Recent change and grammaticalization*
Manfred Krug & Ole Schützler
University of Bamberg

1. Introduction and methodology

In this chapter, we will deal with a recent change in English, namely an ongoing functional and formal change in the construction *the idea is* when it is followed by a clausal complement, as in the following example from the British National Corpus (hereafter BNC):

(1) The other example is known as the ‘could-sell’ concept. *The idea is to compare* a store’s performance not just with its historical record, but with some more objective measure of how well it should be doing. (BNC CRC 115)

We will interpret the changes discussed below as evidencing a construction grammaticalizing into an intention marker, which competes with and complements well-known verbal constructions like *want to, be going to or plan to*. In section 2, we focus on the older, purely verbal intention markers, while in the remainder of this chapter we will focus on the intention marker *the idea is*. The prototypical complements of *the idea is* constructions are also infinitival *to*-clauses (though there are others which will receive due attention in later sections). It is for such functional factors, which are reflected in semantic and structural parallelism with verbal intention markers, that we believe the verb phrase to be *the* (or at least *a*) proper locus for a discussion of the constructions under investigation.

In defining the subject matter onomasiologically and in terms of form-meaning pairs, we adopt a construction grammar framework for our present investigation (cf. e.g. Goldberg 1995; Kay & Fillmore 1999; Croft 2001; Trousdale 2008). As for meaning, the broad functional domain has been defined as ‘intention marking’, but overlap with adjacent semantic fields must be expected in a discussion of grammaticalization. As for form, the structural definition can feature at different levels of schematicity. When we use the all-embracing term ‘*idea*-constructions’, we make no assumptions regarding meaning or syntactic patterns but refer to the potentially infinite number of constructions that involve the noun *idea*. The constructional schemas that are more immediately relevant to the functional domain under scrutiny are provided in (a) to (d) below:

---

* We are grateful to Julia Schlüter, an anonymous referee and the editors for many valuable comments. We have also profited from discussions at conferences in Toulouse, Leuven and London, where earlier versions of this paper were presented.
a) determiner + abstract noun + BE + clause
b) determiner + idea + BE + clause
c) the idea is/was + that/to-clause
d) the idea is + to-clause

In order to enhance our understanding of the most prototypical and most frequent intention-marking idea-construction — the string given as (d) above — we will at various points of this chapter have to discuss more general levels of schematicity, including what is the maximally general constructional schema in the relevant functional domain (a). This may feature other abstract nouns that can be used for intention marking (e.g. point, thing, aim) as well as following main and that-clauses. To anticipate some of our conclusions, an increase in syntagmatic fixation, bondedness and phonological specialization (and thus, by concomitance, an increase in grammaticalization, cf. Lehmann 1995: 164) correlates with decreasing ambiguity. It will be seen, then, that constructions including idea are most likely to signal intention-marking when:

i) the determiner is the definite article,
ii) the noun idea figures in the singular and is strictly adjacent to the verb is, which is
iii) immediately followed by a to-clause,
iv) idea is reduced to a disyllabic word and stressed on the first syllable,
v) idea and is are linked by intrusive /r/ (in non-rhotic accents).

None of the above properties reflects a necessary or sufficient criterion in an Aristotelian sense, but the probability for an idea-construction to signal intention marking is close to 100% if all characteristics apply at the same time.

In what follows, we will adopt a multi-level approach and argue our case from several perspectives, most notably diachronic and synchronic corpus linguistics and, within the latter, the study of phonological (prosodic) variation. One aspect of interest in this investigation is the process of ongoing layering within the domain of intention-marking constructions in English as we witness the introduction of a construction of the type NP + be. The idea is... is thus interesting in many respects and shows that layering and grammaticalization need not be restricted to paradigmatic choices of the same word class (like verbs) but can extend across phrase boundaries and sentence constituents. In the case at hand, the change indeed extends across two constituent boundaries, as the complete structure is NP + VP + Complement (clause).

In addition, we see evidence that the formal change accompanying this functional change follows a different strategy than is normally the case. Rather than contraction and morphophonological reduction (as in classic cases like gonna, wanna), the construction under investigation involves changes in stress-pattern and liaison, more specifically in the form of
intrusive /t/ (see Allerton 2000: 575 for different liaison strategies). We believe that we may be looking at a phenomenon that, while its path of change is less obvious than in many well-known cases of grammaticalization in intention markers, is very special in augmenting the inventory of what can grammaticalize and that also shows more generally new ways of how grammaticalization can be formally indicated.

The present investigation was sparked by the informal observation of phonological variation in *idea* with stress sometimes being on the first, sometimes on the second syllable. The initial research hypothesis was that this might simply reflect a change from a Romance to a Germanic stress pattern (similar to AmE *address*) which would be mediated and thus speeded up by rhythmic factors like stress clash avoidance (along the lines of the *Cambridge English Pronouncing Dictionary*, 2006, cited in Figure 9 below). However, the matter seemed more complex. Our intuition was that in constructions of the type *The idea is/was*, where the complement takes the form of a to-infinitive as shown in (1) and (2), a stress shift from the second to the first syllable in the word *idea* occurred rather frequently and, moreover, that a stress shift was less likely in constructions of the type *[The idea is / was + NP / AdjP]*, even if they had a similar rhythmic structure, as in example (3):

(2) The Idea is / was to go and meet them.

(3) The idEA is / was a very good one.

A wide range of factors can potentially explain such variation. For example, social factors (in our case the sex and age of different speakers) and regional factors (in our study the two broad categories of British and American English) may play a role. There may also be language-internal factors that can be divided into phonological factors (prosodic environment, e.g. stressed neighbouring syllables and potential /t/-liaison sites), and syntactic factors (clause-final or prepausal position of *idea*). But of course the construction *idea is* + complement clause itself – or, more specifically, a potentially ongoing functional change affecting it – must be suspected as a factor at least partly responsible for the observed phonological variation. In other words: The observed formal change may very well point to ongoing grammaticalization. Thus, although at the outset we observed rhythmic variation and were puzzled as to its motivation, we are now more interested in this variation as a symptom of a much more fundamental, namely grammatical, change. Our argument will begin with a discussion and quantitative analysis of *the idea is/was* as a construction type with certain (changing) functions and later move on to an analysis of its formal development.
2. Intention-marking constructions in English

Examples (4) to (8) below all express the subject’s intention to meet John. All of them are clearly still current, but they emerged at different times in the history of the language, and thus can be viewed as a point in case of Hopper’s (1991: 22) first principle of grammaticalization, layering:

Within a broad functional domain, new layers are continually emerging. As it happens, the older layers are not necessarily discarded, but may remain to coexist with and interact with the newer layers.

As has been pointed out above, the “broad functional domain” for the present study is intention-marking, which overlaps with volition and future marking and which is adjacent to these notions on grammaticalization paths (Bybee et al. 1994: Chs. 6f.). All three semantic notions belong to the domain of auxiliaries (Krug 2011), and different layers of intention markers in English are illustrated below. All of them can claim membership of that domain by either having intention as one of their original lexical meanings, or by virtue of the semantic changes they have undergone through grammaticalization. We will briefly consider the forms in their chronological order of appearance in English, though, for expository purposes, they will always be adapted to the same modern example (past tense forms are given in brackets).

The use of will (would) for the expression of volition and intention, as in (4), is attested from the earliest stages of English, i.e. since pre-Old English times. Notice, however, that the intention reading of the historical past tense would – which for Shakespeare would still regularly mean ‘wanted to’ – is rather marked in present-day English, where would is now almost exclusively reserved for other grammatical functions like the expression of irrealis, hypotheticality, past habitual or for tense shifting in indirect speech:

(4) We will (would) meet John at eight.

According to the OED, the use of the verb intend (5) emerged during Middle English times, around the year 1400, while the use of be going to (6) as a semi-auxiliary verb to indicate the status of an action as planned or intended can be dated back to the 15th century (Danchev & Kytö 1994; Mair 2004):

(5) We intend (intended) to meet John at eight.

(6) We are (were) going to meet John at eight.

