Re-evaluating the importance of discourse-embedding for specificational and predicative clauses

Wout Van Praet

UNamur | KU Leuven | F.R.S-FNRS

This paper studies the discourse-embedding of specificational clauses, in contrast with predicative ones. Specificational clauses – which express a variable – value relation – are assumed to have a 'fixed' information structure. This follows from the widespread definition of information structure in terms of a presupposition – focus contrast, which is often conflated with the variable – value contrast, on the one hand, and with a given – new contrast, on the other. Against these conflations, this study demonstrates that the specification is a separate layer of meaning, which not only shows variation in terms of focus-marking (Van Praet and O'Grady 2018), but also in terms of its embedding in specific contexts of use. These findings urge us to revisit not only the basis for distinguishing specificational clauses from predicative ones, but also to separate out the different layers of coded and pragmatic meaning that have been conflated under the header of 'information structure'.

Keywords: discourse-familiarity, specificational and predicative clauses, information structure, presupposition and focus, reversibility, mirativity

1. Introduction

In studies of information structure, the assumption exists that specificational clauses, like (1), are syntactic structures with a 'fixed' information structure (e.g. Mikkelsen 2005; Lehmann 2008). Their meaning can be glossed as setting up a variable, e.g. *the world capital* in (1), for which a value is specified, e.g. *London* (e.g. Akmajian 1970; Declerck 1988).[1] This variable – value relation has been

---

1. The terms 'variable' and 'value' are used, in this paper, in their standard meaning taken from mathematical logic (e.g. Akmajian 1970; Higgins 1979). Here, the term 'variable' refers to the

viewed in terms of a presupposition – focus contrast, with the variable being inherently 'presupposed' and the value being the focus (e.g. Declerck 1988; Mikkelsen 2005; Patten 2012).

(1)   Through most of the 19th Century the world capital was London.

(Wordbanks*Online*, henceforth WB)

Moreover, the 'fixed' information structure of specificational clauses is believed to set them apart from predicative clauses, like (2), which have a 'free' information structure (e.g. Declerck 1988; Moro 1997; Mikkelsen 2005, 51, 162–163; Patten 2012). The function of predicative clauses is to describe an entity (henceforth, the 'describee', e.g. *London*) by attributing representational (e.g. *town; city*) and, potentially, individuating qualities (e.g. *great; beautiful; with lots of different things to do*) to it. Unlike in specificational clauses, neither part of the predicative clause is believed to be intrinsically presupposed, nor is either of them inherently focal.

(2)   London is a great town, a beautiful city with lots of different things to do.(WB)

To the different meanings of specificational and predicative clauses correspond different grammatical structures. This is evidenced by the different formal behaviour of the two clause types. Predicative clauses have one-place predicates that code intransitive relations (Langacker 1991, 66; Davidse and Van Praet 2019, 5): the describee is the only participant in the predicative relation, while the role of the description is to specify the content of the predicative relation (Davidse and Van Praet 2019, 5). Structurally, the predicate nominative (i.e. the NP serving as description) is therefore interpreted as part of the predicate and, hence, cannot become subject (Langacker 1991, 68), e.g. (2′).

(2′)   *A great town is London / it.

Specificational clauses, by contrast, code transitive relations (Davidse and Van Praet 2019, 15): both the variable and the value are participants in the specificational relation and, hence, either one can become subject (Halliday 1985, 123). This is illustrated by the felicity of (1′) as an alternate for (1) (in answer to the same implied question 'what was the world capital'). (Throughout this paper, specificational clauses with variable subject, like [1], will be referred to as 'non-

---

more abstract entity for which a more concrete 'value' is sought. The use of these terms differs from how Halliday (1967) uses them, which is exactly the other way round. In Halliday's semiotic model, the more abstract term is the 'value', as in Saussure's 'valeur', and the more concrete entities that can realise it are 'variables' or, later, 'tokens'.

reversed' specificational clauses; the ones with value subject, like [1'], will be called 'reversed' clauses.)

(1')   Through most of the 19th Century London was the world capital.

Since reversed specificational clauses appear superficially similar to predicative ones – i.e. both clauses construe the more specific NP (with the semantic roles of 'value' or 'describee') as subject –, information structure has been advanced as an important distinguishing criterion between the two clause types: specificational clauses require the variable NP to express 'presupposed' old information, while the value NP is 'focal' and, hence, 'new'.

Behind the assumption that specificational clauses have a 'fixed' information structure is the idea that the variable – value relation is the grammaticalisation of an information-structural presupposition – focus relation. Such a view, however, runs the risk of collapsing different layers of coded and pragmatic meaning. This is problematic, not only because it glosses over the significance of the different layers that are collapsed, but also because it sets a false basis for distinguishing between specificational and predicative clauses.

A first issue is that the semantic role of 'value' – e.g. *London* in (1) – is often conflated with the information-structural notion of 'focus' (e.g. Declerck 1988, 12; Patten 2012, 35). This goes back to Chomsky (1969) and Akmajian (1970), who view focus as a 'semantic' component representing 'new' information. Its non-focal counterpart is the presupposition, which is assumed "to be known to the hearer" (Akmajian 1970, 192). Akmajian (1970, 192) adds that the "semantic prominence [of the focus] is correlated with prosodic prominence" (see also Chomsky 1969, 26).[2] Hence, in (3), *Mitchell* is believed to be prosodically marked as 'focus', because it specifies the variable x in the 'presupposed' open proposition [x *urged Nixon to appoint Carswell*] (Lambrecht 1994, 212).

(3)   **Mitchell** urged Nixon to appoint Carswell.   (Lambrecht 1994, 212)[3]

Against the conflation of focus and value, and of presupposition and variable, Halliday (1967, 226) demonstrates that specificational semantics and focus-marking are two separate layers of meaning: not only can specificational values be non-

---

2. Throughout this paper, I will use square brackets to indicate that additions or reformulations in quotes are my own, and not the author's.

3. In cited examples, boldface is used to indicate focus-marking, e.g. (3) and (4). Underlining will be used to draw attention to a specific part of an example, for instance to highlight a copular clause within a larger stretch of discourse, e.g. in (6) below.

focal, e.g. *that* in (4), but variables can also be marked as focus,[4] in which case they are not information-structurally presupposed, e.g. *what he m\eant* in (4).

