

# Directionality in the development of Future grams

Marvin Martiny

December 10, 2022

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Definitions . . . . .	2
1.2	Research questions . . . . .	2
<b>2</b>	<b>State of research</b>	<b>2</b>
2.1	Development of future grams . . . . .	2
2.2	Onwards: Movement in the development of future . . . . .	3
<b>3</b>	<b>Method</b>	<b>4</b>
3.1	Data sources . . . . .	4
3.2	Data retrieval & processing . . . . .	5
3.2.1	Semantic grouping . . . . .	7
3.2.2	Syntactic grouping . . . . .	8
<b>4</b>	<b>Results</b>	<b>8</b>
<b>5</b>	<b>Discussion &amp; conclusion</b>	<b>9</b>
5.1	Movement and directionality . . . . .	9
5.2	Argument structure . . . . .	10
5.3	Conclusion . . . . .	10
<b>6</b>	<b>Outlook</b>	<b>10</b>
<b>7</b>	<b>Sources</b>	<b>12</b>

# 1 Introduction

## 1.1 Definitions

There are countless definitions of tense and future. Plainly speaking, future tense is the grammaticalized expression of future time reference. Here, I refer to two common definitions:

- (i) Bybee, Perkins and Pagliuca (1994: 244):

We regard the focal use of future as equivalent to a prediction on the part of the speaker that the situation in the proposition, which refers to an event taking place after the moment of speech, will hold [...].

- (ii) future = {S,R} before E (Reichenbach 1947)

These definitions should be seen as complementary rather than exclusive with regards to one another.

Drawing on definitions like these, we might come to a working definition like the following:

Future (tense) is the name for a morpheme, which is grammaticalized to a certain extent<sup>1</sup> and expresses as one of its core functions the notion of PREDICTION (and may or may not include the expression of intention). In order for something to be a prediction, the instance of speech needs to precede the expressed predicate in time.

## 1.2 Research questions

The idea for this study was ignited, among others, by two important observations on the grammaticalization of future grams:

- Futures evolve from a fairly restricted range of lexical sources (Bybee et al. 1994: 244).
- Movement verbs figure more prominently as sources than verbs or other lexical material of any other type (Bybee et al. 1994: 267).

The aims of the study are based on the above statements. They may be formulated as follows:

- (i) Do these statements hold true in the face of new data?
- (ii) What may be the reason for the major role of movement verbs?

# 2 State of research

## 2.1 Development of future grams

**Primary vs. secondary future:** Bybee, Perkins, Pagliuca (1994) distinguish between primary futures and aspectual futures, or futures which develop “directly” from lexical material and futures which develop from already grammatical aspectual morphs, respectively. Importantly, non-grammatical elements of temporal reference, like temporal adverbs, are counted as being lexical elements, thus giving rise to primary futures.

Whereas Bybee and colleagues include futures which develop via modality meanings in the group of primary futures, I deviate from that by distinguishing primary and secondary futures more strictly – development via grammatical expressions of modality are reckoned to be secondary futures as well.

Secondary futures may in principle develop from many kinds of grams, however, they often derive from other

---

<sup>1</sup>In my view, being grammaticalized first and foremost means to be integrated into a paradigm.

grams in the realm of verbal categories. Since there have been instances of future from all canonical verbal categories except evidentiality, we have to differentiate between the following types of secondary future at least:

- (1) aspectual future
- (2) modal future
- (3) future from non-past

**Future from Intention:** All primary futures develop via the meaning of INTENTION. This claim by Bybee, Perkins and Pagliuca (1994: 253–254) has been debated e.g. in Hilpert (2008).

While Hilpert (2008) showed that some movement-based futures develop via aspectual notions like inchoative, this does not falsify the claim, since despite its wording (“[...] all futures go through a stage of functioning to express the intention [...]”), only primary futures were included. And although movement-based futures usually are primary, as soon as a grammaticalization path is developing into future via an aspectual function (like inchoative), it is no longer a primary future. However, in the *Evolution of Grammar* primary futures include temporal adverbs and modal expressions (“agent-oriented modality”) and thus the claim is extended to futures from temporal adverbs without providing substantial evidence for this. Usually, these elements first signal the intention of the speaker, then of the (animate) agent. Once the use expands to inanimate agents, this is reanalysed as prediction of the speaker (since inanimate agents are usually not conceptualized as having intentions) and thus deal with a futur gram in the sense of our working definition, based on the notion of PREDICTION (cf. Bybee et al. 1994: 254, 270, in passim).

Paths of primary and modal futures are shown in a reduced semantic map in Figure 1.

