn] <i>yijing</i>	241	-0.18	0.15

...

	Freq	Dim1	Dim2
Kuki-Chin			
Lamkang [lmk] <i>dok</i>	1063	-0.02	0.05
Lushai [lus] <i>tawh</i>	567	-0.11	0.04
Mün Chin [mwq] <i>pyi</i>	690	-0.07	0.0
Zou [zom] <i>zo</i>	543	-0.03	0.09
Zyphe Chin [zyp] <i>thah</i>	294	-0.16	0.14
Trans-New Guinea			
Angan			
Safeyoka [apz] <i>nto</i>	514	-0.12	0.07
Chimbu			
Chuave [civ] <i>mora</i>	841	-0.11	0.05
Eastern Highlands			
Alekano [gah] <i>mota</i>	347	-0.2	0.08
Yaweyuha [yby] <i>alo</i>	381	-0.18	0.07
Finisterre-Huon			
Dedua [ded] <i>bic</i>	622	-0.14	0.04
Madang			
Amele [aey] <i>wele</i>	646	-0.1	0.09
Tucanoan			
Tucanoan			
Waimaha [bao] <i>mee</i>	444	-0.19	0.05
Uto-Aztecan			
Cahita			
Yaqui [yaq] <i>jaibu</i>	372	-0.21	0.05
Creoles and Pidgins			
Belize Kriol English [bzj] <i>don</i>	647	-0.08	0.06
Eastern Maroon Creole [djk] <i>-kaba-</i>	1158	-0.1	0.02
Hawai'i Creole English [hwc] <i>-ready</i>	424	-0.19	0.08
Papiamentu [pap] <i>kaba</i>	277	-0.2	0.13
Lower left quadrant			
Afro-Asiatic			
Biu-Mandara			
Mofu-Gudur [mif] <i>cay sem</i>	1250	-0.04	-0.05
East Chadic			
Dangaléat [daa] <i>ko</i>	1714	-0.04	-0.03
Kimré [kqp] <i>nijn</i>	874	-0.16	-0.06
Algic			
Algonquian			
Algonquin [alq] <i>aja</i>	735	-0.17	-0.08
Arawakan			
Eastern Arawakan			
Palikúr [plu] <i>kuwis</i>	1011	-0.02	-0.01
Northern Arawakan			
Garifuna [cab] <i>-ali</i>	905	-0.04	-0.01
Austronesian			
Barito			
Inabaknon [abx] <i>na</i>	1951	-0.04	-0.17
Chamorro			
Chamorro [cha] <i>esta</i>	741	-0.06	-0.05
Greater Central Philippine			
Agusan Manobo [msm] <i>on</i>	2092	-0.03	-0.18
Cebuano [ceb] <i>na</i>	1444	-0.1	-0.17
Hiligaynon [hil] <i>na</i>	993	-0.16	-0.16
Kinaray-A [krj] <i>run</i>	1860	-0.05	-0.23
Tagabawa [bgs] <i>dán</i>	997	-0.12	-0.16
Waray (Philippines) [war] <i>na</i>	1783	-0.07	-0.24
Malayo-Sumbawan			
Madurese [mad] <i>la</i>	777	-0.0	-0.06
Oceanic			
Iamalele [ym] <i>akonadi</i>	709	-0.11	-0.0
Iwal [kbn] <i>ande</i>	1061	-0.05	-0.07
Lote [uvl] <i>lo</i>	829	-0.1	-0.04
Mangseng [mbh] <i>lale</i>	965	-0.05	-0.03
Mbula [mna] <i>kek</i>	1051	-0.02	-0.03
Mutu [tuc] <i>wa</i>	898	-0.06	-0.05
Nehan [nsn] <i>manas-</i>	945	-0.12	-0.04
Tawala [tbo] <i>amaka</i>	735	-0.08	-0.01
Waima [rro] <i>aba</i>	1063	-0.04	-0.1

...