(4)   ^that's what he **m\eant**#   (London-Lund Corpus of Spoken English, LLC)

These claims were corroborated by empirical studies of focus-marking (Van Praet and O'Grady 2018; Van Praet 2019), which found a correlation but no one-to-one correspondence between variable – value semantics and focus assignment.

Secondly, the notion of presupposition itself also collapses different layers of meaning. While the variable may not be presupposed in the prosodic sense of the term, the variable is commonly argued to be presupposed in a 'pragmatic' sense. To the variable attaches a presupposition of existence (Declerck 1988, 14–16). This presupposition – i.e. that there is an actual individual that corresponds to the description given by the variable NP – is pragmatically inferred from the semantic role 'variable' (Davidse and Van Praet 2019, 28–31). This presupposition of existence is pragmatic (Stalnaker 1973): it attaches by default to the variable, but it is not intrinsic to it and can, therefore, be cancelled, e.g. (5).

(5)   And the winner is … nobody.
                (https://foreignpolicy.com/2009/09/01/and-the-winner-isnobody/)

Finally, the idea that the variable carries a pragmatic presupposition of existence has led to the assumption that it is also presupposed in the sense of being 'discourse-old' (e.g. Akmajian 1970, 192; Declerck 1988, 14; Lambrecht 1994, 212–213). Declerck (1988, 11–12), for instance, claims that "in any specificational sentence the value part represents 'new' information while the variable part expresses 'old' information", which he links explicitly to the value being focus and the variable being presupposed. For Mikkelsen (2005, 135), the assumption that the variable is discourse-given even holds the promise of explaining "why specificational clauses exist at all". She views specificational clauses as expressing the same semantic relation as predicative clauses, so that the difference between the two clause types resides ultimately in their information structure. Mikkelsen (2005, 135) claims, more specifically, that the variable NP is a non-referential expression, which normally does not function as subject. This default may be overridden, she argues, for information-structural reasons, viz. when the 'non-referential' NP is discourse-familiar and serves as 'topic' of the clause (ib.). Mikkelsen (2005) thereby likens specificational clauses to Birner's (1996) 'inversion' constructions (e.g. *In the garden was an old man*) (see also Patten 2012). Inversion serves a discourse-connective function: it allows the speaker to present

---

4. Focus-marking is signalled prosodically by the assignment of the 'tonic' accent (Halliday 1970), indicated in bold, and by the place of the tone, e.g. fall \, in (4).

relatively familiar information before relatively unfamiliar information (Birner 1996, 60). This leads Mikkelsen (2005, 134)[5] to conclude "that being discourse-old is a precondition for being topic, and hence specificational subjects [i.e. variables] must be discourse-old, at least relative to the [value] complement". To my knowledge, however, no study so far has actually tested these hypotheses on larger sets of attested corpus data.

Therefore, though widespread, the assumption that the variable is discourse-old, in contrast with the value's discourse-newness, remains a moot point in the literature. A particular concern is the possibility of the variable being realised by an indefinite rather than a definite NP, e.g. *one example* in (6).

(6)   Property in the "wrong location" – <u>one example is London Town's The Bridge</u>, which sits next to a gasworks, railway line and very busy road in Battersea.

<div align="right">(WB)</div>

Declerck (1988, 19) notes that the variable NP is 'normally' definite. He takes this to follow from "the fact that the presupposition represents old (even given) information" and it is "well-known that indefinite NPs are used to introduce new information" (Declerck 1988, 19). For that reason, Mikkelsen (2005, 154) argues that indefinite variable NPs must reconcile two conflicting discourse conditions, namely the variable's discourse-connective function and the (assumed) discourse-newness of the indefinite NP. The solution that both Declerck (1988, 19) and Mikkelsen (2005, 157) propose is that indefinite variable NPs, e.g. *an example of this* in (7), introduce overall new information but include modifiers that express old information or information that is somehow linked up with the preceding context.

(7)   (Most wars in history have had economic causes.) An example of this is World War II.   (Declerck 1988, 19)

This, Mikkelsen (2005, 154) puts forward, explains why some indefinite NPs are infelicitous as specificational variable: "these particular indefinites fail to contain any discourse-old material", e.g. (8).

(8)   #A doctor is John.   (Mikkelsen 2005, 159)

However, as Mikkelsen (2005, 159) admits, "there are still other factors that seem to play a role in determining the felicity of indefinite specificational subjects". This is illustrated by (9), in which the second occurrence of the NP *a doctor* expresses discourse-old information, but despite its indefinite NP form, it apparently fails

---

5. Mikkelsen (2005) only considers non-reversed specificational clauses, in which the variable is subject.

to felicitously introduce a discourse-new referent. Mikkelsen (2005, 159) regrets, however, "to leave these very interesting observations as questions for further research".

(9) Bill is a doctor. #A doctor is John (too).   (Mikkelsen 2005, 159)

Therefore, since specificational clauses with indefinite variable NP are still relatively uncharted domain, these clauses will form the starting point of this study. To understand their meaning and use correctly, it is necessary to reconsider specificationals with indefinite variable NP as a positively motivated option in the copular clause system, rather than something semi-aberrant that needs to be explained away. This study will do so by focussing specifically on the discourse-embedding[6] of these clauses and comparing and contrasting them with specificational clauses with definite variable NP, on the one hand, and with predicative clauses with indefinite predicate nominative, on the other. In addition, since specificational clauses allow for either the variable or the value to be subject, both 'non-reversed' and 'reversed' clauses – in which the indefinite variable NP is subject vs complement, respectively – will be examined. This allows us to tackle the additional question – which has so far not been answered, if raised at all – whether discourse-embedding interacts with the choice for the variable or the value as subject.