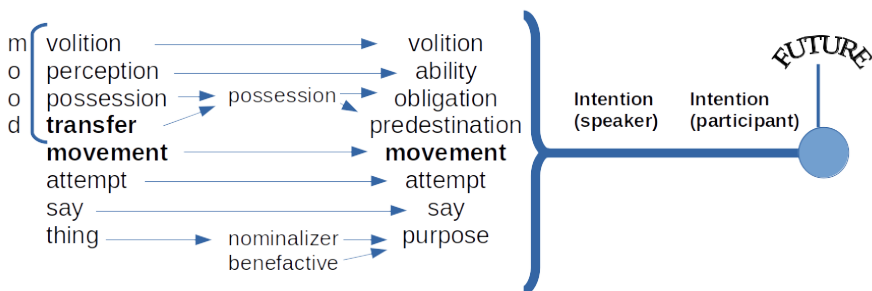


Figure 1: Semantic Map based on Kuteva et al. 2019 & Bybee et al. 1994.

Do all primary futures under my definition develop the function of prediction via the notion of intention? Not necessarily, since for primary futures from direct temporal reference (e.g. temporal adverbs) there is not sufficient evidence for this claim.

**Future from human agent predicates:** All modal and movement futures begin in the above process with the intentions of human agents and only subsequently expand to mark the intentions of other animate beings and prediction. The same may be true for the obligatorification of temporal adverbs (cf. Bybee et al. 1994: 254, 270).

## 2.2 Onwards: Movement in the development of future

In the *Evolution of Grammar*, Bybee et al. (1994:280) claim the following:

The more common sources for futures are those that yield the intention inference most easily — desire, strong obligation, and movement toward a goal. Weak obligation, ability, attempt, and temporal adverbs are not as susceptible to the intention inference and thus do not yield futures as readily.

Setting aside the issue mentioned above, this implies that the semantic distance that has to be bridged in order to yield the intention interpretation is the (main) reason for the respective frequency of sources. Following this interpretation, we can deduct that movement toward a goal has to yield the intention interpretation most readily, more so than even volition.

Only verbs which convey movement towards a goal can develop into future. This ‘allative component’ may be either inherent in the semantics of the verb or explicit in the construction (cf. Bybee et al. 1994: 268).

The construction may entail progressive or imperfective aspect or present tense, as the English going-to future does (Bybee et al. 1994: 268).

Kuteva and her colleagues (2019/2020:12) also find that the whole of the construction is of importance, and that movement verbs may also develop into past tense under different circumstances. Thus, the kind of argument(s) a verb takes or may take determines the possibilities of their development.

This brings us to another potentially interesting feature of movement verbs: they are verbs.

Not only do verbs almost always take arguments which receive a certain pragmatic role in accordance with the lexical semantics of the verb. But they may be monovalent, bivalent, trivalent or more, selecting a certain number of arguments whose semantics are at least statistically determined by their pragmatic role (e.g. benefactives will mostly be animate, not many beneficiaries do things for (as in: in favor of) a stone.

Thus, the number and the kind of arguments that a verb may take are areas of interest.

## 3 Method

### 3.1 Data sources

In order to answer the research questions – to check the relative frequency of movement verbs and explore potential reasons for the distribution – I put together a collection of grammaticalization paths leading to future.

For the purpose of this exploratory study, all grams which fall under the working definition and whose source construction has been described in the literature is taken into account. Three earlier works have been used to gather the lion’s share of the data. These are:

- (i) Bybee et al. 1994 (The evolution of grammar)
- (ii) Kuteva et al. 2019/2020 (World lexicon of grammaticalization, 2nd edition)
- (iii) Bisang et al. (forthcoming) (MAGRAM database) largely based on Bisang & Malchukov 2020 (Handbook of grammaticalization scenarios, eds.)<sup>2</sup>

I have used these, since all of them contain a large number of grammaticalization paths from around the world. In addition, I complemented this data with information I gathered for my M.A. thesis on grammaticalization in Amazonia (unpublished).<sup>3</sup>

From this investigation, 163 (paths to) future grams emerged for which there is at least basic information on the functional development.

In Table 1 below, you can see how these paths distribute over the data sources. If the same path is mentioned by two or more sources, this is counted separately in this representation.

---

<sup>2</sup>The MAGRAM database is not yet a digital publication, however, the data can be accessed [here](#) already.

<sup>3</sup>You can have a look at the data of the original project [here](#).

Data source	Number of grams
Bybee et al. 1994	46
Bybee et al. 1994; Bisang et al. forth.; Kuteva et al. 2019	1
Bisang et al. forth.	30
Kuteva et al. 2019	79
Kuteva et al. 2019; Bisang et al. forth.	1
MA	6
Total Result	163

Table 1: Grams per data source.

