

Chapter 5

Some (new) thoughts on grammaticalization: Complementizers

Anna Roussou

University of Patras

Grammaticalization creates new grammatical exponents out of existing (lexical) ones. The standard assumption is that this gives rise to categorial reanalysis and lexical splits. The present paper argues that categorial reanalysis may not be so pervasive and that lexical splits may also be epiphenomenal. The set of empirical data involves the development of (Indo-European) complementizers out of pronouns. The main claim is that the innovative element (the complementizer) retains its nominal feature; thus strictly speaking, there is no categorial reanalysis, but a change in function and selectional requirements, allowing for an IP complement as well. As a complementizer, the pronoun is semantically weakened (the nominal core), and phonologically reduced (no prosodic unit). In its pronominal use, it may bind a variable (interrogative/relative) and defines a prosodic unit. What is understood as a lexical split then reduces to a case of different selectional requirements, followed by different logical form (LF) and phonetic form (PF) effects.

1 An overview

According to Meillet (1958 [1912]), the two basic mechanisms for language change are *grammaticalization* and *analogy*. While grammaticalization creates new grammatical material out of “autonomous” words, analogy develops new paradigms by formal resemblance to existing ones. Grammaticalization has received great attention in the literature (for an overview see Narrog & Heine 2011), raising the question whether it is a mechanism of change or an epiphenomenon. The answers provided mainly depend on the theoretical framework adopted and the



view on how “grammar” is to be defined. Thus, in functional approaches, grammaticalization is a mechanism that leads to the formation of “grammar” (or of grammatical structures), while in formal approaches, grammaticalization is either denied altogether (Newmeyer 1998; Lightfoot 1998; 2006; Janda 2001; Joseph 2011) or considered an epiphenomenon (Roberts & Roussou 2003; van Gelderen 2004).

Despite the different views on the topic, it is generally accepted that grammaticalization (or whatever it reduces to) has a visible effect cross-linguistically. There are common tendencies and patterns in how the lexical to functional change may take place (see Heine & Kuteva 2002 for a wide range of examples). For example, complementizers can have their origin in pronouns (interrogatives, demonstratives, relatives), verbs (*say, like, etc.*), nouns (*thing, matter*), or prepositions (allative). Within functional/typological perspectives, grammaticalization is primarily viewed as a “semantic” process where concepts are transferred into constructions (for an overview, see Hopper & Traugott 2003). Once the relevant elements are used as grammatical markers, they show semantic “bleaching” (loss of primary meaning) and phonological reduction. According to Traugott (2010), grammaticalization is not only a matter of reduction, but also of pragmatic expansion in terms of content. At the same time, a typological account shows that some lexical items are more amenable than others to give rise to grammatical markers, although this is not deterministic in any respect. Still, this observation points towards an interesting direction with respect to how the lexicon interacts with (morpho-)syntax.

Within the formal approach to grammaticalization, the basic assumption is that it is an epiphenomenon. More precisely, grammaticalization is argued to derive through the loss of movement steps. In more technical terms, it is a change from internal to external merge. This change gives rise to the creation of new exponents of functional heads, along with structural simplification (see for example Roberts & Roussou 2003; van Gelderen 2004). The notion of simplification is built on the idea that external merge draws directly on the lexicon, while internal merge draws on lexical items already present in the structure.¹ Thus internal

¹The change from internal to external merge is rather simplified here. As Roberts & Roussou (2003) point out, this change may involve additional steps, including the “movement” of features from a lower to a higher position; this is, for example, the case with the development of the subjunctive marker *na* in Greek, where the expression of mood changes from being an inflectional/affixal feature to being a modal marker (*na*) in the left periphery, p. 73–87). In all cases though, the change known as grammaticalization is selective, affecting a subset of lexical items, as also pointed out by an anonymous reviewer; a more thorough discussion is provided in Roberts & Roussou (2003: Ch. 5).

merge gives rise to displacement (movement) and requires at least two copies of the same lexical item in different structural positions. The change from internal to external merge implies a single copy in the higher position and elimination of the lower one. This single copy becomes the new exponent of the higher (functional) head. Since merge is bottom-up, it follows that internal merge will also follow this upward (and leftward) path, and the change from internal to external merge will also affect the upper parts of the structure.

In standard terms, irrespectively of the framework adopted, a basic tenet is that grammaticalization involves a change from lexical to functional, or from functional to functional, as in (1):

- (1) Content word > grammatical word > clitic > inflectional affix

In (1) above, what appears on the left hand side of the arrow “>” indicates a preceding stage. Assuming that lexical categories (content words) are embedded under functional projections (grammatical morphemes), the order in (1) is consistent with the view that “grammaticalization” is accounted for in a bottom-up fashion. More precisely, a lexical item α can start as part of a lexical projection, and by internal merge occur in a higher functional position β . The loss of movement steps has an effect in the categorial status of α , which now becomes the exponent of β . The change from grammatical word to clitic does not affect the functional status but affects the morphosyntactic status of α . The same holds for the final stage (from clitic to an inflectional affix), where α becomes part of the morphological structure, as best summarized in Givón’s (1971: 413) quote “today’s morphology is yesterday’s syntax”.