A yet more recent addition to the inventory of intention-marking constructions is the use of want + to-infinitive (7), which can be found in this meaning from about 1700 onwards and whose contracted form wanna + infinitive has been attested since 1896 (Krug 2002; OED s.v. want and wanna):
We want (wanted) to/wanna meet John at eight.

Finally, according to the OED, the verb plan to for expressing intention was introduced in the 18th century (8):

We plan (planned) to meet John at eight.

As no major losses have been reported for the inventory of English intention-marking constructions, the number of options seem to have grown over the centuries, and the expression of intention comes in essentially two grammatical types, either by means of a main verb or a (quasi-)auxiliary verb: Virtually all analyses classify will as a (central) modal auxiliary; be going to and want to/wanna are increasingly regarded as constructions near the auxiliary pole or as belonging to an intermediate focal point on the gradient from auxiliary to full verb status (cf. Quirk et al. 1985: 13; Krug 2000; Verplaetse 2003; Ziegeler 2008). *Intend to* and *plan to*, finally, are generally seen as closer to the main verb pole on the cline of auxiliariness (cf. Bolinger 1980).

Given the appropriate context, the second sentence in (9) below is yet another rough semantic equivalent of the above examples. According to the OED (s.v. *idea*) it is by far the youngest, having emerged only as late as the 1890s. Adapted to the examples above, we obtain:

We (thought we) should get moving. The *idea is (was) to* meet John at eight.

A rough equivalent to a *to*-complement is a *that*-clause with varying use of *should*:

The idea is (was) that we (should) meet John at eight.

If we accept such *idea*-constructions as another more recent addition to the set sketched above, this would indicate that the options for intention marking in English are still expanding, but also that the formal diversity of the relevant markers is increasing in present-day English. Compared with the above types (4) to (8), the construction under investigation is quite an exception in that it consists of the fixed subject NP *the idea* followed by the copula and a choice between a *to-* and *that-*complement clause. It differs significantly in other ways, too: For one, the analogue to the agent in examples (4) to (8) does not have to be specified, though it can in principle, as in the following (much less grammaticalized) construction:

*Our idea is (was) to* meet John at eight.

In this respect, *idea is constructions resemble passives, which can occur with and without a by-agent:

He was seen on the train (by them).
The idea is construction, therefore, can fill a functional gap when the source of the intention is unknown, unclear, complex, backgrounded or when it is intentionally left vague or unspecified by the speaker. The naturally occurring example (11) from the BNC above, whose central part is repeated here for convenience, may serve to illustrate this:

(11) The other example is known as the ‘could-sell’ concept. The idea is to compare a store’s performance…

The source of the intention could be the economist or group of economists who developed the ‘could-sell’ concept as well as the author of the passage or a business adopting the concept. In addition, the idea is construction can exemplify the ‘could-sell’ concept. It could also be a combination of all the above. It is noteworthy, too, that the ‘could-sell’ concept is a third-person inanimate NP and thus unlikely to figure as an agent with any of the verbal intention markers given above as (4) to (8), all of which sound odd or ungrammatical in this context:

(11)’ ??The other example is known as the ‘could-sell’ concept. It will / wants to / intends to / plans to compare a store’s performance…

Due to different degrees of grammaticalization and semantic generalization, selection restrictions (like animacy) on the subject in different intention markers come in shades. Compare the examples below:

(12) This paper will show that…
(13) This paper is going to show that…
(14) ??This paper wants to show that…
(15) ??This paper intends to show that…
(16) ??This paper plans to show that…

The last three are marginally acceptable perhaps only because of the metonymy of paper, which may imply the author(s) of the paper. By contrast,

(17) The idea of this paper is to show that…

seems more acceptable than (14) to (16), which supports the idea that this construction can indeed fill a functional gap. In a number of ways, then, the idea is construction differs from and complements the older verbal intention layers.

Section 2.1 will discuss previous quantitative approaches to verbal intention markers, and in the remainder of this paper we will concentrate on the construction the idea + be with a clausal complement. To anticipate our conclusions, we will argue that this construction is
indeed joining or has to some extent joined the intention-marking domain and displays some of the typical patterns of change generally attributed to constructions in grammaticalization:

- A change in meaning and the acquisition of a more grammatical function (i.e. divergence in the sense of Hopper 1991),
- some degree of specialization in the direction of that emerging grammatical function,
- an increase in discourse frequency as an epiphenomenon or even a driving force of grammaticalization (cf. Bybee & Thompson 1997; Krug 2000; Bybee 2010; Bybee 2011),
- a certain loss of syntagmatic and morphological variability alongside some phonological changes which in combination point to greater bondedness.

2.1 Previous quantitative approaches to intention markers with verbal nuclei

Both the constructions want to and (BE) going to underwent a dramatic increase in discourse frequency between 1850 and 1950, while the preceding 200 years had seen more moderate frequency gains (cf. Figure 1, adapted from Krug 2000). With a gradual spread that later gains pace, the increase follows roughly the classic S-curve pattern. For want, we observe a clear case of divergence: After the decline of impersonal constructions meaning ‘lack’ that had been available in Middle English times, from about 1700 grammatical uses of want – i.e. those with to-infinitives expressing necessity, volition and intention – gained ground besides the older lexical meanings of ‘lack’ and ‘need’, which took nominal direct objects (for details see Krug 2002). Similarly, going to diverges from the original meaning of movement and directionality to intention and futurity, as has been widely documented in the literature (e.g. Hopper & Traugott 2003: Ch.1).
Mair (2004) proposes two categories for different kinds of grammaticalization processes: “dynamic” grammaticalization characterized by dramatic increases in frequency as the construction undergoes a change in function, and a “static” one which is less cataclysmic and characterized by the sporadic use of formerly fully lexical items to transport a grammatical function. An example of dynamic grammaticalization can not only be observed in the ARCHER corpus (Figure 1 above), but also in Mair’s data from the digital quotation base of the OED (Figure 2). It is evident that the frequency of going to in these data steadily increases from the early seventeenth century to the mid 1800s in a semi-static process. While the evolution of future uses for be going to began as early as the fifteenth century, as Danchev & Kytö (1994) show, the early changes were apparently essentially restricted to syntax and semantics but the construction did not display remarkable frequency gains. From the middle of the nineteenth century, however, the process becomes much more dynamic, in Mair’s terminology. The increase from 1600 to 1850 is sizeable, it gains pace in the late nineteenth century, but the quantum leap is observable only during the early twentieth century. The sudden rise in frequency, which Mair interprets as an indicator of grammaticalization, is accompanied by the emergence of contracted forms (gonna), which probably occurred in the spoken mode somewhat earlier, i.e. before such forms made it into written records. Bybee (2011) interprets such univerbations as an automatic, yet driving force in grammaticalization. In both verbs, then, the forms under investigation undergo semantic and syntactic divergence in the early stages where the increase in frequency is moderate, and they specialize phonologically when the frequencies of the respective constructions soar.
While Mair (2004) takes a diachronic corpus-linguistic approach to show long-term grammaticalization as a diachronic process, other methods from the field of sociolinguistics can be used to shed light on more recent changes, especially if a short-term development is suspected and we want to make statements about change in progress. Thus Krug (2000) looks at the frequency of the contracted forms wanna and gonna in six age-groups within the spoken part of the BNC. Although caution is required in interpreting orthographic evidence as phonological change, Figure 3 suggests that there is an unbroken quasi-linear increase of the frequency of wanna from the older to the younger speakers, and a parallel increase in gonna, which levels off for the two youngest age groups, though. In these apparent-time studies, then, younger speakers appear as innovators and the increasing frequencies of the two contracted forms support the hypothesis that univerbation in British English, as one indicator of grammaticalization, is still ongoing for want to, but has reached a plateau stage for going to, where the remaining uncontracted 30% seem to serve for stylistic, regional and social differentiation. Only further studies and time can tell whether this is a transitory or longer-term situation.

Figure 2. Frequency (per 10,000 quotes) of going to and gonna in OED quotes (Mair 2004: 129).
Figure 3. Going to and want to in an apparent time study based on the spoken BNC (Krug 2000: 175):
Percentage of contracted forms.