	Freq	Dim1	Dim2
Philippine			
Paranan [prf] <i>dán</i>	1819	-0.06	-0.2
South Halmahera – West New Guinea			
Ambai [amk] <i>ampa</i>	877	-0.01	-0.05
Basque			
Basque [eus] <i>aqv</i>	1099	-0.12	-0.03
Border			
Amanab [amn] <i>atí</i>	1247	-0.02	-0.0
Camsá			
Camsá [kbh] <i>ya</i>	465	-0.23	-0.03
Chiquito			
Chiquitano [cax] <i>ti-</i>	1623	-0.04	-0.03
Guaicuruan			
Pilagá [plg] <i>sóxote</i>	576	-0.15	-0.02
Toba [tob] <i>mashe</i>	939	-0.12	-0.15
Mayan			
Chol [ctu] <i>-ix</i>	836	-0.03	-0.01
Kaqchikel [cak] <i>yan</i>	751	-0.11	-0.01
Mosetenan			
Tsimané [cas] <i>aty</i>	1867	-0.05	-0.1
Niger-Congo			
Adamawa			
Dii [dur] <i>sú-</i>	732	-0.07	-0.0
Bantoid			
Mahongwe [mhb] <i>lale</i>	948	-0.05	-0.02
Swati [ssw] <i>se-</i>	1474	-0.05	-0.05
Yamba [yam] <i>larj</i>	625	-0.13	-0.01
Gbaya-Manza-Ngbaka			
Northwest Gbaya [gya] <i>kad'i</i>	575	-0.18	-0.01
Ubangi			
Sango [sag] <i>awe</i>	916	-0.08	-0.09
Nilo-Saharan			
Nilotic			
Mabaan [mfz] <i>doki</i>	835	-0.12	-0.04
Oto-Manguean			
Mixtecan			
Magdalena Peñasco Mixtec [xtm] <i>já</i>	769	-0.13	-0.05
Ocoatepec Mixtec [mie] <i>ja</i>	673	-0.15	-0.06
Silacayoapan Mixtec [mks] <i>sa</i>	1189	-0.08	-0.07
Tepeuxila Cuicatec [cux] <i>āā</i>	1303	-0.08	-0.06
Popolocan			
Chiquihuitlán Mazatec [maq] <i>ha</i>	1391	-0.07	-0.05
Huautla Mazatec [mau] <i>je³</i>	782	-0.12	-0.04
Jalapa De Díaz Mazatec [maj] <i>ja</i>	1114	-0.05	-0.15
San Jerónimo Tecóatl Mazatec [maa] <i>jye-</i>	1533	-0.01	-0.15
San Juan Atzingo Popoloca [poe] <i>ó</i>	1532	-0.04	-0.07
San Marcos Tlacoalco Popoloca [pls] <i>o-</i>	1672	-0.08	-0.1
Zapotecan			
Coatecas Altas Zapotec [zca] <i>la</i>	813	-0.14	-0.08
Santa María Quiegolani Zapotec [zpi] <i>che</i>	1482	-0.01	-0.19
Yatee Zapotec [zty] <i>ba</i>	1241	-0.0	-0.08
Panoan			
Panoan			
Shipibo-Conibo [shp] <i>moa</i>	1845	-0.08	-0.13
Quechuan			
Inga [inb] <i>ña-</i>	1181	-0.12	-0.06
North Junín Quechua [qvn] <i>nā</i>	1166	-0.1	-0.1
South Bolivian Quechua [quh] <i>-ña</i>	762	-0.18	-0.09
Sino-Tibetan			
Kuki-Chin			
Falam Chin [cfm] <i>zo</i>	780	-0.04	-0.01
Haka Chin [cnh] <i>cang</i>	660	-0.12	-0.05
Ngawn Chin [cnw] <i>zo-</i>	1045	-0.04	-0.03
Siyin Chin [csy] <i>zo</i>	893	-0.05	-0.02
Zotung Chin [czt] <i>-vae</i>	1160	-0.05	-0.07
Tarascan			
Purepecha [tsz] <i>-ia</i>	962	-0.09	-0.01

...