In the present paper I retain the basic view of Roberts & Roussou (2003) on grammaticalization, namely that it is an epiphenomenon; I also use the term “grammaticalization” in a rather loose way, as the development of grammatical elements out of existing ones. I take complementizers with a pronominal source as the exemplary case, a pattern which is very typical of the IndoEuropean languages. The primary question raised is whether the change from pronoun to complementizer implies categorial reanalysis. The secondary question is whether this reanalysis gives rise to a lexical split that ends up creating homophonous lexical items (i.e., pronoun vs complementizer). The claim put forward here is that the grammaticalized element retains (or at least may retain) its categorial core, thus eliminating homonymy in the lexicon. In §2, I discuss the dual status of some lexical items as pronouns and complementizers, arguing that to a large extent the distinction is functional and not really formal. In §3, I consider the properties of Greek declarative complementizers in connection with their pronominal

uses, showing that we can account for the differences in terms of their logical form (LF) and phonetic form (PF) properties. In §4, I consider the implications of this distribution for grammaticalization, and argue that what looks like a change from pronoun to complementizer indicates a change in selection (expansion) and scope, with visible PF effects also. §5 concludes the discussion.

2 On complementizers and pronouns

Kiparsky (1995) argued that the development of complementizers in Indo-European shows a change from *parataxis* to *hypotaxis*: a previously independent clause becomes dependent on a preceding matrix predicate. This change is linked to a previous one, namely the development of the C position as manifested by V2-phenomena. Another way to view this change is as an anaphoric relation between a pronoun in the first clause which refers to the second (paratactic) clause. Roberts & Roussou (2003) and van Gelderen (2004) argue that in this configuration, the pronoun is reanalyzed as part of the second clause, with the latter becoming part (hypotaxis) of the now main clause since it is embedded under the matrix predicate. This can happen in two steps: first, the pronoun retains its pronominal status and qualifies as a phrase (in a Spec position), and second, it is reanalyzed as a C head, as in (2):

- (2) [IP [VP V pronoun]] [IP] > [IP [VP V [pronoun [IP]]]]

Roberts & Roussou (2003: 118) argue that although this looks like “lowering”, the reanalyzed structure can still be construed in an upward fashion, since the boundary of the second clause shifts to the left (hence upward) to include the pronoun. In their terms, this kind of change is both categorial (pronoun > complementizer) and structural (creating a complement clause headed by the reanalyzed pronoun).² A further aspect of this change is that it has created a new exponent for the C head.

The use of pronouns as complementizers is quite pervasive in Indo-European languages. English *that* is related to the demonstrative *that* (*that book*), Romance

²Kayne (2005: 238) argues that as a complementizer *that* merges above the VP, while as a pronoun it merges inside the VP, accounting for the fact that as a pronoun it may inflect (in Germanic) for case, while as a complementizer it cannot. Within this framework the change from parataxis to hypotaxis would involve merger of *that* in different positions, signalling embedding under the Kayne’s requirement that “For an IP to function as the argument of a higher predicate, it must be nominalized” (Kayne 2005: 236). The complementizer status further implies a silent N.

que/che is related to the interrogative pronoun ‘what’ (*che fai?* ‘what are you doing?’), Greek *oti* is related to a relative pronoun, while *pos* is related to the interrogative ‘how’, to mention just a few examples (see also Rooryck 2013 on French *que* as a single element). In recent approaches to complementation, the relation between pronouns and complementizers is argued to hold synchronically as well. There are basically two ways of analyzing sentential complementation: either to reduce complement clauses to some form of relatives (e.g., Arsenijević 2009; Kayne 2010; Manzini & Savoia 2011), or to reduce relative clauses to an instance of complementation (e.g., Kayne 1994). Either way, the link between the two types of clauses is evident. If indeed there is structural similarity between relatives and complement clauses and the assumption is that complementizers somehow retain their (pro)nominal feature, then what has been considered as categorial reanalysis in the context of grammaticalization may have to be reconsidered.

In their discussion, Roberts & Roussou (2003) point out that one of the differences between D *that* and C *that* has to do with the different complements they embed. In particular, demonstrative *that* takes an NP complement (a set of individuals), while complementizer *that* takes an IP complement (a set of situations/worlds). Manzini & Savoia (2007; 2011) and Roussou (2010) argue that complementizers of this sort are (pro-)nominal. They merge as arguments of the (matrix) selecting predicate and take the CP/IP as their complement; strictly speaking then, they are outside the complement clause. This kind of approach maintains the view that there is embedding, mediated by the “complementizer”, but essentially the relevant element, being a pronominal of some sort (demonstrative, relative/interrogative) occurs as the argument of the predicate. On this basis, it is arguable whether the pronoun changes formally or just functionally. To be more precise, the question is how “real” the D > C reanalysis is. The alternative is to assume that the new element classified as a complementizer retains its nominal property, but expands in terms of selection, allowing not only for an NP but for an IP complement as well.

The approach just outlined regarding complementation is very close to Davidson’s (1997 [1968]: 828–829) view according to which

sentences in indirect discourse, as it happens, wear their logical form on their sleeves (except for one small point). They consist of an expression referring to a speaker, the two-place predicate “said”, and a demonstrative referring to an utterance.

So the sentence in (3a) has the logical structure in (3b):

- (3) a. Galileo said that the earth was round.
b. Galileo said that: the earth is round.
c. Galileo [_v/VP said [that [_{CP/IP} the earth is round]]]

The logical structure in (3b) can translate to the syntactic structure in (3c) where the complementizer is the argument of the selecting predicate. If *that* is construed as a pronoun in (3c), then it retains its nominal feature. This is reminiscent of Kayne's (1982) claim that complementizers have the role of turning the proposition to an argument (also Kayne 2005; see fn. 2). It also recalls Rosenbaum's (1967: 25) analysis, according to which complementizers "are a function of predicate complementation and not the property of any particular sentence or set of sentences". In Rosenbaum's analysis, complementizers are introduced transformationally, and complement clauses are sentences dominated by an NP node.