2.2 The idea is and related nominal constructions

While the previous sections concentrated on the relationship between idea is vis-à-vis intention markers with verbal nuclei like want to and be going to, the present section attempts to shed some light on the relationship between the idea is and related constructions with nominal nuclei. Consider the following example from the BNC, which is first given in its original context with a that-complement and subsequently transformed into a semantically equivalent to-clause.

(18) Here they re-live the life of a pioneer in the wilderness, when the nearest living soul was 20 miles away. This is not cheap: housing costs eat up 44% of a typical wage-earner’s income, as opposed to 14% in the 1950s. A few brave souls are about to try reversing the trend, through ‘co-housing’. The idea is that people band together to buy property, then design and build a cluster of single-family homes. (BNC ABH 582)

(18)’ The idea is (for people) to band together to buy property.

Structurally, the idea is displays parallels with constructions like the thing is (on which see Boyland 1996) and the point is, which also have a Subject-Verb-Complement structure (SVC in the terminology of Quirk et al. 1985) and where the complement also consists of a clause. Semantically, the intention-marking function of idea is is thrown into relief when seen in
comparison to other focusing constructions like *the thing/point is* and pseudo-cleft sentences, as in (19) and (20) below:

(19) $[\text{The thing / point}]_S [\text{is}]_V [(\text{that}) \text{ people band together to buy new property}]_C$

(20) $[\text{What’s important}]_S [\text{is}]_V [(\text{that}) \text{ people band together to buy new property}]_C$

Like *thing* and *point* in the preceding two examples, *idea* in (18) is semantically empty, and like *point*, *idea* has lost its former lexical meaning and assumed grammatical meaning in the given context. However, unlike *thing is* and *point is* (19) or the pseudo-cleft sentence (20), which function solely as focusing constructions, *idea is* has two grammatical functions: that of structuring discourse (i.e. focusing) and that of marking intention, while avoiding the statement of an overt agent (cf. the discussion following example (9) above and that in section 4.4 below):

(21) The idea is to band together to buy new property.

There are a number of structurally parallel constructions which prima facie had a better starting point to become the dominant nominal intention marker, as the meaning of the nouns was ‘intention’ from the moment they entered the English language:

(22) *The intention is* to band together to buy new property.

(23) *The aim is* to band together to buy new property.

(24) *The plan is* to band together to buy new property.

(25) *The purpose is* to band together to buy new property.

Unlike (22) to (25), *the idea is* retains a considerable degree of ambiguity regarding the grammatical meaning. But while this may seem to be a drawback at first sight, general meaning and polysemy of items make them more versatile and thus even increase their likelihood to enter grammaticalization paths (Heine 1993: 30; Bybee et al. 1994: 8).

Even though the case of *the idea is* with a following complement clause is still very much in flux, we see clear evidence of ongoing grammaticalization: First, as has been pointed out above, *idea* has lost its lexical meaning in this construction, has taken on grammatical meaning (intention) and functions at the same time as a focusing device.

Second, morphosyntactically, *idea is* displays a restricted productivity in several ways. In (26) *idea* is pluralized, because there are two intended courses of action each of which could be conceived as an individual idea. The result is an odd sentence bordering on the ungrammatical:

(26) ??The ideas are (were) first to band together and to buy new property later.
Much more natural and idiomatic is the same sentence employing idea is, given as (26)’ below. Pluralization thus appears to be virtually impossible when idea is functions as an intention marker:

(26)’ The idea is (was) first to band together and to buy new property later.

Third, the grammatical construction is not fully compatible with different determiners and varying degrees of definiteness, as is shown by example (27). While the lexical source noun idea readily combines with this, an or some, these determiners sound odd with idea is + to-clause and, crucially, if they replace the definite article, intention marking is lost:

(27) ??This/An/Some idea is to band together…

The only possibility if one wants to retain intention marking with different determiners is to replace the definite article with possessives, which may figure in the form of pronouns, as in (27)’:

(27)’ Our/Peter’s idea is to band together…

Fourth, the intention reading is weakened considerably when the full verb exist replaces be, as in (28); other copulas seem to result in ungrammatical structures, as in (29):

(28) ??The idea exists to band together…

(29) *The idea seems to band together…

In summary, each contributing syntactic constituent of the construction (article, noun, copula) allows no or only a very restricted set of paradigmatic choices. In other words, the construction the idea is shows a considerable degree of idiomaticity and is thus clearly not fully compositional. Instead, it has lost in syntagmatic variability and displays fixation and increased bondedness and thus exhibits a number of Lehmann’s (1995: 164) parameters of grammaticalization. Overall, there are also strong signs that Hopper’s (1991: 22) principle of decategorialization applies, most obviously so to idea:

Forms undergoing grammaticalization tend to lose or neutralize the morphological markers and syntactic privileges characteristic of the full categories Noun and Verb.

Thus far there is, then, strong qualitative evidence suggesting that idea is + to-clause qualifies as a construction under grammaticalization:

- It operates in a local context as a relatively fixed phrase (with a small number of alternative fixed phrases) that displays a certain internal order of elements.
In this construction, *idea* appears to be in the process of dissociation from its source noun on the levels of semantics and morphosyntax.

In the next section, quantitative corpus-linguistic data will be presented to check whether a grammaticalization claim is also supported by frequency gains similar to those found by Krug (2000) and Mair (2004) for other English intention-marking constructions.

3 *Idea is* and *idea was*: Synchronic and diachronic discourse frequency

3.1 Diachronic developments

This section capitalizes on the recent emphasis placed by grammaticalization theorists on the role of frequency and frequency gains as catalytic forces in grammaticalization. One of the major advocates is Joan Bybee (e.g. 2003, 2005, 2010, 2011):

> Frequency is not just a result of grammaticization, it is also a primary contributor to the process, an active force in instigating the changes that occur in grammaticization. ... I will argue for a new definition of grammaticization, one which recognizes the crucial role of repetition in grammaticization and characterizes it as the process by which a frequently-used sequence of words or morphemes becomes automated as a single processing unit. (2003: 602-3)

The first corpus we look at systematically is the *Time Magazine Corpus of American English* (Davies 2007-2010), which comprises 100 million words from texts dating from 1923-2006. A search for the strings *idea is* and *idea was* was conducted to see if an increase in discourse frequency over time can be shown, as this could be taken as an indicator of progressing entrenchment and grammaticalization. For both phrases, the following analyses were conducted:

- All occurrences of *idea is* and *idea was*.
- Occurrences followed by a *to*-clause.
- Occurrences followed by a *that*-clause.

The grammaticalizing contexts of interest are those followed by *to* and *that*-clauses since they typically convey a future or intended action. Thus the exploration of grammaticalization as a diachronic process is done by tracing the frequency of these constructions through time. On the other hand, we look not only at the increase in frequency of these constructions, but also at the extent of specialization they undergo. Therefore we not only ask “How frequent are they?”, but also “How much competition do they (still) face from other constructions involving *idea*?” The first question is answered by measuring normalized discourse
frequency, the second by dividing the frequency of *to/that* complement clauses by the overall frequency of *idea is* to arrive at an index that describes the degree of isomorphism, i.e. a measure (first introduced in Krug 2005) indicating the degree of specialization of the constructions *the idea is* and *the idea was*.

We propose that synchronically both frequency and specialization (observable in its function, the degree of isomorphism, i.e. the share of grammatical constructions seen against all *idea is* and *idea was* sequences) are relevant parameters for the assessment of overall grammaticalization. We therefore use an integrated grammaticalization score (labelled GRAL index in Table 1), which is the product of both. Using the example of *the idea is to/that*, the GRAL index is mathematically defined by the following formula, where \( F_{idea \ is \ to/that} \) is the frequency of *the idea is* followed by *to- or that*-complement clauses, and \( F_{IDEA \ IS} \) is the total frequency of the higher-level, more general construction to which it belongs (here: *the idea is*):

\[
GRAL_{idea \ is \ to/that} = F_{idea \ is \ to/that} \times \frac{F_{idea \ is \ to/that}}{F_{IDEA \ IS}} = \left( \frac{F_{idea \ is \ to/that}}{F_{IDEA \ IS}} \right)^2
\]

Using this index, greater specialization and higher frequency will result in higher scores of grammaticalization (see above).