	Freq	Dim1	Dim2
Ticuna			
Ticuna [tca] <i>marú</i>	980	-0.07	-0.11
Totonacan			
Coyutla Totonac [toc] <i>aya</i>	975	-0.09	-0.1
Highland Totonac [tos] <i>aya</i>	865	-0.1	-0.07
Trans-New Guinea			
Angan			
Angaataha [agm] <i>aimi</i>	1405	-0.03	-0.07
Ankave [aak] <i>rixa</i>	1235	-0.07	-0.09
Eastern Highlands			
Inoke-Yate [ino] <i>ako</i>	1366	-0.02	-0.01
Kanite [kmu] <i>ago</i>	1086	-0.03	-0.01
Keyagana [kyg] <i>ago</i>	1106	-0.05	-0.01
Finisterre-Huon			
Numanggang [nop] <i>agarŋ</i>	810	-0.12	-0.02
Wantoat [wnc] <i>gwa</i>	711	-0.1	-0.02
Yau (Morobe Province) [yuw] <i>urop</i>	659	-0.17	-0.03
Kamula			
Kamula [xla] <i>m[oai]--</i>	1521	-0.03	-0.06
Tupian			
Tupi-Guaraní			
Paraguayan Guaraní [gug] <i>-ma</i>	657	-0.17	-0.06
Uo-Aztecan			
Aztecán			
Central Huasteca Nahuatl [nch] <i>ya</i>	1506	-0.06	-0.02
Eastern Huasteca Nahuatl [nhe] <i>ya</i>	1235	-0.07	-0.01
Northern Puebla Nahuatl [ncj] <i>y[fo]-</i>	1039	-0.1	-0.04
Tetelcingo Nahuatl [nhg] <i>ye</i>	1226	-0.0	-0.14
Western Huasteca Nahuatl [nhw] <i>ya</i>	1280	-0.07	-0.03
Zamucoan			
Ayoreo [ayo] <i>e</i>	1336	-0.05	-0.11
Zaparoan			
Arabela [arl] <i>tari</i>	670	-0.14	-0.03
Creoles and Pidgins			
Pijin [pis] <i>finis</i>	1218	-0.01	-0.01
Tok Pisin [tpi] <i>pinis</i>	1155	-0.01	-0.04
Upper right quadrant			
Afro-Asiatic			
Masa			
Masana [mch] <i>-awa</i>	1203	0.05	0.08
Austro-Asiatic			
Katuic			
Eastern Bru [bru] <i>khoiq</i>	1967	0.19	0.03
Viet-Muong			
Vietnamese [vie] <i>đã</i>	2524	0.29	0.1
Austronesian			
Central Malayo-Polynesian			
Manggarai [mqy] <i>poli</i>	2157	0.26	0.09
Sabu [hvn] <i>all-</i>	1718	0.21	0.03
Uab Meto [aoz] <i>-lali</i>	821	0.07	0.11
Malayo-Sumbawan			
Iban [iba] <i>udah</i>	2992	0.32	0.08
Indonesian [ind] <i>telah</i>	1583	0.29	0.09
Malay [zlm] <i>sudah</i>	1528	0.22	0.06
North Borneo			
Mainstream Kenyah [xkl] <i>-pa</i>	2033	0.16	0.05
Oceanic			
Gilbertese [gil] <i>tia</i>	1426	0.13	0.12
Kahua [agw] <i>bani</i>	1883	0.2	0.05
Kara (Papua New Guinea) [leu] <i>fo</i>	2113	0.24	0.08
Kwamera [tnk] <i>raka</i>	1496	0.03	0.03
Madak [mmx] <i>-am</i>	1538	0.07	0.03
Maori [mri] <i>kua</i>	1205	0.13	0.01
Patpatar [gfk] <i>te</i>	1620	0.14	0.02
Tungag [lcm] <i>-tala</i>	1366	0.2	0.1
South Sulawesi			
Buginese [bug] <i>pura-</i>	1247	0.23	0.07
Makasar [mak] <i>le'ba(')ka-</i>	1235	0.18	0.09

...