Despite containing a large number of data points taking into account the intricate nature of the information, it has to be said that this is not a controlled sample, neither in genealogical nor geographical terms. This is due to the fact that only Bybee et al. (1994) tried to sample their data for a morphosyntactic typological study, whereas the MAGRAM project (Bisang & Malchukov 2020, eds.) and the World lexicon of grammaticalization (Kuteva et al. 2019/2020) put emphasis on cross-linguistic diversity, but did not use strict sampling methods because of the nature of the respective publications (collaborative, handbook/lexicon). My master thesis on the other hand (Martiny, unpublished) was focused solely on Amazonian data, and thus only complements the rest of the data, in which South America was (even more) underrepresented. In a follow-up study, a statistical analysis will be conducted with even more data (since future is thankfully often transparent enough to get the most basic data on development) which will counteract the sampling biases in order to validate the outcomes of the present and other studies.

For now, the number of datapoints (grams) per macro-area is displayed in Table 2 and the number of datapoints per language family is shown in Table 3.

Macro-Area	Number of grams
Africa	70
Australia	1
Eurasia	59
North America	14
Papunesia	9
South America	10
Total Result	163

Table 2: Grams per macro-area.

As can be seen from the tables, Africa and Eurasia are overrepresented, while especially Australia is strongly underrepresented. Similarly, especially Indo-European and also Atlantic-Congo and Kru languages are overrepresented. At least, the Indo-European languages belong to several *genera*, i.e. are not all closely related. In addition, contact between languages of the dataset and especially within the same family is not controlled for, either. The geographical distribution of languages per family is visualized in Figure 2.

### 3.2 Data retrieval & processing

Information from the above sources was collected in the basic format of:

(lexical) `source`  $\rightarrow$  `target`

This includes the function or meaning of both source and target, and the morphosyntactic realization of the constructions in question.

Family	Number of grams
Indo-European	38
Atlantic-Congo	27
Kru	17
Sino-Tibetan	13
Nilotic	10
Afro-Asiatic	6
Mande	6
Mayan	4
Austronesian	3
Nakh-Daghestanian	3
Athabaskan-Eyak-Tlingit	2
Basque	2
Dravidian	2
Eskimo-Aleut	2
Kadugli-Krongo	2
Nadahup	2
Sepik	2
Tucanoan	2
Austrasiatic	1
Austroasiatic	1
Chibchan	1
Guaicuruan	1
Iroquian	1
Iwaidjan Proper	1
Koreanic	1
Mongolic	1
Nambiquaran	1
Nuclear Trans New Guinea	1
Panoan	1
Quechuan	1
Siouan	1
Takanan	1
Timor-Alor-Pantar	1
Tungusic	1
Tupi-Guarani	1
Uralic	1
Wintuan	1
Yeniseian	1
Total Result	163

Table 3: Grams per language family.

Both the meaning or function and the form and thus determined size of source and target constructions was taken from the presentation of the authors suggesting the development. As such, we have to assume that the source given is the earliest attested or reconstructed meaning/function.

Most importantly, every source in the format presented above was assigned a **semantic group** and a **syntactic group**, just as every target was assigned a **specialization**. In addition, every path was annotated for the type of development happening:

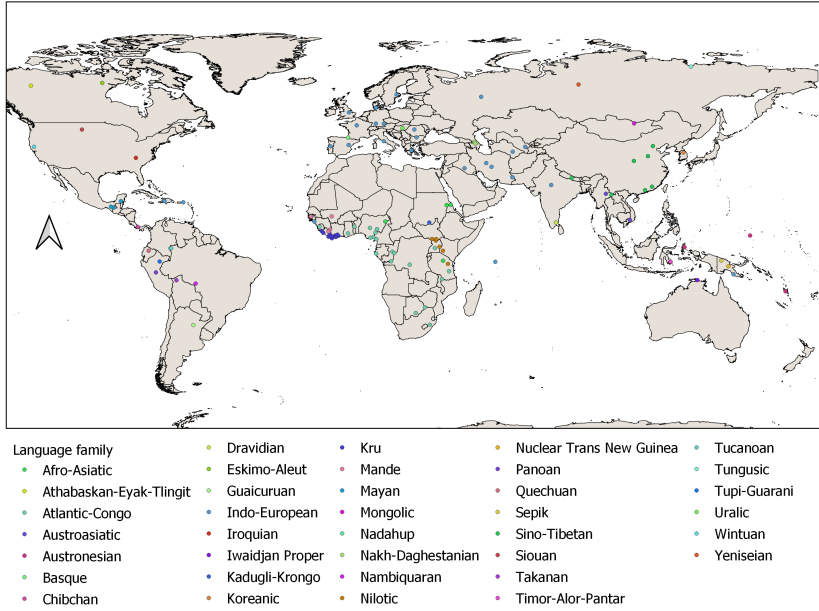


Figure 2: Languages / language families included in the dataset.