Leaving many details aside, the next question that arises is to what extent the complementizer splits apart from the pronoun it originates from, leading a life of its own. Is this a lexical split that diachronically yields two homophonous elements, e.g. demonstrative *that*-complementizer *that*, interrogative *che*-complementizer *che*, interrogative *pos*-complementizer *pos*, and so on? Homonymy is an instance of accidental overlap in form with clearly distinct meanings. However, the phenomenon here is very systematic within and across grammars and as such it cannot be treated as accidental. If we exclude homonymy (synchronically), we still need to account for the differences between the original pronoun and the derived complementizer. Note that while *che* as an interrogative requires a Q operator, *che* as a complementizer is declarative and incompatible with a Q-selecting predicate. The same holds for Greek *pos*, which shows a split between an interrogative and a non-interrogative use, as we will see in the following section.

Interestingly, English *how* shows a similar distribution. Consider the following examples from Legate (2010: 122):

- (4) a. They told me how the tooth fairy doesn't really exist.
b. And don't you start in on how I really ought to be in law enforcement or something proper
(www.ealasaid.com/writing/shorts/nightchild.html).
c. They told me about how the tooth fairy doesn't really exist.
d. * They told me about that the tooth fairy doesn't really exist.

A clear difference between *that* and *how* is that *how* can be embedded under a preposition, while this is not the case with *that*, as in (4c).³ Legate argues that *how*-declarative complements are associated with factivity (see also Nye 2013); structurally, they have an abstract DP-layer (c-selection), and semantically they qualify as propositions (s-selection). Unlike *that*, *how* is excluded in relative clauses. The use of *how* as a complementizer does not affect the use of *how* as an interrogative manner adverbial though, as in “*how* did you fix the car?” (= in what manner/way). The question then is whether complementizer *how* is a grammaticalized version of the manner interrogative and a separate entry in the lexicon.

In relation to the above, note that van Gelderen (2015) discusses another use of *how* in matrix yes/no questions, where it remains interrogative (i.e. restricted to questions) but has no adverbial manner interpretation. Consider the following examples (van Gelderen 2015: 164–165):

- (5) a. How would you like to go to the park?
 b. How would you mind clearing a blocking path for Brando Jacobs, eh?
 (<https://twitter.com/jimshearer/status/178244064238514177>)

As van Gelderen argues, this *how* occurs in matrix yes/no questions, and is neither a manner adverbial nor a complementizer. She further shows that throughout the history of English, *how* was not just restricted to a *wh*-manner adverbial, but also conveyed an exclamative or an emphatic reading. In the latter use it emphatically modifies the modal. Let us illustrate this with the example in (5a). In the manner reading, *how* gives rise to the interpretation “in what way would you like to go to the park?”. In the non-manner reading it expresses the degree to which something may hold, giving emphasis on the modal; the reading is something like “Can it be the case that you’d like to go to the park?”, that is an epistemic one. As van Gelderen shows, the emphatic interpretation is already attested in Old English *hu*, so this is not an innovation. What is an innovation though is the yes/no reading of the question introduced by *how*.

On the basis of the empirical data, van Gelderen argues for the following steps in the development of *how* in yes/no questions (emphatic/epistemic) and complement clauses (complementizer):

³One of the reviewers points out that *how* can be embedded under a preposition because it has a degree feature, which *that* lacks. More precisely, *about* refers to properties which can be provided by the adverbial *how*; *that* refers to truth values, so embedding under *about* results in an empty intersection, hence the ungrammaticality.

- (6) a. Adverbial *how*: movement to Spec,CP as a manner adverb modifying vP >
b. Merge in C [i-degree] (interpretable feature) >
c. Spec to Head (not complete for *how*).

The step in (6b) involves a change from internal to external merge with an interpretable feature. The step in (6c) eliminates specifiers in favor of heads. According to her analysis, this state is not complete for *how*, while it is for *whether*. The steps in the reanalysis of *how* affect the features associated with it. More precisely, van Gelderen argues that, as a *wh*-element, *how* has the feature bundle {i-wh, manner/quantity/degree}. The formal *wh*-feature is interpretable and agrees with the uninterpretable *wh*-feature of C in questions, triggering a *wh*-question reading. If the interpretable feature of *how* is that of [i-degree], as opposed to [wh], then it is emphatic (i.e., to such a great degree). If this latter feature becomes uninterpretable, then *how* is merged directly in C and *how* qualifies as a complementizer. In yes/no questions, as in (5), *how* has an interpretable polar feature. In this latter context, according to van Gelderen, the Spec-to-Head reanalysis is not complete.⁴

The account provided by van Gelderen (2015) highlights different stages of *how* both diachronically and synchronically, by manipulating the repertoire of features associated with *how* and its structural position. Synchronically, this allows for different functions associated with *how* (from *wh*-interrogative to polar interrogative to declarative factive). One way to account for this is to assume that activating different features gives rise to different interpretations. Instead of treating the different *hows* as distinct elements (homonyms), we can treat all instances of *how* as a single but polysemous element, where polysemy is structurally-conditioned. For example, in the context of a Q operator, the only available reading is that of an interrogative, either as a *wh*-element, or as an epistemic (yes/no questions). If there is no Q operator, then no interrogative reading arises and therefore *how* can only be compatible (*modulo* its degree feature) with a declarative context under selection by a certain class of predicates (hence its factive reading). The distinction between a specifier and a head (complementizer)

⁴One of the reviewers suggests that the degree feature of *how* combined with the pragmatics of verbs like *mind*, *like*, etc., as in “I would SO like to be there”, is maintained in yes/no questions as well. Thus (5b) could get the answer “Well, not very much”. Van Gelderen models this change in terms of interpretability (an interpretable feature becomes uninterpretable in its new position); I agree with the reviewer, however, that the degree feature remains interpretable. What is crucial is that yes/no questions introduced by *how* implicate an epistemic (or evidential) reading, which shifts the degree feature from the predicate to the proposition. Note also that in all the examples with matrix *how*, the modal *would* is present.

is a function of the syntactic position of *how* and the dependency it forms either with a constituent or a proposition. Note that the interpretation of *how* seems to be affected by the presence or the absence of an operator in the clause-structure, in a way that is reminiscent of polarity-item licensing. The pronoun then acquires its quantificational force, as a *wh*-phrase, through the presence of a Q operator. Once Q is absent, there is no *wh*-reading either, allowing for a declarative use as a complementizer. We come back to this issue in the following section.