	Freq	Dim1	Dim2
Baining-Taulil			
Baining			
Qaqet [byx] <i>sa</i>	1286	0.05	0.02
Guahiban			
Cuiba [cui] <i>bayatha</i>	1279	0.03	0.07
Indo-European			
Celtic			
Welsh [cym] <i>wedi</i>	1745	0.2	0.12
Germanic			
Danish [dan] <i>Perfect</i>	1129	0.21	0.26
Dutch [nid] <i>Perfect</i>	2424	0.25	0.23
English [eng] <i>Perfect</i>	1812	0.3	0.2
Faroese [fao] <i>Perfect</i>	1703	0.27	0.25
German [deu] <i>Perfect</i>	2323	0.13	0.11
Icelandic [isl] <i>Perfect</i>	3375	0.26	0.18
Norwegian Bokmål [nob] <i>Perfect</i>	3063	0.28	0.22
Swedish [swe] <i>Perfect</i>	2277	0.31	0.2
Indic			
Bengali [ben] <i>-ēch[iaē](na)</i>	2489	0.21	0.08
Romance			
Catalan [cat] <i>Perfect</i>	4062	0.25	0.19
Spanish [spa] <i>Perfect</i>	1677	0.26	0.18
Mayan			
Mam [mam] <i>matxi-</i>	871	0.05	0.01
Yucateco [yua] <i>dzo'oc</i>	931	0.08	0.01
Mixe-Zoque			
Juquila Mixe [mxq] <i>tʰ</i>	2257	0.21	0.09
Niger-Congo			
Bantoid			
Bulu (Cameroon) [bum] <i>-ya</i>	1033	0.02	0.06
Fang (Equatorial Guinea) [fan] <i>yaá</i>	1381	0.19	0.06
Oku [oku] <i>-(n)en e y)?men</i>	1664	0.1	0.0
Suba [sxb] <i>-iire</i>	1957	0.1	0.04
Swahili [swh] <i>-me-</i>	3157	0.13	0.05
Tikar [tik] <i>-ā</i>	1691	0.1	0.01
Tswana [tsn] <i>Perfect</i>	1680	0.04	0.07
Kru			
Kuwaa [blh] <i>ya</i>	1016	0.09	0.01
Mel			
Timne [tem] <i>po</i>	1768	0.26	0.1
Nilo-Saharan			
Nilotic			
Adhola [adh] <i>-tyeko</i>	776	0.04	0.08
Kumam [kdi] <i>-tiēk-</i>	978	0.06	0.01
Luo (Kenya and Tanzania) [luo] <i>-se-</i>	2123	0.2	0.09
Oto-Manguean			
Chinantecan			
Comaltepec Chinantec [cco] <i>nica-</i>	1881	0.13	0.06
Ozumacín Chinantec [chz] <i>-maʔjln-</i>	1765	0.04	0.03
Sepik			
Middle Sepik			
Mende (Papua New Guinea) [sim] <i>angop</i>	1078	0.01	0.0
Upper Sepik			
Abau [aau] <i>-po</i>	1558	0.14	0.02
Sino-Tibetan			
Kuki-Chin			
Ao Naga [njo] <i>-ogo</i>	940	0.04	0.08
Sangtam Naga [nsa] <i>-ko</i>	958	0.03	0.06
Yimchungru Naga [yim] <i>-do</i>	695	0.02	0.11
Northern Naga			
Konyak Naga [nbe] <i>-ki</i>	1048	0.02	0.08
Trans-New Guinea			
Engan			
East Kewa [kjs] <i>abala-</i>	1681	0.04	0.08
West Kewa [kew] <i>aba-</i>	1681	0.04	0.09
Finisterre-Huon			
Rawa [rwo] <i>kuli</i>	1156	0.02	0.08

...