### 3.2.1 Semantic grouping

**Simplification of construction:** From every complex construction, a single simplified representation of source and target was determined. This method was based on the notion of the nucleus of a grammaticalizing construction (Bisang et al. 2020:xi).<sup>4</sup>

**Semantic grouping of sources:** Inspecting the function of each source concept, groups of attested future grams were formed on a semantic basis in the following manner:

- (1) Groups which were mentioned in previous studies as source groups were put together. A notable exception is “possession” which is subsumed under “obligation” or modality in Bybee et al. (1994) since it presumably always develops via this path. However, since possession may develop into future through predestination (and possibly other paths) as well (Bybee et al. 1994: 261–64, 280), we will take it as a group on its own.
- (2) Other source items with clear similarities were clustered into groups, e.g. all expressions with a local reference which do not clearly involve movement or direction, were labelled “locational”.
- (3) All sources which could not easily be put into a single category in step 1 or 2 were handled as *hapax legomena* and labelled with their simplified source label without further abstraction. An example would be STOP.

I intended that by this methodology all labels to be as fine-grained as necessary; however, on the flipside, one has to admit that the labels now obviously are not all on the same level of abstraction, which comes with disadvantages itself.

<sup>4</sup>The most lexical morpheme in a construction is regarded as the nucleus, the idea being that this morpheme bears the brunt of the functional load. However, this may lead to so-called *nucleus mismatches*.

### 3.2.2 Syntactic grouping

The main rationale behind the syntactic grouping was mainly to distinguish between nouns, verbs and adjectives as lexical sources.

For the cross-linguistic applicability of this method, the language-specific word classes were taken into account, e.g. for languages without adjectives expressing attributes nominally ('with bigness') or verbally ('be big') the respective morphosyntactic class would be given.

The rest of the sources was more roughly categorized into nominal and verbal categories being of interest for the development of secondary futures, and a dustbin category 'other' which I later decided to split into adverbs, connectors and (bound) derivational morphemes.

## 4 Results

In Table 4 you can see the 10 most frequent semantic groups of sources as described above with their absolute and relative frequency.

As is apparent, movement is the most frequent source group by far, being three times as frequent as the second most frequent group, volition.

Source: semantic group	Number of grams	Percentage
<b>MOVEMENT</b>	<b>70</b>	<b>43.21%</b>
VOLITION	23	14.20%
TEMPORAL	13	8.02%
POSSESSION	12	7.41%
COPULA	8	4.94%
OBLIGATION	8	4.94%
DIRECTIONAL	5	3.09%
TRANSFER	5	3.09%
CHANGE	2	1.23%
SPEECH	2	1.23%
Total (all) Result	162	100.00%

Table 4: Most frequent semantic groups.

In Table 5 you can see the complete results of the semantic grouping as described above. Note that it encompasses lexical as well as grammatical sources from the collection.

It is roughly sorted into clearly and solely lexical, potentially lexical or grammatical and clearly and solely grammatical source groups.

In Table 6 on the other hand, we see the syntactic groupings. Most importantly, the more noun-like and verb-like lexical source are placed in adjacency to each other, the more grammaticalized adverbs, connectors and derivational morphemes stand in the middle, and the nominal and verbal categories are put on the bottom of the table.

The results clearly show, that verbs are by far the most important syntactic group within the sources.

It is also noteworthy that there are no adjectives among the sources and that also among the core lexical classes (NVA), verbs are more frequent than nouns (70 times!).

Interestingly, also verbal categories are more frequent as sources than nominal categories.



Source: semantic group	Number of grams	Percentage
ATTEMPT	1	0.61%
DO	1	0.61%
MEET	1	0.61%
MOVEMENT	70	42.94%
SPEECH	2	1.23%
STOP	1	0.61%
THING	1	0.61%
TRANSFER	5	3.07%
TREE	1	0.61%
ABILITY	1	0.61%
CHANGE	2	1.23%
COPULA	8	4.91%
COPULA, POSSESSION	1	0.61%
DIRECTIONAL	5	3.07%
EXIST	1	0.61%
LOCATIONAL	1	0.61%
OBLIGATION	8	4.91%
POSSESSION	12	7.36%
POSSESSION, TRANSFER	1	0.61%
TEMPORAL	13	7.98%
VOLITION	23	14.11%
ASPECT	1	0.61%
NOMINALIZER	2	1.23%
PURPOSE	1	0.61%
Total Result	163	100.00%

Table 5: Semantic groups of source concepts.