In what follows, I will expand the empirical base by considering similar data in Greek which has a range of declarative complementizers with a pronominal (interrogative, relative) counterpart. As will be shown below, this “duality” can give rise to ambiguity in some contexts (recall *how* in 5a).

3 The double behavior of pronouns

In the discussion that precedes we saw that a clear-cut distinction between pronouns and complementizers is not so obvious. To put it differently, as the discussion in van Gelderen (2015) shows, the non-manner uses of *how* are also attested in earlier stages of English, so this is not an innovation. One way to understand this is as follows: the non-manner readings are compatible with a core interpretation of *how* that allows it to modify manner in qualitative terms as well (degree > emphasis). The interrogative use depends on the activation of the *wh*-feature in the scope of a Q operator; in fact, it only arises in the scope of Q. The issue of categorial reanalysis now emerges in clearer terms: does it really exist, and if so to what extent? It is interesting to mention that in a framework where lexical items are considered as feature bundles in the lexicon (Chomsky 1995), categorial classification can be viewed in a different perspective, as will be shown below.

Bearing the above in mind, let us now consider Greek which has a range of declarative complementizers. Along with *oti* (‘that’), we also find *pos* (‘how’). This looks very much like English *that* and *how*. There is a crucial difference though: *oti* and *pos* seem to be in free variation and are selected by the same predicates (note that some dialects may show a strong preference for *pos* instead of *oti*). Greek possesses a third declarative complementizer, namely *pu* (lit. ‘where’) which is selected by factive predicates (Christidis 1982; Roussou 1994; Varlokosta 1994). The complementizer *pu* also introduces restrictive and non-restrictive relative clauses, where *oti* and *pos* are excluded:⁵

⁵*How* can also be used in relative clauses, in examples like “The way how she walks”. The equivalent construction in Greek would use the relative pronoun *opos*, which has the prefix *o-* and the *wh*-pronoun *pos* (lit. ‘the how’).

(7) Greek

- a. nomiz-o oti/pos kerdhis-e to vravio
think-1SG that/that won-3SG the prize
'I think that she won the prize.'
- b. thimame pu kerdhis-e to vravio
remember-1SG that won-3SG the prize
'I remember that she won the prize.'
- c. o fititis pu sinandis-es ine filis mu
the student that met-2SG is friend mine
'The student that you met is my friend.'

Greek then has a two-way distinction of three complementizers: *oti/pos* and *pu*. The two-way distinction involves factivity and relativization. Specifically, *pu*-complements are associated with a factive interpretation, while *oti/pos*-complements are selected mainly by non-factives and only some factives (Christidis 1982; Roussou 1994). So there is a one-way implication between sentential complementation and factivity, since not all factive complements are introduced by *pu*. With respect to relativization, *pu* is the only complementizer available; as will be shown immediately below, free relatives behave differently (and exclude *pu*).

Considering *pos* and *pu* in more detail, we observe that they also correspond to *wh*-pronouns, as in the following examples:

(8) Greek

- a. pos tha fij-is?
how FUT leave-2SG
'How will you leave?'
- b. pu tha pa-s?
where FUT go-2SG
'Where will you go?'
- c. pu to=edhos-es to vivlio?
where it=gave-2SG the book
'Where/to whom did you give the book?'

In (8) both *pos* and *pu* occur in matrix questions. They may also introduce embedded *wh*-interrogatives. Both sentences have a *wh*-question (rising) intonation. As (8c) shows, *pu* as an interrogative apart from the locative reading, it

may also realize an indirect (oblique) *wh*-argument. Finally, it is crucial to mention that although *oti* does not have a *wh*-counterpart, it has a relative pronoun one, which in orthographic terms is spelled as *o,ti* (lit. ‘the what’). As a relative pronoun, it is found in free relatives with an inanimate referent, and is excluded from restrictive and non-restrictive relative clauses (the relevant examples are given below).

The picture we have so far with respect to the distribution of English and Greek complementizers and their pronominal counterparts can be summarized as in Table 5.1.

Table 5.1: Pronoun and complementizers (Greek and English)

	Demonstrative	Complementizer	Relative	Interrogative
<i>oti</i>	no	yes	yes	no
<i>pos</i>	no	yes	no	yes
<i>pu</i>	no	yes	yes	yes
<i>that</i>	yes	yes	yes	no
<i>how</i>	no	yes	yes	yes

A quick look at Table 5.1 shows that all five elements qualify as declarative complementizers, despite their different feature specifications. It further shows that all of them have a pronominal use as well, despite differences again. Based on this pattern, I will assume that their core defining property is that of N, i.e., they are essentially nominal elements (see also Franco 2012), which can be construed with different features (D, *wh*, etc.) or different functional layers (Baunaz 2015). In this respect, their core (minimal) categorial content is N – very much like indefinites; this property can account for the fact that they may also distribute like indefinites, subject to operator licensing (polarity-like). I leave this issue open for the time being.