This paper seeks to examine, through attested corpus data, whether discourse-embedding is indeed an essential layer of meaning for specificational clauses, as has been claimed in previous studies (e.g. Declerck 1988; Mikkelsen 2005). This general concern subdivides into the following more specific research questions. First, do specificational clauses always have 'given' variables and 'new' values, or can more variation in discourse status be found for both semantic roles? Can predicative clauses, for which no specific discourse conditions have been posited, be found to display more variation in the discourse status of the semantic roles of 'describee' and 'description'? In addition, are there differences between the discourse statuses of definite variable NPs and indefinite variable NPs, whose analysis has posed difficulties in past research (e.g. Declerck 1988; Mikkelsen 2005; Heycock 2012; Patten 2012)? Finally, how do non-reversed and reversed specificational clauses compare to each other, and to predicative clauses, in terms of discourse-embedding? Can a motivation for choosing one variant of the specificational clause type over the other be found in the discourse status of the variable and/or the value?

---

6. Throughout this paper, the term 'discourse-embedding' will be used to talk about how clauses are integrated with the co- and context, while the term 'discourse-familiarity' will be used to discuss the discourse-givenness, or -newness, of individual clause constituents.

Addressing these questions will not only shed light on the meaning of specificational and predicative clauses, but it will also allow us to disentangle the different layers of coded and pragmatic meanings that have been conflated in many studies of information structure. This will lead to a better appreciation of the interaction between the semantic variable – value relation, on the one hand, and discourse-embedding, on the other hand.

The structure of this paper will be as follows. In Section 2, I will first introduce the model of discourse-familiarity that will serve as the basis for the categories in the corpus study. In Section 3, I will then describe the method used to carry out the corpus research. The results will be analysed in detail in Section 4. The paper will end with a conclusion in Section 5.

2.   A taxonomy of discourse-familiarity

The study of discourse-familiarity in this paper is based on Prince's (1981) account of discourse-familiarity and Kaltenböck's (2005) reconceptualisation of Prince's model. In Prince's (1981, 232) model of 'assumed familiarity', what counts as 'given' is what the speaker "assumes the hearer assumes", since it is the speaker's assumptions about hearer-familiarity that are reflected linguistically. Kaltenböck (2005) takes a more 'empiricist' stance, with a primary focus on the text as a finished product of communication. The corpus study in this paper will adopt Kaltenböck's text-based approach, which has the methodological advantage of providing a more objective basis for analysing corpus data. The rationale behind this is that researchers (normally) do not have (direct) access to the inner states of speakers or hearers. They can, therefore, only draw conclusions from the text itself. These conclusions may feed into hypotheses about the assumptions speaker and hearer make, but these hypotheses cannot be confirmed or rejected based on the text alone.

The model Prince (1981) proposes distinguishes, in a first instance, between NEW, INFERRABLE and EVOKED discourse referents. When a speaker first introduces an entity in the discourse, it counts as NEW information, which can be either BRAND-NEW or UNUSED. In the first case, the hearer is instructed to create a new entity and add it to the discourse-model (Prince 1981, 235), e.g. *a bus* in (10).

(10)   I got on a bus yesterday and the driver was drunk.   (Prince 1981, 233)

(11)   Noam Chomsky went to Penn.   (ib.)

In the second case, the hearer is "assumed to have a corresponding entity in his/her model and has to place it (or copy it) in the discourse-model" (Prince

1981, 235–236), e.g. *Noam Chomsky* in (11). Prince's (1981) distinction between brand-new and unused discourse entities thus pertains to whether a discourse-new entity is assumed to be hearer-new (i.e. brand-new) or hearer-familiar (i.e. unused).

Brand-new entities can be further divided into ANCHORED and UNANCHORED entities, based on whether the expression contains an 'anchor' that links the entity to another discourse entity. For instance, in (12) the NP *a painting I made* is anchored to the discourse by the situationally evoked anchor *I*. In (13), the referent of the indefinite NP is BRAND-NEW UNANCHORED (or just 'brand-new'), since it does not include information linking the referent to the co- or context.

(12)   She wanted to buy <u>a painting I made</u>.

(13)   She wanted to buy <u>a painting made by some French artist</u>.

While Prince (1981) restricts 'anchoring' to cases in which the 'anchor' has been explicitly evoked, a case can be made for 'implied anchoring'. In (14), for instance, *one reason* is interpreted with reference to the prior co-text, which provides information for filling in its precise content (i.e. 'one reason for the fact that Catholic but not Mormon pro-lifers picket SLC's two aborturies').

(14)   Catholic pro-lifers picket Salt Lake City's two aborturies; the Mormons don't. Someone said <u>one reason</u> is that Mormons don't like to be linked to anything the Catholics do.   (WB)

At the other end of the taxonomy, evoked, or given, entities can be either textually or situationally evoked. TEXTUALLY EVOKED entities are explicitly mentioned in the prior text, though perhaps by means of a different description, e.g. the second mention of *Charles* and the pronoun *her* (referring back to *Camilla*) in (15).

(15)   Charles and Camilla are a perfect match. So why doesn't <u>Charles</u> marry <u>her</u>?
                                                                                                    (WB)

SITUATIONALLY EVOKED entities refer to information in "the extratextual context, which includes the text itself" (Prince 1981, 236). This can be the speaker/writer or hearer/reader, but also entities that are physically present in the communicative setting (e.g. *this*) and comments referring to the speech exchange itself (e.g. *your earlier question*).

Prince's (1981) most complex category is formed by the INFERRABLES. A discourse entity is inferrable "if the speaker assumes the hearer can infer it, via logical – or, more commonly, plausible – reasoning" (Prince 1981, 236) from another entity in the discourse, a so-called 'trigger entity', e.g. in (16), where *the driver* is inferrable from *a bus*.

(16)   I got on a bus yesterday and <u>the driver</u> was drunk.   (Prince 1981, 233)

In the final analysis, Prince (1981) ranks the different types of discourse-givenness in her taxonomy on a SCALE of ASSUMED FAMILIARITY, illustrated in (i) from most to least familiar (i.e. hearer-old).