	Freq	Dim1	Dim2
Uralic			
Finnic			
Estonian [est] <i>Perfect</i>	2556	0.34	0.22
Finnish [fin] <i>Perfect</i>	1716	0.3	0.21
Western Fly			
Bine [bon] <i>itu-</i>	1508	0.18	0.15
Wipi [gdr] <i>kea</i>	1675	0.05	0.05
Creoles and Pidgins			
Saint Lucian Creole French [acf] <i>ja</i>	1341	0.14	0.01
Sea Island Creole English [gul] <i>done</i>	1943	0.17	0.04
Seselwa Creole French [crs] <i>fin</i>	2639	0.25	0.17
Lower right quadrant			
Afro-Asiatic			
Biu-Mandara			
Kamwe [hig] <i>wuri</i>	1894	0.05	-0.02
Austronesian			
Barito			
Ma'anyan [mhy] <i>haut</i>	1906	0.3	-0.15
Ngaju [nij] <i>jari</i>	1787	0.29	-0.14
Ot Danum [otd] <i>jari</i>	1633	0.27	-0.12
Celebic			
Banggai [bgz] <i>lapamo</i>	1365	0.27	-0.04
Central Malayo-Polynesian			
Alune [alp] <i>peneka</i>	1381	0.05	-0.07
Bima [bhp] <i>wa'ura</i>	1331	0.21	-0.14
Kisar [kje] <i>me'e</i>	2177	0.03	-0.12
Luang [lex] <i>olek-</i>	1820	0.12	-0.05
Termanu [twu] <i>so</i>	1101	0.04	-0.11
Javanese			
Caribbean Javanese [jvn] <i>wis</i>	2188	0.13	-0.13
Javanese [jav] <i>wis</i>	1597	0.13	-0.09
Lampungic			
Lampung Api [ljp] <i>radu</i>	2109	0.3	-0.14
Malayo-Sumbawan			
Achinese [ace] <i>ka</i>	2962	0.27	-0.11
Central Malay [pse] <i>la</i>	2341	0.24	-0.1
Indonesian [ind] <i>sudah</i>	1587	0.27	-0.16
Minangkabau [min] <i>lah</i>	2989	0.28	-0.13
Sasak [sas] <i>sampun</i>	1875	0.3	-0.12
Sundanese [sun] <i>geus</i>	1906	0.1	-0.12
North Borneo			
Western Penan [pne] <i>lepah</i>	858	0.02	-0.02
Northwest Sumatra-Barrier Islands			
Batak Angkola [akb] <i>madung</i>	828	0.06	-0.03
Batak Dairi [btd] <i>enggo</i>	1595	0.1	-0.13
Batak Karo [btx] <i>enggo</i>	2279	0.25	-0.06
Batak Toba [bbc] <i>-ung</i>	1703	0.08	-0.02
Oceanic			
'Auhelawa [kud] <i>-o</i>	1846	0.04	-0.02
Kilivila [kij] <i>bogwa</i>	1529	0.08	-0.0
Manam [mva] <i>ambe</i>	1432	0.01	-0.14
Misima-Panaeati [mpx] <i>iyaka</i>	1018	0.01	-0.09
Molima [mox] <i>nida</i>	1557	0.06	-0.03
Motu [meu] <i>vada</i>	1806	0.16	-0.01
Paicī [pri] <i>jée</i>	1768	0.05	-0.1
Uripiv-Wala-Rano-Atchin [upv] <i>pa</i>	1379	0.01	-0.04
Palauan			
Palauan [pau] <i>-la</i>	1504	0.12	-0.0
Rejang			
Rejang [rej] <i>bi</i>	1134	0.17	-0.04
Sangiric			
Sangir [sxn] <i>séng</i>	2532	0.25	-0.16
South Halmahera – West New Guinea			
Biak [bhw] <i>kwar</i>	2118	0.2	-0.04
Hmong-Mien			
Hmong Daw [mww] <i>twb</i>	1445	0.05	-0.07

...