Let us now consider the following sentence (I leave *oti* unglossed on purpose in the following example):

- (9) Greek
 pistev-i oti dhjavas-e
 believe-3SG OTI read-3SG
 i. ‘He believes that he has studied.’
 ii. ‘He believes whatever he has read.’

The two sentences above exemplify two different readings. In (9i) *oti* is a complementizer that introduces the complement clause of *pistevi*. In (9ii) it is a relative pronoun construed as the argument (object) of *dhjavase*. The whole clause introduced by *oti* (or *o,ti*) is the internal argument (object) of *pistevi*. What is responsible for this ambiguity? First, a verb like *pistevi* ‘believe’ can take either a noun or a clause as its complement; second, the embedded verb *dhjavazi* ‘read’ can take an implicit argument. So in (9i), the matrix verb selects a sentential complement, and the embedded verb has an implicit argument. In (9ii), on the other hand, the matrix predicate selects a free relative (akin to a noun phrase) while the argument of the embedded verb is not implicit but present in the form of the displaced pronoun. The string of words in *pistevi oti dhjavase* is ambiguous between a complement clause (where *oti* functions as a complementizer) and a relative clause (where *oti* functions as a free relative pronoun). Thus the surface string of words in this case corresponds to two different syntactic configurations.

Similar examples hold with the other two elements, namely *pos* and *pu*, as below (again left un glossed):

(10) Greek

paratiris-a pos jiriz-i o troxos
observed-1SG POS spin-3SG the wheel

- i. ‘I observed that the wheel was spinning.’
- ii. ‘I observed how the wheel was spinning.’

(11) Greek

emath-a pu perpatis-e
learnt-1SG PU walked-3SG

- i. ‘I learnt/found out that he had walked.’
- ii. ‘I learnt/found out where he had walked.’

In the absence of any PF-indication (prosody), each of these lexical items can be construed as a declarative complementizer (i), or a pronoun (ii).

In all examples (9–11) so far, the complementizer construal is possible to the extent that a declarative complement is selected by the matrix predicate. In (10–11) for example, if instead of *paratirisa* and *ematha* accordingly we have an interrogative predicate, such as *rotisa* (‘asked’), then only a *wh*-interrogative reading is available, as expected. So ambiguity arises in certain contexts only. The second property we need to point out is that the interrogative reading in these examples (and accordingly, the free relative in (9)) depends on the availability of a variable in the complement clause, that is an open position that modifies the predicate for manner, place, etc. (or as an implicit argument in (9) for the free relative reading).

I will assume, in line with work in the recent literature (Manzini & Savoia 2007; 2011; Roussou 2010; Franco 2012), that as a complementizer each of these elements merges as the argument of the verb and takes the CP/IP as its complement. On the other hand, as a pronoun it is internal to the embedded clause, at least to the extent that it has a copy inside the clause (at the v/VP level), as illustrated below:

- (12) a. V *oti/pos/pu* [CP/IP ...]
 b. V [*o,ti/pos/pu* [CP/IP ... *o,ti/pos/pu*]]

The different configurations map onto different PFs; thus the ambiguity is resolved prosodically. As complementizers, *oti*, *pos*, and *pu*, are unstressed, i.e., they do not form a prosodic unit. As pronouns, however, they are stressed, in a manner typical of *wh*-questions; i.e. the pronoun defines an L*+H prosodic unit. This holds for all three cases, including the *o,ti* relative function. The pattern with *pu* as a complementizer has one more interesting angle: as expected, *pu* is unstressed, but the preceding predicate is stressed (an L*+H prosodic unit). In other words, selection of a *pu*-complement in this context is associated not only with the semantics of the selecting predicate but also with focus. As expected, focus on the predicate turns the *pu*-complement to the presupposed part, hence its association with factivity (on the interaction of focus with factivity, see Kallulli 2006).

What we observe so far is that the lexical items under consideration have two phonological variants: a stressed one (pronominal) and an unstressed one (complementizer). This kind of alternation is quite common in the pronominal system. For example, in Classical Greek the indefinite pronoun *tis* has an accented variant (*tís*) as an interrogative (also Latin *quis*); in (Modern) Greek negative polarity items like *kanenas* ('anyone') and *tipota* ('anything') acquire a status of universal quantifiers (negative quantifiers) when focused. So the different categorizations of *oti*, *pos* and *pu* as pronouns vs complementizers in relation to their phonological properties comes as no surprise in this respect. But does this property suffice to classify them as distinct lexical items synchronically? The answer seems to be negative, given that their differences can be accounted for independently.

Assuming that the use of *pos* and *pu* as complementizers is an innovation in the path of their diachronic development, the question is whether grammaticalization is at stake or not. So far, I have argued that, strictly speaking, there is no categorial reanalysis as such, in the sense that in either function, these elements retain their nominal core. The activation of the *wh*-feature depends on the presence of a Q operator and involves focusing of the item in question. If this is correct, the interrogative reading is syntactically defined and is read off at the

two interfaces (it introduces a variable at LF, it defines a prosodic unit at PF). On the other hand, as complementizers, they are selected by designated predicates and in turn they select a proposition. The complementizer is externally merged, subject to selection by the matrix predicate. This structure is accordingly read off at LF, given that the complementizer turns the clause to an argument, and at PF, since it has not prosodic properties. This latter characteristic is in accordance with the notion of phonological reduction attested in grammaticalization. What about semantic weakening (or bleaching)? As a complementizer, the pronoun retains its nominal core, and has no additional features (like *wh-*). At the same time, under its complementizer use, the lexical items under consideration expand, on the assumption that they manifest a wider choice in terms of selection; for example, i.e., selection of an NP or a clause (CP/IP). Note that the complementizer status assigned to *oti* is not an innovation, since it is used as a complementizer throughout the history of Greek.