(i)   Evoked (Textual | Situational) > Unused > Inferrable > Brand-New
      (Anchored) > Brand-New (Unanchored)

In his text-based model of discourse-familiarity, Kaltenböck (2005) adopts most of Prince's (1981) categories but he reconceptualises them in terms of textual 'retrievability'. The reason, he argues, is that, from an empirical viewpoint, researchers can only take into account the text itself as a finished product of communication, not the mental states of speaker and/or hearer. Hence, Kaltenböck (2005, 127) argues, for instance, that Prince's category of 'unused' information is empirically suspect: determining whether discourse-new information is hearer-familiar is not possible based on the text itself. Therefore, while 'unused' information may have theoretical validity, Kaltenböck (2005) excludes the category from his 'taxonomy of discourse-familiarity', shown in Figure 1.
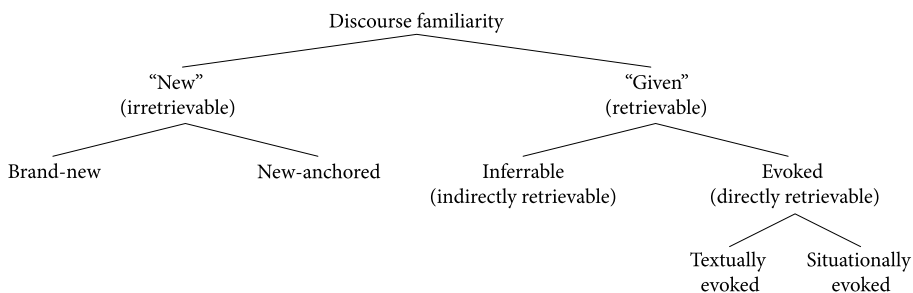

Figure 1. A taxonomy of discourse-familiarity (Kaltenböck 2005, 127)

Kaltenböck's (2005) model of 'retrievability' seems to imply a close connection between information status and (in)definiteness. Prince (1992, 302–303), however, argues convincingly that the two factors may be correlated but they need not coincide. She adduces the following three observations as evidence for her claim.

Firstly, while inferrable entities are normally coded by definite NPs, indefinite NPs can code inferrables as well. This is the case when the indefinite NP introduces a member of a previously evoked set. The set – membership relation can be coded grammatically, e.g. *one* in (17), or it may need to be contextually inferred, e.g. *a page* in (18a). Prince (1992) demonstrates the inferrability of *a page* from *that book* in (18a) by contrasting it with *a cockroach* in (18b): the assumption in (18a)

is that the said page is not just any page but a page from the previously mentioned book; in (18b), by contrast, no such inference can be made for *a cockroach*, which is, therefore, non-inferrable.

(17)   Various reasons are given for Egypt's continued religious tensions […]. <u>One</u> is
the poor state of the economy.   (WB)

(18)   a.   I picked up that book I bought and <u>a page</u> fell out.   (Prince 1992, 306)
b.   I picked up that book I bought and <u>a cockroach</u> fell out.   (ib.)

Secondly, Prince (1992, 302) also notes that definite NPs do not necessarily introduce discourse-old (or retrievable) information. This is illustrated, for instance, by the overall newness of *the people of New Zealand's Chatham Islands* in (19), which was not mentioned in, or inferrable from, the prior context.

(19)   Next to see in the new century were <u>the people of New Zealand's Chatham
Islands</u>.   (WB)

Thirdly, indefinite NPs can also introduce 'given' information. Generic indefinites are a case in point (Prince 1992). While first mentions of generic indefinites are discourse-new, second (and subsequent) mentions can be made either by means of a definite expression (e.g. personal pronoun with the same value for number, gender and animacy) or by means of an indefinite NP, as illustrated in (20). In the second case, the indefinite generic is discourse-familiar.

(20)   A better understanding of the nature of dogs gives us an insight as to why and
when <u>a dog</u> may ignore a call to return: <u>dogs</u> are hard-wired to chase moving
objects.   (WB)

In addition, non-referential indefinite NPs can also draw on given information, e.g. *a subtle-souled psychologist* in (21). While the non-referentiality of NPs like *a subtle-souled psychologist* makes it impossible to establish co-reference between first and second mentions, the description itself is discourse-old. For the purpose of this paper, I will examine the discourse-familiarity of predicate nominatives, like in (21), based on the discourse-familiarity of the nominal description, rather than on the familiarity of its referent.

(21)   Coleridge tries to demonstrate that Shakespeare is a subtle-souled psycholo-
gist. And there's only way [sic] in my opinion in which you can demonstrate
that anybody is <u>a subtle-souled psychologist</u>, that is you look into yourself and
see if it touches you.   (WB)

Taken together, these three observations warrant caution against the idea that (in)definiteness and discourse-familiarity are necessarily interdependent: not only can indefinite NPs, under special circumstances, express familiar informa-

tion, but it is also possible for definite NPs to introduce new information. Therefore, both indefinite and definite NPs can, in principle, have all the different statuses of discourse-familiarity listed by Prince (1992) and Kaltenböck (2005).

Finally, the discussion so far has focussed implicitly on NPs and pronouns, while the discourse status of nominalisations has not yet been mentioned. Nominalisation involves reclassification of a relational category (e.g. infinitives, gerunds, finite clauses), whereby it comes to denote a 'thing' rather than a relation or process.[7] In the case of nominalised finite clauses, for instance, the nominalised clause comes to express a fact or report about a state-of-affairs rather than the state-of-affairs itself, e.g. (22).

(22)   What's the difference between a woman lawyer and a pit bull? Lipstick. […]
       Judy couldn't laugh. She hated that the public made jokes about lawyers.

                                        (WB, as cited in Gentens 2016, 148)

As clausal constituents, nominalisations have a dual discourse status (Kaltenböck 2005, 128; Firbas 1992): on the one hand, they typically consist of multiple individual 'communicative units'; on the other hand, they serve as a single 'communicative unit' at the clause level (Kaltenböck 2005, 128). Their information status is, therefore, assessed not based on the individual component parts but on the composite whole (Kaltenböck 2005, 128). Often, nominalisations include a combination of given and new information, which typically makes them new-anchored. They can, however, be found to introduce referents that span the whole range from evoked to brand-new information, as illustrated by the examples, taken from Gentens (2016, 146–149), in (23)–(26).

(23)   *TEXTUALLY EVOKED*
       Even though *giving birth scares me to death*, I like it that it's frightening. (WB)[8]

(24)   *SITUATIONALLY GIVEN*
       "Look!" One of them pointed. "This boy is naked!" Bardo made his way carefully among the petitioners. […] He hated it that everyone was looking at him.
                                        (WB)

---

7. The terms 'thing' and 'process' are used abstractly (see, for instance, Langacker 1991): 'things' include not only physical objects, but also 'people' and abstract concepts like 'love'; 'processes' include not only actions and events, but also temporalised states (i.e. stative relations located in time, e.g. *he was a doctor*).