	Freq	Dim1	Dim2
Huitotoan			
Huitoto			
Murui Huitoto [huu] <i>jai</i>	2066	0.02	-0.05
Mayan			
Tzotzil [tzo] <i>xa</i>	2104	0.04	-0.13
Niger-Congo			
Bantoid			
Bafut [bfd] <i>mê</i>	1269	0.0	-0.02
Meta' [mgo] <i>ɸe</i>	1880	0.19	-0.06
Noone [nhu] <i>wase</i>	1616	0.04	-0.13
Edoid			
Ivbie North-Okpela-Arhe [atg] <i>she</i>	878	0.03	-0.04
Platoid			
Kutep [kub] <i>pú</i>	1703	0.04	-0.06
Western Mande			
Susu [sus] <i>bara</i>	2049	0.18	-0.01
Yalunka [yal] <i>bata</i>	1716	0.19	-0.03
Nilo-Saharan			
Nilotic			
Acoli [ach] <i>dorj</i>	1645	0.01	-0.13
Lango (Uganda) [laj] <i>dorj</i>	1465	0.0	-0.12
Oto-Manguean			
Otomian			
Central Mazahua [maz] <i>ya</i>	2016	0.01	-0.14
Popolocan			
Ayautla Mazatec [vmy] <i>je</i>	1843	0.05	-0.15
Zapotecan			
Cajonos Zapotec [zad] <i>ba</i>	2163	0.15	-0.09
Isthmus Zapotec [zai] <i>ma</i>	2205	0.04	-0.17
Nopala Chatino [cya] <i>cua'</i>	2537	0.09	-0.11
Santo Domingo Albarradas Zapotec [zas] <i>ma</i>	1862	0.08	-0.17
Tabaa Zapotec [zat] <i>ba</i>	1948	0.07	-0.06
Yalálag Zapotec [zpu] <i>ba</i>	2492	0.14	-0.09
Zoogocho Zapotec [zpq] <i>ba</i>	2457	0.15	-0.05
Sino-Tibetan			
Burmese-Lolo			
Achang [acn] <i>goeus</i>	1456	0.07	-0.05
Lahu [lhu] <i>peu-</i>	1678	0.11	-0.01
Kuki-Chin			
Khumi Chin [cnk] <i>poen</i>	1200	0.01	-0.1
Matu Chin [hlt] <i>poen</i>	1200	0.0	-0.11
Tequistlatecan			
Highland Oaxaca Chontal [chd] <i>joupa</i>	1468	0.03	-0.05
Torricelli			
Marienberg			
Kamasau [kms] <i>pre</i>	1474	0.0	-0.03
Trans-New Guinea			
Eastern Highlands			
Waffa [waj] <i>vaa</i>	1557	0.0	-0.01
Mek			
Nalca [nlc] <i>seleb</i>	1328	0.03	-0.04
Tupian			
Tupi-Guaraní			
Eastern Bolivian Guaraní [gui] <i>ma</i>	1490	0.02	-0.14
Western Bolivian Guaraní [gnw] <i>ma</i>	1571	0.01	-0.15
Uto-Aztecan			
Aztecan			
Southeastern Puebla Nahuatl [npl] <i>yo-</i>	1649	0.03	-0.05
Zacatlán-Ahuacatlán-Tepetzintla Nahuatl [nhi] <i>yo-</i>	1690	0.03	-0.04
Yawa			
Yawa [yva] <i>to</i>	1412	0.04	-0.05
Creoles and Pidgins			
Bislama [bis] <i>finis</i>	1412	0.03	-0.03
Krio [kri] <i>don</i>	2662	0.19	-0.05
Nigerian Pidgin [pcm] <i>don</i>	1611	0.15	-0.06