The above properties can be summarized as follows:

- (13) a. Pronoun: prosodic unit, internal merge/scope
- b. Complementizer: no prosodic unit, external merge.

The development of the complementizer use for *pos* and *pu* under the current approach is consistent with the “change” from internal to external merge. As already pointed out, as displaced pronouns due to internal merge, they bind a lower copy and take scope. As complementizers they merge externally and therefore do not bind a copy. Does this approach account for the idea of “upward reanalysis” of Roberts & Roussou (2003)? Recall that in relation to the schema in (2), Roberts & Roussou assume that this involves a leftward shift of the clause boundary; this means that while the pronoun as a complementizer literally lowers, since it becomes part of the embedded clause, the boundaries of the embedded clause move upwards to include the reanalyzed pronoun. This account though has some shortcomings, given that it takes “upward” in linear and not structural terms. In terms of the claim made in the present paper, the upward reanalysis is accounted for structurally: the pronoun as a complementizer merges as the argument of the predicate (as was before), and the paratactic clause becomes embedded under the pronoun, triggering the change from parataxis to hypotaxis. The pronoun nominalizes the second clause, which now qualifies as an argument. The relation between the pronoun and the clause changes from being anaphoric to being an instance of complementation.

In short, the presentation of the data above points towards a unified account of pronouns and complementizers. The basic line of reasoning is the following: if

two grammatical lexical items look the same, they are (most probably) the same. Further evidence is provided by the fact that this similarity is diachronically and synchronically supported. Diachronically, because we can trace the steps in the development of complementizers, and synchronically, because it is very systematic across grammars, but also within a given grammar, to be simply treated as accidental (as is the case with homonymy).

4 Grammaticalization and syntactic categories

The above discussion on (Greek) complementizers has concentrated on the connection between the “new” functional item and its lexical source. So the question has been whether complementizers retain their core nominal feature or not. So far, I have talked about complementizers and pronouns, assuming that the latter occur in the left periphery of the clause, while the former (potentially) as arguments of the selecting predicate. I have made no reference to the C head as such though. In fact, the approaches that assign a nominal feature to complementizers distinguish it from the C positions as such. If C is a position retained for verbal elements that is part of the (extended) projection of the verb (Manzini & Savoia 2011), then it is not and cannot be realized by nominal-type elements (such as pronouns or complementizers, unless the latter are verbal-like). This line of reasoning allows us to maintain that the pronoun to complementizer reanalysis is not an instance of categorial reanalysis. In other words, it is more of a functional change (affecting the use of the pronoun) and less so of a formal one.

This issue of categorial reanalysis arises in all contexts of grammaticalization. For example, when verbs become modals, do they retain their verbal feature? Does *for*, as a complementizer in English, retain its prepositional feature? Is the infinitival marker *to* in English the same as the preposition *to*? This question can be obviously asked for every single case of grammaticalization, and it is related to the nature of syntactic categories, their repertoire and feature specification. The answers to the question just raised can vary. Consider the case of *for* as in the following example (for a historical account, see van Gelderen 2010):

- (14) a. A present for Mary/her
 b. I prefer for Mary/her to be late

In descriptive terms, *for* in (14a) is a preposition which takes a DP complement (*Mary*) or an accusative pronoun (*her*). In (14b), on the other hand, *for* introduces the infinitival complement, and is usually analyzed as a C element. However, it

can still assign accusative case to the embedded subject (*Mary/her*), at least under standard assumptions in the generative grammar. If so, then it maintains its prepositional property of being a case assigner. This has been further supported by the fact that in Standard English at least, *for* forces the presence of an overt subject and excludes a null one (that is a PRO subject), as in **I prefer for to be late vs. I prefer to be late*. Based on similarities of this sort, (Kayne 1984; 2000) pointed out the affinity between prepositions (P) and complementizers (C), but also determiners (D). This then turns out to be a recurrent theme in the literature.

In the light of the present discussion, the link between apparently different categories is not surprising. To be more precise, if *for* is a preposition in (14a) there is no particular reason why it cannot be a preposition in (14b). The difference between the two instances has to do with the different complements *for* takes in either case. That prepositions can introduce subordinate clauses is rather well established, and can be further illustrated with the following examples:

- (15) a. We went for a walk after the dinner
b. After we had dinner, we went for a walk

Once again, the same element can take different types of complements: a DP or a clause (finite or non-finite). Unlike *for*, *after* can only select for a finite clause, does not interfere with the realization of the embedded subject, and can only introduce adverbial (non-complement) clauses.

Having provided a discussion of the relation between pronominals and complementizers (potentially extending this to prepositions as well), let us discuss a bit more the question that we first raised, namely that of lexical splitting in the context of grammaticalization. Related to this is the categorial identification of the new lexical item. As discussed in the literature, there are many cases where the limits between two categories are not very obvious, or “fuzzy” (see Traugott & Trousdale 2010). In typological approaches to grammaticalization, where grammatical categories are under formation, the notion involved is that of *gradience*. However, in formal approaches where grammatical categories are defined as bundles of features with a role in the syntactic computation, the notion of gradience is problematic. Roberts (2010) partly overcomes this problem by assuming an elaborate functional hierarchy, along the lines of Cinque (2006), allowing for the possibility for the same lexical item to merge on different heads along this hierarchy. This has the advantage of maintaining a core property, thus avoiding the issue of homonymy, while at the same time it derives the different meanings by merger of the same item in different positions. So in this respect, what looks like a lexical split has a syntactic explanation: the same lexical item can realize different positions along the functional hierarchy. One disadvantage of this

approach is that it requires every possible meaning to be syntactically encoded, introducing an immense increase of functional categories, even for those cases where certain readings can be derived pragmatically.