8. In Examples (23)–(26), the underlined texts highlights the nominalisation, while the parts in italics indicate either a previous mention of the referent of the nominalisation, e.g. (23), or the anchor from which the referent can be inferred, e.g. (25).

(25)   *INFERRABLE*
Dejected fans orderly shuffled out of bars and homes in the Charlotte area
after the Super Bowl, with most saying they were proud of *the Carolina Pan-*
thers' performance in a 32–29 loss to the New England Patriots. […] "I hate that
it ended this way."   (WB)

(26)   *BRAND-NEW*
De Pouzilhac […] was born in Sète in the south of France. […]. One of five sib-
lings, he hated school. But his father, a director of a wine company, held
lunches with interesting businessman [sic] and felt his children learnt more
there. "I love it that in life if you choose the wrong way you pay a penalty," he
says. "But in school all you did was learn and repeat."   (WB)

In sum, this study will use the levels of discourse-familiarity that Kaltenböck
(2005) distinguishes in his text-based model, i.e. 'brand-new' < 'new-anchored' <
'inferrable' < 'textually' or 'situationally evoked' (in increasing order of familiar-
ity). Like Kaltenböck (2005), I will interpret these levels of familiarity in terms of
their actual discourse-givenness, or -newness, taking care, however, not to equate
the discourse-familiarity of information with the identifiability of referents (as
expressed by (in)definiteness marking).

3.   Method

The research in this study focusses on a comparison of the discourse-embedding
of non-reversed specificational clauses with indefinite variable subject, which
I compare and contrast with (i) predicative clauses with indefinite predicate
nominative, (ii) non-reversed specificationals with definite variable subject, and
(iii) reversed specificational clauses with indefinite variable complement. For
each construction 250 examples were analysed. The examples were taken from
randomised datasets from the 'Times' (written English) and 'British Spoken'
subcorpora of Wordbanks*Online*. These datasets initially consisted of 100,000
concordances, from which – after randomisation – the first 250 examples for each
of the four clause types were extracted, with equal representations for the two sub-
corpora.

   The extraction of the data was based on three recognition criteria for the dif-
ferent clause types. The first was reversibility (or the possibility of a subject –
complement switch), which is only possible for specificational clauses (e.g. *The*
winner is John/him vs *John/he is the winner*), but not for predicative clauses (e.g.
John/he is a doctor vs **A doctor is John/him*). The second test is cleft-alternation.
Specificational clauses can only be clefted in one way since the value of the cop-

ular clause remains value in the cleft (e.g. *It's John who's the winner*). Predicative clauses have no direct cleft alternate (e.g. *It's John who's a doctor* or *It's a doctor that John is*). Finally, while predicative clauses allow for coordination of *be* + complement, e.g. (27), specificational clauses do not, e.g. (28a,b), at least not without imposing a predicative reading on the clause, e.g. (28b).

(27)   Saddam Hussein's regime <u>was a threat and was a danger</u>.   (WB)

(28)   (What soothes pain?)
     a.   *A good option <u>is vitamin C and is aspirin</u>.
     b.   ???Vitamin C <u>is a good option and is a strong antioxidant</u>.

All relevant examples were analysed in terms of the discourse-familiarity of the two NPs, whose semantic roles are those of 'variable' and 'value' in specificational clauses and 'describee' and 'description' in predicative clauses. For the reasons outlined in Section 2, I adopted Kaltenböck's (2005) classification of discourse-familiarity, distinguishing between 'brand-new', 'new-anchored', 'inferrable', 'textually evoked' and 'situationally evoked' information. Extra care was taken to analyse discourse-familiarity as a factor separate from identifiability of referents, as argued in Section 2.

4.   Quantitative and qualitative analysis

Focussing on specificational clauses with indefinite variable NP, the analysis of discourse-embedding will centre on three dimensions of contrast: (i) predicative vs specificational clauses with indefinite NP (Section 4.1), (ii) specificational clauses with indefinite vs definite variable NP (Section 4.2), and (iii) non-reversed vs reversed specificational clauses with indefinite NP (Section 4.3). All three distinctions yielded significant differences with respect to discourse-embedding, of which a general overview is given in Table 1.

4.1   The embedding of predicative and non-reversed specificational clauses

A first comparison between non-reversed specificational clauses with indefinite variable NP and predicative clauses presents a clear difference in the discourse status of the subject in the two construction types ($\chi^2(4)N = 341.9$, $p < .001$; Cramer's V: 0.828) (cf. Table 1a,b; Figure 2).

    More importantly, a comparison between the statuses of the subject and the complement in each clause type indicates that the hypothesised clause-internal

Table 1. Discourse-embedding of the NPs in the four copular clause types

a. predicative clauses

	Describee		Description	
brand-new	10	(4%)	118	(47%)
new-anchored	25	(10%)	85	(34%)
inferrable	17	(7%)	26	(10%)
textually given	180	(72%)	21	(9%)
situationally given	18	(7%)	0	(0%)

b. Non-reversed specificational clauses with indefinite variable

	Indefinite variable		Value	
brand-new	6	(2%)	92	(37%)
new-anchored	174	(70%)	140	(56%)
inferrable	70	(28%)	8	(3%)
textually given	0	(0%)	9	(4%)
situationally given	0	(0%)	1	(0%)

c. Non-reversed specificational clauses with definite variable

	Definite variable		Value	
brand-new	4	(2%)	95	(38%)
new-anchored	114	(45%)	113	(45%)
inferrable	104	(42%)	19	(8%)
textually given	26	(10%)	22	(9%)
situationally given	2	(1%)	1	(0%)

d. Reversed specificational clauses with indefinite variable

	Value		Indefinite variable	
brand-new	118	(47%)	1	(0%)
new-anchored	36	(14%)	115	(46%)
inferrable	9	(4%)	122	(49%)
textually given	83	(33%)	11	(5%)
situationally given	4	(2%)	1	(0%)

contrast between 'given' vs 'new' information is only borne out for predicative clauses, not for specificational ones with indefinite variable NP.