Table 1 shows that the overall text frequencies of *idea is* and *idea was* increase significantly through the 20th century: While the incidence of *idea is* triples, *idea was* doubles. The most striking gains are to be found for *to*-clauses, which increase sixfold and fourfold in present and past contexts, respectively. In a very parallel fashion, for both *idea is* and *idea was*, the share of following *to*-complements rises from about one in five (21% and 23%) to almost one in two (44% in either context). *That*-complements, by contrast, display a more moderate increase in discourse frequency and are stable or even decline in relative terms. It is therefore essentially the absolute and relative increase in *to*-complements that is responsible for the significant increase of the grammaticalization index GRAL in *idea is/was* constructions.

Compared to the increases in discourse frequency, the increases in the rate of isomorphism (i.e. relative shares of *to- and that*-complements) show greater fluctuations. After the 1980s, for which our data indicate a halt or even temporary reversal of an otherwise unbroken development, the combined proportion of *to/that*-complementation increases more markedly, and consequently the same is true for the composite index of grammaticalization. Overall, our data suggest, then, that both *idea is* and *idea was* – after a gradual increase in frequency over the twentieth century – entered a dynamic grammaticalization phase during
the 1990s. For the past tense form the idea was to/that, however, the patterns are generally less pronounced (cf. Figure 4).

<table>
<thead>
<tr>
<th>Time period</th>
<th>total</th>
<th>total per 10 million words</th>
<th>to-clause per 10 million words</th>
<th>that-clause (of total)</th>
<th>to-and-that-clauses per 10 million words</th>
<th>to-and-that-clauses (of total)</th>
<th>GRAL index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920s</td>
<td>39</td>
<td>50.9</td>
<td>11.7 (23%)</td>
<td>11.7 (23%)</td>
<td>23.4 (46%)</td>
<td>10.8</td>
<td></td>
</tr>
<tr>
<td>1930s</td>
<td>49</td>
<td>30.9</td>
<td>8.2 (27%)</td>
<td>4.4 (14%)</td>
<td>12.6 (41%)</td>
<td>5.1</td>
<td></td>
</tr>
<tr>
<td>1940s</td>
<td>55</td>
<td>35.6</td>
<td>13.6 (38%)</td>
<td>3.2 (9%)</td>
<td>16.8 (47%)</td>
<td>7.9</td>
<td></td>
</tr>
<tr>
<td>1950s</td>
<td>87</td>
<td>51.8</td>
<td>19.1 (37%)</td>
<td>4.8 (9%)</td>
<td>23.9 (46%)</td>
<td>11.0</td>
<td></td>
</tr>
<tr>
<td>1960s</td>
<td>112</td>
<td>69.6</td>
<td>29.8 (43%)</td>
<td>8.7 (13%)</td>
<td>38.5 (55%)</td>
<td>21.3</td>
<td></td>
</tr>
<tr>
<td>1970s</td>
<td>113</td>
<td>83.8</td>
<td>27.4 (33%)</td>
<td>5.9 (17%)</td>
<td>33.3 (40%)</td>
<td>13.2</td>
<td></td>
</tr>
<tr>
<td>1980s</td>
<td>80</td>
<td>70.3</td>
<td>18.5 (26%)</td>
<td>7.0 (10%)</td>
<td>25.5 (36%)</td>
<td>9.2</td>
<td></td>
</tr>
<tr>
<td>1990s</td>
<td>112</td>
<td>115.0</td>
<td>46.2 (40%)</td>
<td>15.4 (13%)</td>
<td>61.6 (54%)</td>
<td>33.0</td>
<td></td>
</tr>
<tr>
<td>2000s</td>
<td>95</td>
<td>147.8</td>
<td>65.3 (44%)</td>
<td>32.7 (22%)</td>
<td>98.0 (66%)</td>
<td>65.0</td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Idea is and idea was in the Time Magazine Corpus (1920s to 2000s): Overall frequency and proportion of to- and that-complementation vs. other types of complementation.

Figure 4. Idea is to/that and idea was to/that in the Time Magazine Corpus (1920s-2000s): Development of the index of grammaticalization composed of discourse frequency and the degree of isomorphism.
Figure 5 shows the individual developments of the grammaticalization index for each of the four constructions, *idea is to*, *idea was to*, *idea is that* and *idea was that*. It reflects the fact that, compared to the past tense forms, the generally higher frequencies of the present tense forms lead to higher grammaticalization scores (cf. Table 1). Furthermore, *to*-complementation becomes far more frequent than *that*-complementation after the 1920s.

![Graph](image)

*Figure 5. Idea is and idea was in the Time Magazine Corpus (1920s-2000s): Development of the index of grammaticalization composed of discourse frequency and the degree of isomorphism, treated separately for complementation with to- and that-clauses.*

We briefly return to a more detailed inspection of the changing degree of isomorphism, i.e. the changing proportion of *idea is/was + to-* or *that*-clause complementation relative to the total number of constructions including *idea is/was*. This is shown in Figure 6.
Figure 6. Proportion of idea is/was + to/that-clause complementation relative to all forms of idea is/was in the Time Magazine Corpus (1920s to 2000s). 95% confidence intervals based on Wilson-scores.

The proportion for each decade is given with a 95% confidence interval based on Wilson scores. As mentioned above, the results for the 1970s and 1980s are somewhat puzzling – in fact, they may well be artifacts of text sampling. Nevertheless the increasing overall trend is clear. We decided to compare three groups: (1) 1920s/30s/40s, (2) 1950s/60s, and (3) 1990s/2000s, skipping the ‘problematic’ 1970s/80s. The (raw) frequencies of alternative forms are given in Table 2.

<table>
<thead>
<tr>
<th>Time period</th>
<th>idea is/was: type of complement</th>
<th>to/that-clause</th>
<th>other</th>
</tr>
</thead>
<tbody>
<tr>
<td>1920s - 1940s</td>
<td>141</td>
<td>207</td>
<td></td>
</tr>
<tr>
<td>1950s - 1960s</td>
<td>189</td>
<td>193</td>
<td></td>
</tr>
<tr>
<td>1990s - 2000s</td>
<td>186</td>
<td>133</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Frequencies of to/that-clause complementation of idea is/was in the Time Magazine Corpus, comparing three selected time periods.

Carrying out Newcombe-Wilson tests (see Appendix 1 in Aarts, Close and Wallis, this volume) at \( p < 0.05 \), we found significant differences between any two of these groups with regard to the proportion of to/that-clause complementation. In the early 20th century to/that-complementation was a (very sizeable) minority form (40.5%), in the middle of the century it accounted for almost exactly half of all forms (49.5%), and in the late 20th and early 21st century most cases of idea is/was were complemented in this way (58.3%).
3.2 Variation with other constructions of the type [the NP + BE] in present-day English

A second approach is to look at idea is/was in comparison to other constructions that can express intention by means of a construction of the type the NP + be in contemporary English. Our question is to what extent idea is/was has grammaticalized in comparison to other related constructions. Adopting the same methodology as for the Time Magazine Corpus above, we measure this by counting all occurrences of a given construction in the spoken BNC, and by determining what proportion of these is followed by a to- or that-clause relative to all occurrences of the sequences the idea is etc. Unlike in the Time Magazine Corpus, where every decade has a different corpus size, there is no need to normalize the absolute frequencies for the BNC analysis because all data stem from the same corpus containing about 10.4 million words. Again we multiply the frequency of the to/that-clause complements by that proportion to arrive at a grammaticalization index (cf. Table 3). In addition to idea is/was, the following four constructions were analyzed, all of which contain what Schmid (2000: 4) classifies as ‘shell nouns’, i.e. members of “an open-ended functionally-defined class of abstract nouns that have, to varying degrees, the potential for being used as conceptual shells for complex, proposition-like pieces of information”:

- the intention is/was
- the aim is/was
- the plan is/was
- the purpose is/was

---

2 We have restricted our investigation to the two most common constructions, i.e. the idea is/was. There are also 19 instances in the spoken BNC of the idea has been, none of which, however, is relevant to the constructions under investigation: 17 are passives, i.e. followed by a past participle; the remaining two are the idea has been around. Non-finite forms like the idea being (of which there are 6 in the spoken BNC) have also been excluded.
Table 3. Determining degrees of grammaticalization in different constructions of the type the + NP + BE in the spoken BNC (c. 10.4m words).