Another issue that opens up under the current approach concerns the lexical vs. functional divide. If so-called grammaticalized elements can retain their verbal or nominal (in a broad sense) features, then the question is what sort of implications this has for the lexicon, the syntax and the view of parametric variation, among other things. One possible answer is that this basic distinction between two classes of lexical items is not primary but secondary. This is consistent with the view that the lexicon consists of lexical items with no a priori characterization (see Marantz 2001). If something is a predicate, i.e. assigns a property or expresses a relation, it has all the typical characteristics to qualify as a core lexical category. Languages allow these elements to generate quite freely in the lexicon. At the same time, whether an element functions as a predicate or not is to a large extent determined configurationally under current minimalist assumptions. Similarly, whether the same element is “less lexical” or “more functional” is also determined configurationally. This is indeed captured in Cinque’s (2006) approach, and is to some extent implicit in Roberts’s (2010) account. So “more functional” in current terms is understood as being associated with a high position (and scope) in the clause structure. But if this is correct, it does not really tell us much about this distinction as an aspect of the lexicon. As Manzini & Savoia (2011: 5) put it “There is no separate functional lexicon – and no separate way of accounting for its variation”.

Consider again the case of verbs, which are typical examples of predicates in natural languages. The verb expresses its argument structure in connection with certain positions, realized by nominals (giving rise to expressions of transitivity, case, etc.). The typical I and C positions associated with the verb are essentially scope positions (relating to the event, the proposition, or various types of quantification over possible worlds, etc.). Nominals also have a predicative base, carry inflectional properties and become arguments in relation to a predicate. What actually lies in the heart of this discussion is the categorization of concepts. Different choices give rise to different lexica cross-linguistically. Interestingly, it is in this respect that grammaticalization in functionalist frameworks makes sense, since the idea is that concepts acquire a grammatical form and consequently grammatical categories are defined functionally. In formal approaches, syntactic (grammatical) categories are meant to be well-defined, and languages differ as to which concepts map onto which categories and how. This raises the question of how well-defined categories are. Looking at complementizers and how they

develop out of pronouns can shed some light into this question. So is complementizer a formal category after all, or is it a functional classification of a nominal (or a verbal in other languages) element? grammaticalization phenomena then allow us to have a better understanding of how syntactic category could be defined, how they are realized cross-linguistically and how they are manipulated by narrow syntax.

In short, grammaticalization phenomena can tell us something about the diachronic development of grammatical elements, especially with respect to morphosyntax. At the same time, they force us to pay closer attention to what actually a syntactic category is. The answers are not easy either way, but the empirical data is there to be further explored. Taking a view towards grammaticalization along the lines suggested here, where its core property of categorial reanalysis is put into question, invites us to reconsider syntactic change and focus more on how certain elements change the way they do, even if they retain their categorial status (thus no categorial reanalysis).

5 Concluding remarks

In the present paper I have mainly focused on the notion of categorial reanalysis, and in this respect I have outlined an account which, at least to some extent, casts some doubts on this standard view. The empirical set of data was restricted to the development of complementizers out of pronouns. The basic argument has been that formally, the innovative element, namely the complementizer, retains its nominal categorial feature. In its new function as a complementizer, the pronoun externally merges with the selecting predicate. The change attested involves properties that affect the interfaces, such as phonological reduction and selectional/scope requirements.

Abbreviations

1	first person	LF	logical form
2	second person	PF	phonetic form
3	third person		
FUT	future	SG	singular

Acknowledgements

This paper is dedicated to Ian Roberts on the occasion of his 60th birthday. It reflects on our joint work on grammaticalization, adding a new angle on categorial reanalysis. Working with Ian has been inspirational and a good source of agreements and (productive) disagreements! The present version of the paper has benefited from the constructive comments of two anonymous reviewers. I thank them both.

References

- Arsenijević, Boban. 2009. Clausal complementation as relativization. *Lingua* 119(1). 39–50. DOI: 10.1016/j.lingua.2008.08.003.
- Baunaz, Lena. 2015. On the various sizes of complementizers. *Probus* 27(2). 193–236. DOI: 10.1515/probus-2014-0001.
- Chomsky, Noam. 1995. *The Minimalist program*. Cambridge, MA: MIT Press.
- Christidis, Anastasios Ph. 1982. *Oti/pos-pu*: Complementizer selection in modern Greek. *Greek. Studies in Greek Linguistics* 2. 113–177.
- Cinque, Guglielmo. 2006. *Restructuring and functional heads*. Vol. 4 (The cartography of syntactic structures). Oxford: Oxford University Press.
- Davidson, Donald D. 1997 [1968]. On saying *that*. In Peter Ludlow (ed.), *Readings in the philosophy of language*, 817–832. Reprinted from *Synthese* 19, 130–146. Cambridge, MA: MIT Press.
- Franco, Ludovico. 2012. Against the identity of complementizers and (demonstrative) pronouns. *Poznań Studies in Contemporary Linguistics* 48. 565–596.
- Givón, Talmy. 1971. Historical syntax and synchronic morphology: An archaeologist's field trip. *Chicago Linguistic Society* 7. 394–415.
- Heine, Bernd & Tania Kuteva. 2002. *World lexicon of grammaticalization*. Cambridge: Cambridge University Press.
- Hopper, Paul J. & Elizabeth Closs Traugott. 2003. *Grammaticalization*. Second edition. Cambridge: Cambridge University Press.
- Janda, Richard D. 2001. Beyond “pathways” and “unidirectionality”: On the discontinuity of language transmission and the counterability of grammaticalization. *Language Sciences* 23(2–3). 265–340. DOI: 10.1016/S0388-0001(00)00023-1.
- Joseph, Brian D. 2011. Grammaticalization: A general critique. In Heiko Narrog & Bernd Heine (eds.), *The Oxford handbook of grammaticalization*, 193–205. Oxford: Oxford University Press.