In predicative clauses (Table 1a), the subject entity, with the semantic role of 'describee', tends to convey familiar information, typically textually evoked (72%).
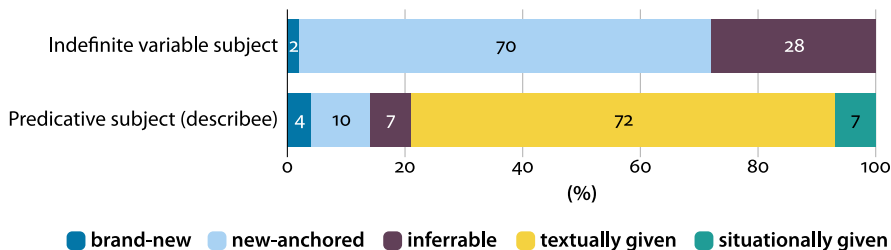
Figure 2. The discourse status of the subjects in predicative and non-reversed specificational clauses with indefinite variable

The complement, on the other hand, describes the subject entity typically in terms of new information, which is either brand-new (47%) or new-anchored (34%). In (29), for instance, the textually given *he* is characterised in terms of the brand-new description *a has-been*; in (30), the already evoked *neighbourhood committees* are described in terms of new information that is anchored to the prior discourse (via *crisis in Argentina*).

(29)   I know that bloke. It's Paul Daniels. I can't stand him. <u>He's a has-been</u>.   (WB)

(30)   Argentina's neighbourhood committees are filling in the gaps left by a severe political and economic crisis. [...] The committees have been sustained by a political class that seems reluctant to take the reins and get the country back on track [...] This suggests that <u>the neighbourhood committees are more a sign of the perpetuation of crisis in Argentina, than the seed of some new movement pointing the way out of it</u>.   (WB)

The typical patterns of discourse-embedding for predicative clauses can be contrasted with those for non-reversed specificational clauses with indefinite variable NP (Table 1b). In this second clause type, the roles of both variable and value are typically realised by NPs conveying new information, predominantly new-anchored information, viz. 70% for the variable and 56% for the value. This is illustrated in (31), where the variable is anchored to the prior discourse via *this result* and the value via *the two countries*.

(31)   US salespeople feel that the pace at which their sales managers require them to perform their job tasks and assignments is more fair than their Japanese counterparts. <u>A possible explanation for this result is the work environment present in the two countries</u>.   (WB)

If we add to these numbers the percentages of brand-new information – which were much higher for the value (37%) than for the variable (2%) –, then both the indefinite variable subject (72%) and its value (93%) are highly likely to express new information. The fact that the role of 'variable' is rarely realised by an NP that

presents *brand*-new information is consistent with claims made in previous studies that the variable is unlikely to be used 'out of the blue' (Declerck 1988), that it serves a discourse-connective function (Mikkelsen 2005), or that the semantic specifications of the variable are interpreted with reference to the local discourse context (e.g. Davidse and Van Praet 2019).

At the same time, the shared preference of indefinite variable subjects and the corresponding values for new-anchored information suggest that the variable and value NPs often have the same level of discourse-familiarity, viz. in 43% of the cases. This goes against previous claims (e.g. Mikkelsen 2005, 135) that the felicity of variable subjects depends on the requirement that they are more discourse-familiar than the value complement (see also Vallduvi 1992, 21). Further evidence against this is found in a handful of specificational clauses (6%), in which the variable subject is of a lower level of familiarity (i.e. 'newer') than the value complement. In (32), for instance, the textually evoked value NP *the country's rail network* is specified for a variable NP that expresses overall new information but contains an inferrable anchor (i.e. *terrorist air attacks in America*).

(32)   Amid fears over air safety [after 9/11], America's neglected rail system is making a comeback, says David Wickers. <u>One of the few beneficiaries of the terrorist air attacks in America has been the country's rail network</u>. With travellers reluctant to fly on domestic routes, Amtrak has reported passenger increases of up to 40% on some of its key services.   (WB)

In sum, predicative clauses and non-reversed specificational clauses with indefinite variable NP are typically embedded in the discourse in different ways. Predicative clauses predominantly have a textually evoked subject referent as the 'describee', of which the predicative complement gives a description offering new information. Such a clear given – new contrast between subject and complement is rarely found in non-reversed specificational clauses, which instead normally introduce overall new information. This goes against the idea that the semantic role of variable is associated with presupposed and therefore discourse-old information (contra, for instance, Declerck 1988), or that it *must* be at least more familiar than the value complement (contra Mikkelsen 2005, 208). Instead, from a discursive viewpoint, non-reversed specificational clauses (with indefinite variable NP) often shift the discourse to a new subtopic, on which the information serving as value elaborates.

4.2   The embedding of indefinite vs definite variable subjects

A second contrast that can be drawn is that between the embedding of non-reversed specificational clauses with indefinite vs definite variable NP. This com-

parison, visualised in Figure 3, brings to light subtle differences in how the (in)definiteness of the variable NP interacts with its discourse status.
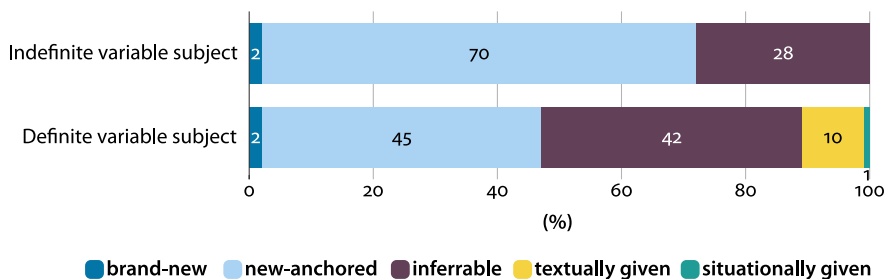

Figure 3. The frequencies of different discourse statuses of (in)definite variable subjects

Like its indefinite counterpart, the definite variable subject has a preference for new-anchored information (45%), e.g. *the remarkable man most responsible for bringing Rushdie to the faith* in (33).