Table 3 shows that idea is is by far the most frequent, followed by idea was. All four alternative constructions are much less frequent than idea is/was, which suggests that they are largely restricted to written language. Translated into our grammaticalization index, both constructions with idea rank highest, with intention is being almost on a par with idea was. These three constructions are followed by an intermediate group comprising aim is, aim was, intention was and plan is, while plan was, purpose is and purpose was score very low both in frequency and their degree of isomorphism and thus have very low grammaticalization indexes. These are calculated as in the previous section: The sum of the absolute occurrences of the grammatical constructions (here: the NP + to/that complement clauses) is multiplied with the degree of isomorphism (i.e. the proportion of the NP + to/that complement clauses measured against the total frequency of the higher-level construction, here: the NP is/was).

Figure 7 illustrates that – with the exception of intention – constructions using the same shell noun within the NP are immediate neighbours in terms of rank, with the present tense form invariably in the lead. This latter fact comes as no surprise given that the present tense is the unmarked option in natural language production. More surprising is the fairly categorical division between the different shell nouns. Most striking, however, is the immense gap between idea is and the remaining constructions. Idea is towers above all other options among intention-marking constructions using NPs in spoken discourse. Compared to the other structurally related intention markers investigated here, idea is/was is not only more frequent in absolute terms, but also more highly grammaticalized according to its grammaticalization scores.

---

3 Cross-checks in the BNC revealed that the text frequencies of these constructions are actually considerably higher in the written part than in the spoken part of the corpus.
A brief discussion of the point is and the thing is vis-à-vis the constructions analyzed here seems indicated, since these two common constructions were mentioned as related in section 2.2 above. They were excluded from the present quantitative analysis because their prime function is not one of marking intention. From a constructionist and grammaticalization point of view it is interesting to note that the thing is and the point is are essentially non-intentional and hardly ever take a to-infinitival complement. Both favour main clause and that-clause complements, which contrasts markedly with unambiguous intention markers like the plan/aim/intention is and their common to-complements (about 30%, see Table 3 for quantitative detail). It is important to note that to-complement clauses themselves derive historically from purposive – i.e. intention-marking – constructions (Haspelmath 1998; Hopper & Traugott 2003: 3). We can conclude, therefore, that both the noun (e.g. thing vs. plan, idea) and the complement (to vs. that or zero) retain traces of their older meaning – as predicted by Hopper’s (1991: 22) principle of persistence. All components thus contribute to the overall meaning of the entire construction and to the development of its selection restrictions. This holds not only across different shell nouns like thing and plan but also for different constructional strategies of idea, which is more likely to be non-intention-marking with that or main clause complements, as in (30), where complementizer that is optional:

(30) So the idea is that you might consider that the letter X is made up of two features.

(BNC JSY 329)

---

4 Idea is attested (OED s.v. idea, no. 2, 4) with ‘plan/aim/intention’ meanings since the late 16th century.
Example (30) demonstrates that *idea is* followed by a *that*-clause may function similarly to the focusing construction *thing is* and merely highlight the following proposition, albeit in a slightly more formal register. Not all *that*-clauses, however, are non-intention-marking, which is why they were included in the present analysis. Compare:

(30)’ So the idea is that you (might/should) consider the letter X, not Y.

In summary, semantically, there seems to be a cline of increasing intention marking, where *the idea is* is situated roughly halfway between virtually intention-free *the thing is* and clear intention markers like *the plan is*:

(31)  the purpose/plan/intention/aim is > the idea is > the point is > the thing is

The figures above underpin the qualitative observations made in section 2.2, where we hypothesized that specific semantic source meanings may be an obstacle to dramatic frequency gains, while more general meanings can more easily enter into grammaticalization paths in domains of which they were originally only marginal members. *The idea is* can thus be interpreted as having undergone paradigmaticization, one of Lehmann’s (1995: 164) parameters of grammaticalization by which an item that participated loosely in a semantic field becomes a member of a smaller integrated paradigm. At the same time, however, *the idea is* participates in at least one more grammatical paradigm, that of focusing constructions.

To summarize thus far in more concrete terms: *plan, intention, aim* and *purpose* are unambiguous intention markers but have remained rather infrequent. The polysemous item *idea*, by contrast, with its more generalized meanings has filled large parts of a functional gap that other nouns would have been more likely to fill from a semantic perspective alone; at the same time, *idea is* has made inroads into the territory of focusing constructions and has occupied much of the space that combines the two functions of intention marking and discourse organization.

4. The phonology of *idea*

4.1 Hypotheses and citation form

Phonological reduction at a word boundary is a typical concomitant of grammaticalization and often the consequence of increased discourse frequency. It may result in a functional differentiation between full and auxiliary verb. Compare:

(32) It’s *going to* rain.  ~  It’s *gonna* rain.

(33) I’m *going to* London. vs.  *I’m gonna* London.
This is not, however, relevant to the grammaticalization of the intention marker under investigation:

(34) The idea is to go to London. ~ The idea’s to go to London.

(35) Your idea is a very good one. ~ Your idea’s a very good one.

Phonological reduction by cliticization of the copula in idea is as in (34) and (35) is unobtrusive for the hearer, difficult for the transcriber, and certainly not restricted to the intention marker. Considering the fact that 49 occurrences of directly adjacent to- and that-clauses after the idea is contrast with only one clausal complement with a contracted form in the BNC, quoted as (36) below, it seems that, if anything, the share of the full form, rather than that of the clitic, is higher than would be expected, even if one allows for difficulties and inconsistencies in transcription.

(36) The idea’s to teach managers to spot the symptoms of stress early. (BNC K6B 194)

The idea is thus seems parallel to the thing is constructions in avoiding cliticization of is. One might speculate that phonological fusion or reduction affects the phrase the idea is at the first rather than second word boundary, i.e. by reducing the definite article. This would be another parallel with thing is constructions, where article omission is commonly found in speech and even sometimes figures in informal writing (including books, not just internet communication):

(37) Thing is, you’ve got to grab attention with just one phrase as punters rush on by.

(BNC FAS 315)

Clearly, however, omission of the definite article is not pervasive with intention-marking the idea is. And phonetic reduction of the definite article is not investigated here, because rather than grammaticalization being the only potential explanatory factor, there are phonetic factors (e.g. hiatus avoidance) that could equally explain them.

In the phonological investigation undertaken for this section, then, we shall look at formal variation at the level of intonational phonology. More specifically, we will argue that idea often does not carry its main stress on the second syllable, and that this may be a formal low-level change connected to a functional shift towards the more grammatical meaning of intention-marking.

---

5 The thing is, however, is a rather special case because of the word-initial homorganic fricatives /θ/ and /ð/, which, given further reduction and deletion of /ð/ in the definite article, will readily fuse into a single sound, possibly lengthened. Thus, the underlying structure could even be argued to remain phonologically intact.
We first turn to the two most widely used pronunciation dictionaries of contemporary English, *Longman's Pronunciation Dictionary* (hereafter: LPD, Wells 2008, 2000, 1990) and the *Cambridge English Pronouncing Dictionary* (hereafter: CPD, Jones 2006, 1997). Both of these clearly place the lexical stress of *idea* on the second syllable. The first edition of LPD (1990) lists a possible secondary stress for the first syllable in General American, the second and third editions (2000, 2008) report that 14% of the members of a 1993 poll panel for American English preferred *idea* with first-syllable stress (cf. Figure 8). Crucially, in the examples presented to the panellists *idea* was in sentence-final (i.e. prepausal) position (Shitara 1993: 201), i.e. in a position where the construction we investigate is not possible. It would thus seem that the finding by Wells indicates variability of lexical stress in near-citation form, possibly pointing to a general difference between American and British English.

![Figure 8. Longman Pronunciation Dictionary (Wells 2008): s.v. idea.](image)

![Figure 9. Cambridge Pronouncing Dictionary, 17th ed. (Jones 2006): s.v. idea.](image)

CPD (1997 and 2006, Figure 9) reports a stress shift in the word *idea* for British English in contexts where a stress clash is avoided. Hints at occasional first-syllable stress due to rhythmic motivations can be found earlier for both British English (Jones 1969) and American English (Kenyon & Knott 1953). Both sources, however, motivate this variation in stress not functionally but phonologically.