Source: Syntactic group	Number of grams	Percentage
nouny	2	1.23%
<b>verb</b>	<b>140</b>	<b>85.89%</b>
adverb	12	7.36%
adverb, connector	1	0.61%
derivational	1	0.61%
nominal category	1	0.61%
verbal category	6	3.68%
Total Result	163	100.00%

Table 6: Syntactic group of source concept.

## 5 Discussion & conclusion

### 5.1 Movement and directionality

The ranking of the most important source groups for (primary) future tense exactly affirms the ranking by Bybee, Perkins and Pagliuca (1994): Movement, volition and temporal (adverbs). Movement is the most frequent source group by a huge margin.

If one adds the directional expressions which are not strictly speaking expressions of movement, the new larger group DIRECTIONAL even exceeds a share 50% (cf. Table 7).

One may also interpret the relative frequency of the concepts within this group. Not only are almost all types and tokens verbs by nature, but in fact the very general verbs of movement COME and GO are by far

the most common, the former contributing 33 or 34 instances, the latter contributing 31 instances.

Source: concept (gloss)	Number of grams	Percentage
APPROACH, COME_NEAR	1	1.33%
BE_NEAR, REACH/ARRIVE	1	1.33%
COME	33	44.00%
COME, DO	1	1.33%
GO	31	41.33%
RETURN	1	1.33%
SURROUND	1	1.33%
ANDATIVE	1	1.33%
ASSOCIATED MOTION	1	1.33%
HITHER	1	1.33%
LOOK_FOR	1	1.33%
VENTIVE	2	2.67%
Total Result	75	100.00%

Table 7: Movement & directional sources.

## 5.2 Argument structure

But it also shows in the results that verbs are so much more important than other source concepts that the importance of the ‘short’ semantic bridge necessary for movement or directionality seems like a small contributing factor in comparison. As Table 6 tells, verbs make up almost 86% of the source concepts.

I suggest that the reason could be called RELATIONALITY or argument structure. In the most simple terms, that verbs take arguments may be of importance.

Any grammatical structure is relational by nature. But in order for the necessary reanalysis to occur, there have to be plausible syntactic contexts, not only plausible semantic contexts. Future constructions more specifically need two open argument slots for the typical route from a lexical item: one slot for the verb and one slot for the S/A argument. This is something that especially bivalent verbs can provide.

## 5.3 Conclusion

I hypothesize that (directional) constructions with one **agent** and one **goal** argument are thus semantically and syntactically the ideal candidates to be reanalysed as future tense markers.

I further suggest that verbs with this structure are more likely to undergo this reinterpretation because their argument structure is provided by the lexical morpheme and not by other parts of the construction, so that there are more contexts in which the reinterpretation may draw from. Movement verbs often exhibit such an argument structure, and would thus be not the only construction profiting from these circumstances, but the most prominent one (especially considering their overall frequency).

This would explain the importance of verbs for this and other grammaticalization processes as well as the importance of the semantic groups of directional and more specifically movement constructions at the same time.

## 6 Outlook

From the above certainly preliminary conclusions spring several plans for furthering this research.

First and foremost, the dataset should be analysed more closely using statistical methods in order to counteract the biases introduced by the non-independent data.

Some of the hypotheses that could then be tested regarding movement, directionality and the argument structure would be:

- (H1) Inherently directional constructions with the respective argument slots are more readily available for reanalysis and thus more common than elements which may carry directional meaning only in limited number of the constructions they are partaking in. COME\_TO, ARRIVE, APPROACH etc. will occur more often as sources than e.g. verbs of undirected movement like WALK\_AROUND with allative or illative, or posture verbs (STAND, SIT).
- (H2) If DIRECTIONALITY (goal-directedness) is the relevant property, then also sources like TURN\_TOWARDS or LOOK\_AT are plausible candidates (but frequent only relative to their collostructional frequency).
- (H3) If the argument structure of verbs facilitates the reanalysis, then bivalent verbs and other constructions should be the more common as source constructions than constructions with one or three arguments.

Apart from these hypotheses, there are further questions which could be of interest. For example, how temporal adverbs develop into future, whether there are different paths and whether or not any of them in fact lead through intentionality.

There are tentative claims by Bybee and colleagues (Bybee et al. 1994) that the source and the target type are correlated for future tense; more specifically, that come-derived futures are more likely to develop into imminence or near future markers, whereas go-derived futures do not show this tendency. With the data available one could follow up on this claim.

## 7 Sources