(33) Why did Rushdie embrace the Muslim faith of his persecutors? People have said it was to save his own skin. And Rushdie has joked about it 'as if that was a rather poor thing to want to do' [...] The remarkable man most responsible for bringing Rushdie to the faith was Dr Hesham El-Essawi, [...] one of the first British Muslims to denounce *The Satanic Verses*.   (WB)

The two kinds of variable subject may have a shared preference for new-anchored information, but this preference is much stronger for the indefinite variable NP (70%) than for the definite one (45%). In other words, indefinite variable subjects are mostly concentrated in the category of new-anchored information. Definite variable subjects, by contrast, are distributed almost equally between the categories of new-anchored (45%) and inferrable information (42%). The latter is illustrated in (34) by *the overall winner*, which can be inferred from the previously mentioned *Environmental Awards*. Unlike indefinite variable NPs, definite variable NPs even have a small percentage of uses featuring textually evoked information (10%), e.g. *the emergency* in (35).

(34) During the reception the Lord Mayor presented the annual Environmental Awards: the overall winner was Dr Kathy Lewis, of Hertfordshire University.
                                                                      (WB)

(35) From the moment he arrived, [Taoiseach Bertie] Ahern seemed uncomfortable. After a while, he [...] explained that he would have to rush back to Dublin for an important but unforeseen engagement. On landing, the party officials who shared the return flight were taken aback to discover that the emergency

was, in fact, a pint-drinking session with a group of Ahern's constituency foot soldiers who had been canvassing.   (WB)

A significant difference can, therefore, be observed between indefinite vs definite variable NPs ($p <$ .001, Fisher's Exact Test).[9] However, the difference between the two NPs is by no means discrete. It is not the case, as assumed in the literature, that the difference between indefinite and definite variable NPs is that the former *must* be discourse-new to contrast them with the latter, which supposedly *must* be discourse-old (i.e. inferrable or evoked). Instead, indefinite variable NPs contrast with definite ones in that the former have a strong preference for new-anchored information, whereas the latter are more equally divided between new-anchored (45%) and inferrable (42%) information. In other words, a more accurate conclusion would be that indefinite and definite variable NPs can be characterised as having a strong vs weak preference, respectively, for new-anchored information.

This finding suggests that the choice between construing the variable NP as indefinite or definite is based on other factors than discourse-familiarity alone. An important motivation for choosing one construal over the other is the different implicatures that the two NPs trigger (e.g. Davidse and Van Praet 2019). Since definite NPs imply reference to all instances of a given type in the speech context (i.e. contextual uniqueness) (Hawkins 1978), definite variable NPs trigger an exhaustiveness implicature: the value specifies all the possible items that meet the criteria set up by the variable. Indefinite variable NPs trigger the opposite implicature: since indefinite NPs imply contextual non-uniqueness (Hawkins 1978), indefinite variable NPs trigger a non-exhaustiveness implicature, i.e. other items than the actually specified value could also be specified as values for the same variable. This is evidence that the indefinite variable NP is not something semi-aberrant for which conditions making them acceptable have to be met (e.g. discourse-familiarity); instead, the indefinite variable NP offers a contrastive option to the definite variable NP that is motivated by the implicature of non-exhaustiveness that the indefinite variable triggers.

Finally, the value complements that are specified for indefinite and definite variable subjects also show some (minor) variation in discourse status ($\chi^2(1) = 12.532$, $p = 0.005$; Cramer's V: 0.158)[10] (Figure 4). The value complements specified for definite variable NPs have a slightly higher percentage of inferrable (8%) and evoked (9%) information than the ones for indefinite variable NPs

---

9. Because of the low numbers for brand-new information and situationally evoked information, the *p*-value was calculated by means of a Fisher's Exact Test.

10. Because of the very low numbers for situationally evoked information, they were here taken together with textually evoked information as forming one 'evoked' category.

(respectively, 3% and 4%). Overall, however, the value complement is most likely to be either new-anchored or brand-new in both clause types.
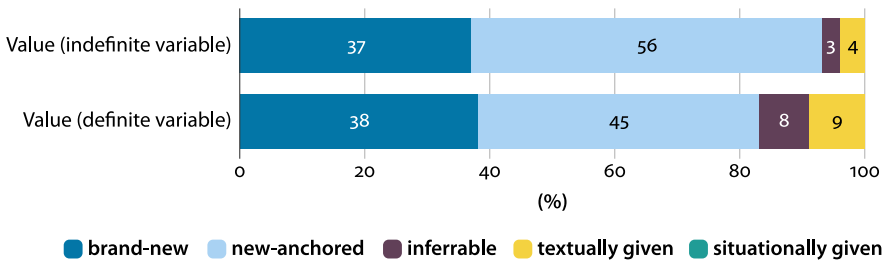

Figure 4. The discourse status of the value complement in non-reversed specificational clauses with indefinite vs definite variable

To conclude, no absolute contrast in discourse-embedding appears between specificational clauses with indefinite vs definite variable NP. Instead, the (in)definiteness of the variable NP interacts with its discourse status in more subtly different ways. The variable subjects in both clause types have the same preference, not for 'old', or evoked, information – as has typically been argued (e.g. Declerck 1988, 14, 19) – but for new-anchored information. Indefiniteness strengthens this preference, while definiteness makes it less pronounced. Therefore, it would seem that semantic role, i.e. being 'variable', takes precedence in influencing the discourse-familiarity of the NP: that is, variable NPs – whether definite or indefinite – tend to be new-anchored. The (in)definiteness of the 'variable' NP plays an additional mediating role: indefiniteness adds to the preference for new-anchored information; definiteness tempers this preference. Therefore, the choice between construing the variable NP as indefinite or definite does not hinge (solely) on discourse-familiarity. Instead, I argued that this choice is motivated by the contrasting implicatures of exhaustiveness and non-exhaustiveness they trigger. In sum, therefore, it is not the case that the variable *must* be discourse-old (contra, for instance, Declerck 1988, 14, 19; Mikkelsen 2005, 135).