---

6 Reproduced with permission from Pearson Longman.

7 Reproduced with permission from Cambridge University Press.
4.2 Stress-patterns of *idea* in ICE-GB

We next looked at the spoken part of the British component of the *International Corpus of English* (hereafter: ICE-GB) to approach the phonology of *idea* empirically. This corpus consists of 600,000 words and was compiled no earlier than 1989. It has the advantage of accessible sound files – this was important as the corpus is not phonologically transcribed and an original auditory analysis was therefore necessary.

Of a total number of 144 cases of *idea*, 133 were considered for analysis. In the remaining 11 cases, *idea* was masked by overlapping utterances or noise. We speculated that some raters might not be able to distinguish between first-syllable and level stress and therefore introduced a conflated category labelled "early stress".\(^8\) This combination of first-syllable stress and level stress seems consistent with both perceptual factors and a cognitive approach to categorization, i.e. one in terms of absence or presence of the prototype.

Table 4 gives the frequencies of all stress patterns for *idea* across all observations, for *the idea is/was*, and for all observations except *the idea is/was*.

<table>
<thead>
<tr>
<th></th>
<th>[1st]</th>
<th>[level]</th>
<th>early (=1st + level)</th>
<th>2nd</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>15/133</td>
<td>28/133</td>
<td>43/133</td>
<td>90/133</td>
</tr>
<tr>
<td></td>
<td>11.3%</td>
<td>21.1%</td>
<td>32.3%</td>
<td>67.7%</td>
</tr>
<tr>
<td><em>the idea is/was</em></td>
<td>2/7</td>
<td>4/7</td>
<td>6/7</td>
<td>1/7</td>
</tr>
<tr>
<td></td>
<td>28.6%</td>
<td>57.1%</td>
<td>85.7%</td>
<td>14.3%</td>
</tr>
<tr>
<td>all others</td>
<td>13/126</td>
<td>24/126</td>
<td>37/126</td>
<td>89/126</td>
</tr>
<tr>
<td></td>
<td>10.3%</td>
<td>19.0%</td>
<td>29.4%</td>
<td>70.6%</td>
</tr>
</tbody>
</table>

*Table 4. Stress pattern of *idea* in ICE-GB.*

Across all observations of *idea*, about two thirds (67.7%) carry stress on the second syllable, about one third (32.3%) fall into the early stress category. Of these observations, one third (11.3% of the total) are clearly stressed on the first syllable. The pattern for *the idea is/was* is in marked contrast to this general picture, in fact it is more than the inverse: here, six out of seven observations are characterized by early stress, and exactly one third of these by first-syllable stress. Comparing this construction to the remaining observations instead of the grand total brings out the contrast even more clearly: the proportion of first-syllable stress in *idea*

---

\(^8\) The authors were no exception: For us, level stress was very often sufficient to make a token so noticeably different from the citation form that the first spontaneous rating was “first-syllable” rather than “level”.

24
is/was is about 3 times higher than in the remaining cases. In the Appendix we present a logistic regression analysis in which we investigate several factors that can help predict early stress in idea, including the construction type.9

4.3 Qualitative analysis of idea is in the spoken ICE-GB

For this section we took a closer look at all idea is and idea was sequences in ICE-GB irrespective of what precedes them. The eight examples discussed below include five cases of the idea is, (38), (39), (40), (41) and (44), two occurrences of the idea was, (42) and (43), and one case of idea was with a preceding attributive adjective phrase (45).10 Six of the eight examples display level or first-syllable stress (i.e. early stress). Intriguingly, the two which display second-syllable stress differ functionally in that they do not convey intention-meaning. We will now discuss the individual examples qualitatively.

Sentence (38) conveys the hope that the magazine in question will appeal to various different contributors: It is the intention of the publishers to reach a wide contemporary audience, i.e. we interpret the sentence that way, although the agent is not overtly stated. In fact, intention/futurity is doubly marked: through the idea is and through going to.

Semantically, the idea is is roughly equivalent to we intend / it is intended or we hope:

(38) But the idea is that uhm it’s a magazine that’s going to be a kind of forum for a lot of the things that are happening. (ICE-GB S1A-096#089 – first-syllable stress in idea)

In (39), the coordinated clause (“and the idea is...”) directly builds upon the first clause: The two theatres are programmed together in a certain way with the aim, plan or intention for one of them to attract a certain kind of company. The planner is most likely the understood subject (we) from the first sentence, which is not overtly stated as “our idea”:

(39) We are programming the two theatres together and the idea is that the Lyceum is a home for touring productions. (ICE-GB S1B-050#012 – first-syllable stress in idea)

While (40) is incomplete, its structure is best interpreted in terms of an unfinished that-complement clause following the idea is:

(40) no because the idea is that he’s... (ICE-GB S1A-098#295 – level stress in idea)

9 For applications of logistic regression in variationist research and a more detailed description of that statistical model than the one provided in the Appendix, see Schützler (2010, 2011).

10 The sentence given below as (45) was not classified as an idea is/was construction for Table 4 for the very reason that an adjective intervenes between the and idea. This is why the number of relevant cases in that table is seven, not eight.
Similarly to (38), example (41) is interesting because it features not only idea is but another VP that semantically reinforces the intention-marking function of the whole: to try implies the intention of attaining a goal, in this case the publication of a PhD thesis:

(41) so the idea is to try and publish it in a much more commercial publishers [sic!] than the normal. (ICE-GB S1A-066#123 – level stress in idea)

Sentence (42), paraphrased as (42)' and (42)'' below, shows the close relationship between intention markers, volition and deontic modality in its obligation sense:

(42) The idea was you know we were supposed to do all these graphs and stuff. (ICE-GB S1A-008#049 – level stress in idea)

(42)' The idea was for us to do all these graphs.

(42)'' Somebody wanted us to do all these graphs.

Unless there are coreferential subjects in intention-marking constructions (as, for example, in we wanted to do all these graphs), we are dealing with directives: ‘Somebody wants/intends for somebody else to do something’. Notice that the person(s) imposing the obligation is/are not given by the idea is construction, a strategy which is copied structurally by use of the passive we were supposed to do. Thus, the idea is construction is more vague and potentially less face-threatening (the people imposing the obligation could, after all, include the addressee) than the corresponding verbal constructions, which would need to specify who wants whom to do something and thus also who is in authority.

Deontic modality in (42) is made explicit by the modal construction be supposed to, but it need not be expressed that way. This can be seen in the paraphrase (42)' or example (43) below:

(43) The idea was for me to see the material and say well yes I can do it ... (ICE-GB S1A-060#012 – level stress in idea)

Interestingly, out of context (43) allows for the obligation reading (with somebody else wanting the speaker to see the material) or the intention reading (where the speaker intends to see the material). Again, therefore, the idea is construction is ambiguous in leaving the source of the intention open, which in this sentence even includes the speaker herself.

In example (44) the idea is features with a main clause, and is both structurally and functionally equivalent to the thing is or the point is (cf. the discussion of example (30) above):
(44) Well yes I mean *the idea is* I’m not interested really in uh museums more generally. (ICE-GB S1A-066#063 – second-syllable stress in *idea*)

Interestingly, the raters were for a long time uncertain whether to classify this example as second-syllable or level stress, with the final vote going to the former. In hindsight, the default stress pattern becomes at least partly explainable by this example’s function, which, we would argue, is focusing rather than intention-marking.

As in (41) the verb *try* is used in example (45), seemingly to reinforce the notion of intentionality:

(45) That’s why I thought *a good idea was* to try and apply sort of a model. (ICE-GB S1A-064#049 – second-syllable stress in *idea*)

Here, however, *a good idea* suggests that there were several options open to the speaker. Hence *idea* is used in a different function, and this is reflected both in the attributive use of *good* and in the fact that *idea* is stressed on the second syllable. Perhaps even more importantly, example (45) has *good* and *idea* in juxtaposition. First-syllable stress in this situation would result in stress clash, which would in turn contribute to the selection of second-syllable stress.