Prototypical uses

EUROPFCT 2nd ITERATION – 20 top segments

- *"I have revealed your name to the men* whom you gave me out of the world. They were yours, and you have given them to me, and they have kept your word. (43017006\$4)
- What then? Do we have an advantage? Not at all. *For we have already charged* both Jews and Greeks are all under sin, (45003009\$18)
- Philip found Nathanael and said to him, *"We have found* the one whom Moses wrote about in the law, *and* the prophets wrote about Jesus son of Joseph from Nazareth!" (43001045\$12)
- *And I made known to them your name, and* will make it known, in order that the love with which you loved me may be in them, *and* I may be in them." (43017026\$4)
- Do not lie to one another, *because you have taken off the old man together with his deeds,* (51003009\$11)
- Then Peter answered and said to him, *"Behold, we have left everything* and followed you. What then will there be for us?" (40019027\$14)
- I have written to you, children, because you have known the Father. I have written to you, fathers, because you have known the One who is from the beginning. I have written to you, young men, because you are strong, and the word of God resides in you, and *you have conquered the evil one.* (62002014\$62)
- *if indeed you have heard about the stewardship of God's grace* given to me for you. (49003002\$5)
- *And we have seen and testify* that the Father has sent the Son to be the Savior of the world. (62004014\$4)
- *because he has set a day on which* he is going to judge the world in righteousness by the man who he has appointed, having provided proof to everyone by raising him from the dead." (44017031\$4)
- He is not here, *for he has been raised,* just as he said. Come, see the place where he was lying. (40028006\$10)
- For the Father does not judge anyone, but *he has given all judgment to the Son,* (43005022\$12)
- The Father loves the Son and *has given all things into his hand.* (43003035\$8)
- *Behold, I have given you the authority* to tread on snakes and scorpions, *and over all the power of the enemy, and* nothing will ever harm you. (42010019\$5)
- Peter began to say to him, *"Behold, we have left everything* and followed you." (41010028\$13)
- among whom are Hymenaeus and Alexander, *whom I have handed over to Satan,* in order that they may be taught not to blaspheme. (54001020\$11)
- *Now the chief priests and the Pharisees had given orders* that if anyone knew where he was, they should report it, in order that they could arrest him. (43011057\$9)
- The one who eats must not despise the one who does not eat, and the one who does not eat must not judge the one who eats, *because God has accepted him.* (45014003\$33)
- Not that anyone has seen the Father except the one who is from God *this one has seen the Father.* (43006046\$18)
- I have said these things to you so that in me you may have peace. In the world you have affliction, but have courage! *I have conquered the world."* (43016033\$30)

IAMITIVES 2nd ITERATION – 20 top segments

- Therefore whenever you practice charitable giving, do not sound a trumpet in front of you, as the hypocrites do in the synagogues and in the streets, in order that they may be praised by people. Truly I say to you, *they have received their reward in full!* (40006002\$48)
- Do not lie to one another, *because you have taken off the old man together with his deeds,* (51003009\$11)
- And whenever you pray, do not be like the hypocrites, because they love to stand and pray in the synagogues and on the corners of the streets, in order that they may be seen by people. Truly I say to you, *they have received their reward in full!* (40006005\$49)
- But I say to you that everyone who looks at a woman *to lust for her has already committed adultery with her in his heart.* (40005028\$19)
- "Whenever you fast, do not be sullen like the hypocrites, for they make their faces unrecognizable in order that they may be seen fasting by people. Truly I say to you, *they have received their reward in full!* (40006016\$39)
- But I say to you *that Elijah has already come,* and they did not recognize him, but did with him whatever they wanted. In the same way also the Son of Man is going to suffer at their hands." (40017012\$10)
- But I tell you *that indeed Elijah has come,* and they did to him whatever they wanted, just as it is written about him." (41009013\$9)
- I have said these things to you so that in me you may have peace. In the world you have affliction, but have courage! *I have conquered the world."* (43016033\$30)
- Father, glorify your name!" Then a voice came from heaven, *"I have both glorified it, and* I will glorify it again." (43012028\$17)
- He replied to them, *"I told you already and* you did not listen! Why do you want to hear it again? You do not want to become his disciples also, do you?" (43009027\$8)
- *But if I expel demons by the finger of God, then the kingdom of God has come upon you!* (42011020\$18)
- "But woe to you who are rich, *because you have received your comfort.* (42006024\$13)
- And he said to her, "Because of this statement, go! *The demon has gone out of your daughter."* (41007029\$18)
- who have deviated concerning the truth by saying *the resurrection has already taken place,* and they are upsetting the faith of some. (55002018\$13)
- saying, "Get up, take the child and his mother and go to the land of Israel, for those who were seeking the life of the child *are dead."* (40002020\$31)
- *and concerning judgment, because the ruler of this world has been condemned.* (43016011\$13)
- *But if I expel demons by the Spirit of God, then the kingdom of God has come upon you!* (40012028\$18)
- But when they came to Jesus, after they saw *he was already dead,* they did not break his legs. (43019033\$12)
- So when he arrived, Jesus found *he had already been four days in the tomb.* (43011017\$11)
- Know *that our brother Timothy has been released, with whom* I will see you, if he comes quickly enough. (58013023\$8)