- Kallulli, Dalina. 2006. Triggering factivity: Prosodic evidence for syntactic structure. In Donald Baumer, David Montero & Michael Scanlon (eds.), *Proceedings of the 25th West Coast Conference on Formal Linguistics*, 211–219. Somerville, MA: Cascadilla Proceedings Project.
- Kayne, Richard S. 1982. Predicates and arguments, verbs and nouns. *GLOW Newsletter* (8). 24.
- Kayne, Richard S. 1984. *Connectedness and binary branching*. Dordrecht: Foris.
- Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, MA: MIT Press.
- Kayne, Richard S. 2000. A note on prepositions, complementizers and word order universals. In *Parameters and Universals*, 314–326. Oxford: Oxford University Press.
- Kayne, Richard S. 2005. Antisymmetry and Japanese. In *Movement and silence*, 215–240. Oxford: Oxford University Press.
- Kayne, Richard S. 2010. Why isn't *this* a complementizer? In *Comparisons and contrasts*, 190–227. Oxford: Oxford University Press.
- Kiparsky, Paul. 1995. Indo-European origins of Germanic syntax. In Allan Battye & Ian Roberts (eds.), *Clause structure and language change*, 140–169. Oxford: Oxford University Press.
- Legate, Julie-Anne. 2010. On how *how* is used instead of *that*. *Natural Language and Linguistic Theory* 28(1). 121–134. DOI: 10.1007/s11049-010-9088-y.
- Lightfoot, David W. 1998. *The development of language: Acquisition, change, and evolution*. Oxford: Blackwell.
- Lightfoot, David W. 2006. *How new languages emerge*. Cambridge: Cambridge University Press. DOI: 10.1017/CBO9780511616204.
- Manzini, M. Rita & Leonardo M. Savoia. 2007. *A unification of morphology and syntax: Studies in Romance and Albanian dialects*. London: Routledge.
- Manzini, M. Rita & Leonardo M. Savoia. 2011. *Grammatical categories: Variation in Romance languages*. Cambridge: Cambridge University Press.
- Marantz, Alec. 2001. Word and things. Ms., MIT.
- Meillet, Antoine. 1958 [1912]. L'évolution des formes grammaticales. In *Linguistique historique et linguistique générale*, 130–58. Reprinted from *Scientia* 12, 384–400. Paris: Champion.
- Narrog, Heiko & Bernd Heine (eds.). 2011. Oxford: Oxford University Press.
- Newmeyer, Frederick J. 1998. *Language form and language function*. Cambridge, MA: MIT Press.
- Nye, Rachel. 2013. Complementizer-like *how* clauses and the distribution of finite clausal complements in English. Paper presented at ConSOLE XXI, 8–10 January 2013, University of Potsdam.

5 *Some (new) thoughts on grammaticalization: Complementizers*

- Roberts, Ian. 2010. Grammaticalization, the clausal hierarchy and semantic bleaching. In Elizabeth Closs Traugott & Graeme Trousdale (eds.), *Gradience, gradualness and grammaticalization*, 45–73. Amsterdam: John Benjamins. DOI: 10.1075/tsl.90.05rob.
- Roberts, Ian & Anna Roussou. 2003. *Syntactic change: A minimalist approach to grammaticalization*. Cambridge: Cambridge University Press.
- Rooryck, Johan. 2013. *Configurations of sentential complementation: Perspectives from romance languages*. London: Routledge.
- Rosenbaum, Peter S. 1967. *The grammar of English predicate complement constructions*. Cambridge, MA: MIT Press.
- Roussou, Anna. 1994. *The syntax of complementisers*. University College London. (Doctoral dissertation).
- Roussou, Anna. 2010. Selecting complementizers. *Lingua* 120(3). 582–603. DOI: 10.1016/j.lingua.2008.08.006.
- Traugott, Elizabeth Closs. 2010. Grammaticalization. In Silvia Luraghi & Vit Bubenik (eds.), *The Continuum companion to historical linguistics*, 271–285. London: Continuum.
- Traugott, Elizabeth Closs & Graeme Trousdale (eds.). 2010. Amsterdam: John Benjamins. DOI: 10.1075/tsl.90.
- van Gelderen, Elly. 2004. *Grammaticalization as economy*. Amsterdam: John Benjamins. DOI: 10.1075/la.71.
- van Gelderen, Elly. 2010. Features in reanalysis and grammaticalization. In Elizabeth Closs Traugott & Graeme Trousdale (eds.), *Gradience, gradualness and grammaticalization*, 129–147. Amsterdam: John Benjamins. DOI: 10.1075/tsl.90.08gel.
- van Gelderen, Elly. 2015. The particle *how*. In *Discourse-oriented syntax*, 159–174. Amsterdam: John Benjamins. DOI: 10.1075/la.226.
- Varlokosta, Spyridoula. 1994. *Issues on Modern Greek sentential complementation*. University of Maryland. (Doctoral dissertation).