In summary, the two instances of *idea is/was* in ICE-GB that do not exhibit early stress can be motivated: (44) throws into relief a focusing function; and (45) is doubly exceptional with *a good idea* featuring an attributive adjective and an indefinite article. In other words, this example combines less syntagmatic fixation with the absence of intentional meaning. We would argue that this is no coincidence and submit therefore that early stress in *idea is/was* becomes more likely when the construction is actually used as an intention-marking device. In addition, all examples except (44) and (45) shift the agent out of focus. This turns *idea is* and *idea was* into a less specific intention marker than the older purely verbal constructions like *I intend(ed) to* or *I aim(ed) to*.

In conclusion, the grammaticalization of *idea is* would appear as something quite typically accompanied by a formal (here prosodic / phonological) change that happens later than the semantic changes, i.e. the bleaching of lexical meaning and the concomitant assumption of grammatical meaning. However, unlike other well-known morpho-phonological reduction processes that have been discussed in the literature on grammaticalization, here the formal change is not readily explained as a loss of phonological complexity or attrition (Lehmann 1985: 307). In the phrase *The idea is to...*, it makes no sense to assume that the stress pattern *the Idea is* should be any less complex or more economical than *the IDEA is*. However, if a secondary stress is introduced on the word *is* (or, by analogy, *was* in *the idea was*), this results
in the stress-pattern “unstressed – stressed – unstressed – stressed”: the Idea IS, which indeed appears much more regular than the iDEA is (in which idea is stressed on the second-syllable and all function words are unstressed). The whole phrase also becomes more prototypical in optimality-theoretic terms since a regular alternating stress-pattern is created (cf. Selkirk 1984: 37; Couper-Kuhlen 1986: 60; Nespor & Vogel 1989: 69, 82; Kager 1999: 161f.). It appears, then, that each individual component of the phrase The idea is gives up its independent status and stress pattern during the creation and amalgamation of a more fixed grammatical unit (as is also common in the creation of complex lexemes), within which the principle of rhythmic stress alternation is followed more closely than outside such units. This may be related to the reduction of the erstwhile canonical trisyllabic pronunciation of idea to a disyllabic one (cf. the entries from the pronouncing dictionaries quoted above), which parallels typical cases of phonological attrition in grammaticalization (cf. Lehmann 1995: 164). Further phonological factors will be discussed in the next section.

Let us try to integrate the present study into more general patterns of language change and typology. In the qualitative analysis, we have seen that the grammatical function of idea is was repeatedly reinforced by other markers in the context, notably by a second intention or obligation marker (try, be going to, be supposed to). We interpret this as a further indication that we are dealing with the early stages of grammaticalization, as new meanings tend to evolve in heavily redundant contexts (e.g. Traugott 2000: 289f). If we adopt a bird’s eye perspective, finally, this English innovation in the domain of intention marking seems to fit into a more general typological and historical pattern: English has repeatedly developed (certainly when compared to German) into a more ‘loose-fit’ language, i.e. one in which hearers have to invest a good deal of interpreting and processing effort because semantic roles are not mapped consistently onto certain syntactic positions, and because clause boundaries and voice distinctions have become blurred. Compare such notorious examples as This tent sleeps five people; Tomorrow will be cloudy; The book sells; I believe him to be liar (Hawkins 1986; Mair 1990; König & Gast 2009). In a similar fashion, the syntactic subject (i.e. the idea) of the construction The idea is to do X is not the agent. Worse for the hearer, the construction leaves the source of the intention unspecified. It has to be retrieved from the context or may ultimately remain non-recoverable.

4.4 Rates of intrusive /r/ in idea is and idea of

We propose that a formal change which constructions undergo in conjunction with processes of grammaticalization may essentially apply at any level. Thus, instead of morphological or
phonological attrition we have observed a tendency for grammaticalizing contexts to shift the stress of idea to an earlier position in order to achieve a more regular and in that specific context more optimal structure. We will now look at another general process in the high-frequency phrase idea is, namely intrusive /rl/. When we analyzed rates of intrusive /rl/ in our data, we found that the sequence idea of also scores high on this feature. This is in fact even more frequent than idea is. We therefore treat idea is and idea of separately in this section.

The attraction of higher rates of intrusive /rl/ to grammaticalized or grammaticalizing forms has, to the best of our knowledge, not been demonstrated or proposed before and may at first appear to be a rather speculative notion. However, it can be argued that liaison between idea and is (and idea of), apart from facilitating pronunciation, may also bind the phrase more tightly together. This roughly corresponds to coalescence leading to greater bondedness in the sense of Lehmann (1995: 164). While in this case the coalescing elements form a tighter articulatory unit, they still seem to be two separate words. However, it is precisely the intrusive /rl/ which obscures the former word boundary as it belongs to neither of the words it links. It is therefore more appropriate to speak of a special type of coalescence.

Perhaps this aspect can contribute to the discussion on the boundary between lexicalization and grammaticalization (on which see Brinton & Traugott 2005) because the above considerations lend further support to the hypothesis made in the previous section: The component items lose their independence in the creation of the intention-marking idea-construction, whose internal structure is optimized, very much like in the creation of complex lexemes. We would argue, however, that we are indeed dealing with a case of grammaticalization, not lexicalization. Firstly, intention marking is a grammatical meaning, so that from an onomasiological perspective it belongs to the grammar of English (cf. Krug 2011). The same can be said of modal adverbs like perhaps or possibly, which mark epistemic modality, but which in our view belong more marginally to the domain of grammar. Secondly and more importantly, unlike modal particles, the idea is construction exhibits (when marking intention) a higher degree of clausal integration with its typical complements, to- and that-clauses (cf. Hopper & Traugott 2003: Ch. 7; Brinton & Traugott: Chs. 2 and 5).

In ICE-GB, the data indeed suggest a correlation of construction type and the rate of intrusive /rl/. As Table 5 shows, the rates for intrusive /rl/ are far higher for two high-frequency strings — idea of (22 out of 38 cases) and idea is (3 out of 5) — than for the remaining non-prepausal idea examples followed by any other element beginning in a vowel (2 out of 8).¹¹

This suggests that high discourse frequency of adjacent items triggers not only erosive processes at word boundaries as in I have > I’ve or is not > isn’t (Krug 1998: 289-298); in

¹¹ These were idea followed by the words a, actually (twice), and (twice), as, either, and if.
addition, such figures support the hypothesis that sandhi phenomena in general apply to high-frequency sequences more frequently (cf. Kaisse 1985).\(^\text{12}\) According to recent definitions of grammaticalization by Bybee (2010; 2011) cited in sections 2 and 3 above, discourse frequency is both a force and indicator of grammaticalization. Our figures seem to support this position as the two most frequent (and in this definition most entrenched and grammaticalized) *idea*-constructions attract intrusive /r/ at a likelihood of around 60%, while other liaison sites on average actually realize liaison at less than half that rate.

<table>
<thead>
<tr>
<th>linking context</th>
<th>sites</th>
<th>/r/</th>
<th>liaison rate</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>idea is</em></td>
<td>5</td>
<td>3</td>
<td>60%</td>
</tr>
<tr>
<td><em>idea of</em></td>
<td>38</td>
<td>22</td>
<td>58%</td>
</tr>
<tr>
<td><em>idea</em> + other word beginning with vowel</td>
<td>8</td>
<td>2</td>
<td>25%</td>
</tr>
</tbody>
</table>

*Table 5. Intrusive /r/ by construction in the spoken ICE-GB.*

It is true that the absolute numbers of occurrences are low, in particular for *idea is*, and this at least partially accounts for the fact that the difference between *idea is/of* and other potential liaison sites involving *idea* fails to reach statistical significance at the 5% level in ICE-GB. Notice, however, that *idea is* sequences account for a high share of the remaining potential liaison sites after *idea of* is taken out of the count. We would be surprised if both did not account for the majority of liaison sites in larger corpora too, and a preliminary analysis (ignoring *h*-initial words) of the spoken BNC indeed confirms this intuition (see Table 6). Unfortunately, the spoken BNC cannot be analysed acoustically, but 70% of about 1,000 potential liaison sites are *idea of* or *idea is*.