Examples of segments highly correlated with Dimension 2

- a) So also you, when you see all these things, know *that he is near, at the door.* (40024033\$15)
 - b) "Now learn the parable from the fig tree: Whenever its branch has already become tender and puts forth its leaves, you know that *summer is near.*" (41013028\$28)
 - c) So also you, when you see these things happening, know *that he is near, at the door.* (41013029\$15)
 - d) *And the Passover of the Jews was near, and Jesus went up to Jerusalem.* (43002013\$7)
 - e) *Now the Passover, the feast of the Jews, was near.* (43006004\$11)
 - f) *Now the Passover of the Jews was near, and many went up to Jerusalem from the surrounding country before the Passover, so that they could purify themselves.* (43011055\$7)
 - g) And not being weak in faith, he considered his own body as good as dead, *because he was approximately a hundred years old, and the deadness of Sarah's womb.* (45004019\$20)
 - h) for you were formerly darkness, but *now you are light in the Lord.* Live like children of light (49005008\$10)
 - i) *For I am already being poured out as a drink offering, and the time of my departure is imminent.* (55004006\$6)
 - j) You also be patient. Strengthen your hearts, *because the coming of the Lord is near.* (59005008\$16)
-

Examples of 'universal perfect' contexts

- a) *Therefore I have experienced help from God until this day,* and I stand here testifying to both small and great, saying nothing except what both the prophets and Moses have said were going to happen. (44026022\$4)
 - b) By teaching these things to the brothers, you will be a good servant of Christ Jesus, trained in the words of the faith and of the good teaching *that you have followed faithfully.* (54004006\$34)
 - c) And looking intently at the Sanhedrin, Paul said, "*Men and brothers, I have lived my life in all good conscience before God to this day.*" (44023001\$18)
 - d) Jesus replied to him, "*I have spoken openly to the world. I always taught in the synagogue and in the temple courts where all the Jews assemble, and I have said nothing in secret.*" (43018020\$9)
 - e) You did not give me a kiss, but *from the time I entered, she has not stopped kissing my feet.* (42007045\$19)
 - f) For this reason *that field has been called the Field of Blood until today.* (40027008\$8)
 - g) *Therefore my dear friends, just as you have always obeyed, not as in my presence only but now much more in my absence, work out your own salvation with fear and trembling.* (50002012\$11)
 - h) Jesus replied to him, "I have spoken openly to the world. *I always taught in the synagogue and in the temple courts where all the Jews assemble, and I have said nothing in secret.*" (43018020\$17)
 - i) So they took the money and did as they were told, and *spread abroad this report among the Jews until this very day.* (40028015\$11)
 - j) And Simon answered and said, "Master, although *we worked hard through the whole night, we caught nothing. But at your word I will let down the nets.*" (42005005\$12)
 - k) in order that what was spoken through the prophet would be fulfilled, who said, "I will open my mouth in parables; I will proclaim *what has been hidden since the creation.*" (40013035\$32)
-