<table>
<thead>
<tr>
<th>linking context</th>
<th>sites</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>idea is</em>(^\text{13})</td>
<td>107</td>
</tr>
<tr>
<td><em>idea of</em></td>
<td>590</td>
</tr>
<tr>
<td><em>idea</em> + other word beginning with vowel</td>
<td>304</td>
</tr>
</tbody>
</table>

*Table 6. Liaison sites in the spoken BNC: idea + word beginning with vowel (orthography only).*

\(^\text{12}\) Bybee (2001) makes a related point for liaison in modern French (as in *je vais aller*). This, however, is essentially a conservative feature that is retained due to entrenchment and thus different in nature.

\(^\text{13}\) Included in these are 11 occurrences of *isn’t*. 
Although we cannot draw far-reaching conclusions on the basis of such small data sets alone, we accept them as indicators suggesting that, other things being equal, grammaticalizing constructions tend to realize intrusive /r/ to a greater extent than other potential liaison sites do. However, as the tendencies found tie in with previous studies in a usage-based tradition (Krug 1998; Bybee 2011), we are more confident in arguing that the sequences idea of and idea is are more tightly bonded than the intervening phrase boundary between NP and PP or NP and VP would suggest. In other words, we would argue that due to high co-occurrence of their component elements the sequences idea is and idea of form closer units in processing and production than either of or is with the potentially infinite number of open-class items that follow them (cf. Bybee 2010: Ch. 3). Increased rates of intrusive /r/ are thus perhaps a secondary formal change following the emergence of early stress.

5. Summary and conclusions

In the quantitative corpus-linguistic study of the Time Magazine Corpus of American English, we found a rather dramatic real-time increase of idea is and a somewhat smaller but still very sizeable increase of idea was from the 1920s to the early 2000s. The proportion of grammaticalizing to- and that-clauses relative to other types of complementation also increased throughout that period and, in particular, to-complement clauses account for a substantial share of the overall frequency gains. These factors lead to a very marked increase in the grammaticalization index we propose, a measure combining frequency of syntactic strings (hence bondedness) and specialization. From a study of the BNC we conclude that intention-marking constructions other than idea is/was have not grammaticalized to a similar extent: Their frequencies are lower overall, and although some of them are frequently followed by to- or that-clauses, they do not score as highly as constructions containing idea on the grammaticalization scale. We found that — unlike the idea is/was — the constructions the intention/aim/plan/purpose is/was are largely confined to written language.

The picture is complicated by the fact that idea is/was may fulfil different functions and will usually even blend them: Apart from its development towards a grammaticalized intention-marking construction, it may alternatively or simultaneously function as a semantically empty focusing device and the boundaries between these two are not always sharply defined.

Our phonological analysis of idea in context in the ICE-GB suggests that early stress becomes considerably more likely in the constructions the idea is and the idea was. We would argue that this reflects – on a formal level – the functional change described above. A second
development in the pronunciation of *idea* in context is the observation of increased rates of intrusive /t/ in *idea is*. In principle, we find it not very surprising that chunking leads to increased liaison rates in high-frequency constructions. However, the combination of increased rates of stress shift, resulting in a more ideal alternating stress pattern in *the idea is* (especially if *is* is given secondary stress) and the indication of closer bondedness between *idea is* through liaison seems quite exceptional. We conclude therefore that the intention-marking *idea*-construction not only diverges semantically from its source construction, but also formally in taking a more coherent, more prototypical and more fixed phonological form in certain (grammaticalizing) contexts.

Our hypothesis would of course be that patterns roughly equivalent to those found by Krug (2000) for the real- and apparent-time changes in the verbal constructions *going to* and *want to* in the spoken BNC and ARCHER also apply to *the idea is* with regard to stress pattern, reduction of syllables and rates of intrusive /t/. However, quantitative explorations of these aspects are currently impeded by the lack of the necessary corpora of spoken English, i.e. corpora of sufficient size, with accessible soundfiles, and covering a relatively long period of time.

Our final, most general, and therefore perhaps most interesting conclusion is that, within the intention-marking domain, grammaticalization can apparently follow more varied routes than previously supposed. While it is well known that verbal constructions grammaticalize, our research shows that there are reasons to believe that constructions of the type NP + *be* + to-infinitive also grammaticalize and that they may adopt rather different strategies to formally mark the functional changes they undergo. We believe that future investigations within the intention-marking domain that aim to take a broadly onomasiological approach need to include both construction types. Also, if they aim to be more comprehensive than the present paper, they need to look at different text types and degrees of formality, social, regional, internal factors, and ideally more varieties of the English language.
Appendix: Logistic regression predicting early stress in *idea* in the spoken ICE-GB

N = 128 (133 valid cases, 5 missing in this analysis because not coded for gender)
Predictors: IDEA IS/WAS, IDEA OF, PREPAUSAL, FEMALE

Summary of final model (cf. Table 7): Only FEMALE, IDEA IS/WAS and PREPAUSAL remain as predictors, IDEA OF was excluded in the backward logistic regression. IDEA IS/WAS and PREPAUSAL are statistically significant at the .05 level. Prepausal position strongly disfavours early stress and the idea is/was very strongly favours early stress. Male speakers tend to use early stress more frequently than female speakers. Although not statistically significant at the 5% level, this effect is retained in the model as indicating a trend.

<table>
<thead>
<tr>
<th>N = 128</th>
<th>B</th>
<th>SE</th>
<th>p</th>
<th>odds ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEMALE</td>
<td>-.836</td>
<td>.467</td>
<td>.073</td>
<td>.433</td>
</tr>
<tr>
<td>IDEA IS/WAS</td>
<td>2.695</td>
<td>1.134</td>
<td>.018</td>
<td>14.802</td>
</tr>
<tr>
<td>PREPAUSAL</td>
<td>-1.317</td>
<td>.586</td>
<td>.025</td>
<td>.268</td>
</tr>
<tr>
<td>CONSTANT</td>
<td>-.366</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7. Logistic regression predicting level or first-syllable stress. N = 123; Excluded: IDEA OF.

The coefficients (B) in the model indicate by how much the log-odds in favour of the outcome early stress change if the respective predictor takes the value one. The log-odds are useful as an independent variable because they can theoretically take values between $-\infty$ and $+\infty$, but will invariably be convertible into values of $0 < p < 1$. In our model, for example, if a female speaker utters the construction the idea is, the log-odds in favour of early stress are the sum of the coefficients for CONSTANT, IDEA IS and FEMALE, namely $-0.366 + 2.695 - 0.836 = 1.493$. This quantity can be converted into a $p$-value using the logistic function thus:

If FEMALE = 1 and IDEA IS = 1, $p($stress = early$) = \frac{e^{(1.493)}}{1+e^{(1.493)}} = \frac{4.450}{5.450} = .817$

The probability of a female speaker producing *idea is* with early stress is expected to be 81.7%. Predictions of the probabilities of early stress under different conditions are shown in Table 8.

<table>
<thead>
<tr>
<th>Probability of early stress</th>
</tr>
</thead>
<tbody>
<tr>
<td>male speaker, <em>idea is/was</em></td>
</tr>
<tr>
<td>female speaker, <em>idea is/was</em></td>
</tr>
<tr>
<td>male speaker, <em>idea in</em> any other non-prepausal context</td>
</tr>
<tr>
<td>female speaker, <em>idea</em> in any other non-prepausal context</td>
</tr>
<tr>
<td>male speaker, <em>idea</em> in prepausal position</td>
</tr>
<tr>
<td>female speaker, <em>idea</em> in prepausal position</td>
</tr>
</tbody>
</table>

Table 8. Probabilities of the occurrence of early stress under different conditions; based on Table 7

Leaving aside additional factors like FEMALE or PREPAUSAL and concentrating on construction types, a 2×2 Newcombe-Wilson test of the distribution of early and second-syllable stress across the idea is/was vs. all other constructions is significant at an error level $p < 0.05$. 

33


Bybee, Joan. 2005. ‘The impact of usage on representation: Grammar is usage and usage is grammar’. Presidential Address at the 79th Annual Meeting of the Linguistic Society of America (Oakland, USA).