4.3   The embedding of non-reversed vs reversed specificationals with indefinite variable NP

The third and final comparison focusses only on specificational clauses with indefinite variable NP and examines the differences between the ones in which the variable is subject and those where it is complement. Significant differences can be found for the discourse status of each of the two NPs serving as variable and value in the two clause types. The results for the value in particular will be

advanced as offering insights into what could motivate the speaker to construe the specificational clause in its reversed form rather than in its default non-reversed form.

The differences in discourse status between variable subject vs complement are visualised in Figure 5. These differences are significant ($\chi^2(2) = 38.082$, $p < 0.001$; Cramer's V: 0.278). Interestingly, they go against the expectation that the clause-initial position of indefinite variable subjects makes them more likely than variable complements to link up with the preceding text and, hence, to have an overall higher degree of discourse-familiarity.
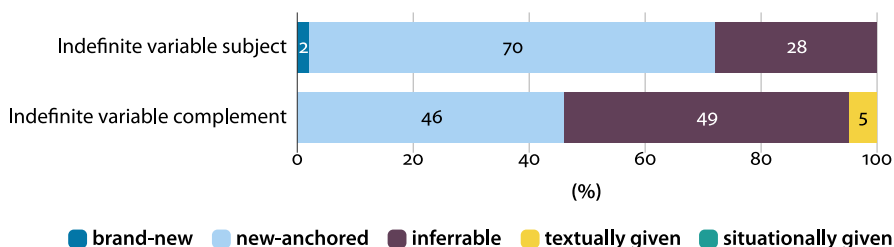
Figure 5. The discourse status of indefinite variable subjects vs indefinite variable complements

Instead, when construed as complement, the indefinite variable NP does not show the unequivocal preference for new-anchored information (46%) that it does as subject (70%). Rather, the variable complement is most likely to express inferrable information (49%), e.g. *causative factors* in (36).

(36)   Repeated sports training and performance may lead to inflammation and stress fractures. <u>Poor technique such as running flat-footed or striking the heel hard are causative factors</u>. (WB)

In general, the indefinite variable complement has a slight preference for recoverable information (54%), while the indefinite variable subject favours new information (72%) ($\chi^2(1) = 32.865$, $p < .001$; Cramer's V: 0.26).[11] Therefore, for this particular dataset, the findings go against the expectation, based on previous studies for different clause types (e.g. Birner 1994), that information that is presented more towards the end of the clause tends to be newer than clause-initial information.

---

11. These are the results of a chi-square test for a binary recoverable vs non-recoverable contrast.

The discourse status of value complements vs subjects further complicates the picture (Figure 6). Here, too, significant differences can be found ($\chi^2(3) = 125.86$, $p < 0.001$; Cramer's V: 0.502).
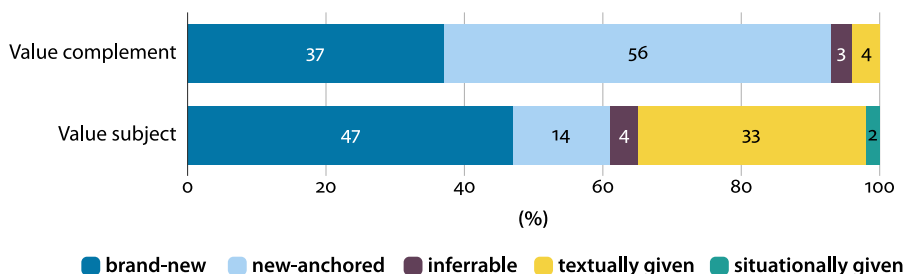

Figure 6. The discourse status of the value in non-reversed vs reversed specificational clauses with indefinite variable

In non-reversed clauses, the expectation that the value is most likely non-recoverable from the prior context is confirmed (93%): the value complement conveys either brand-new (37%) or new-anchored (56%) information. Reversed clauses, however, are split between two patterns: one where the value subject is textually evoked (33%) – e.g. *this* in (37) – and another in which it is brand-new (47%) – e.g. *Peter Roberts*, etc. in (38). In the remainder of this section, the two different patterns for reversed clauses will be described in terms of two different discourse-pragmatic purposes for which the reversed specificational clause can be used.

(37)   What Shelley is in a sense trying to give us is not so much a tranquil reflection of experience as much as the experience itself white hot. Also, it is splendidly lyrical this opening, so that it is almost as if the words […] create the experience itself. And this is one of the things that Shelley tries to do. (WB)

(38)   Small companies are also feeling the pinch of rationalisation by service providers such as banks and post offices. They point to poor co-ordination between government departments, rural development agencies and local authorities. Peter Roberts, a potter from Tywyn in Mid Wales, is one business owner who has felt the full force of rural decline. Having built his business over two decades […], he finally shut up shop this summer. (WB)

Firstly, the fact that a large number of value subjects express textually evoked information (33%) corresponds to what Halliday (1967, 205) and Birner (1994) have argued is the unmarked presentation of recoverable before non-recoverable information. The positioning of recoverable information in initial position creates coherence in the text and ensures a smooth transition between sentences. In most

such cases, the value subject is a demonstrative pronoun, found in 26% of the reversed specificational clauses, e.g. (39).

(39) (NARRATOR) Why is it, when we've got, across the rest of Eastern Europe, countries marching away from the old-style Communism – why does Bulgaria seem to be so different? (RICHARD CRAMPTON) Well, first of all, the Bulgarian Communist Party […] has a long tradition […] Its relationship to the Soviet Union has been popular amongst most of the Bulgarian population. So it has a good deal of historical tradition and strength behind it. <u>That's one reason</u>. Secondly, I think many people were frightened not to vote Communist.   (WB)

An important reason for the frequent use of demonstratives is the fact that they can establish anaphoric reference to information expressed by full sentences and even larger stretches of discourse (Halliday and Hasan 1976, 52), as is the case in (39). Such 'extended text reference' provides a means for summarising (a) prior proposition(s) as forming one single piece of information (Halliday and Hasan 1976, 52). From a discursive viewpoint, reversed specificationals with demonstrative values often summarise what was said in the prior co-text as a way of closing a subtopic, rather than shifting to a new topic like non-reversed clauses often do.

  Secondly, in the other and more frequent pattern, the value subject conveys brand-new information (47%), so that the unmarked presentation of recoverable before non-recoverable information is overridden (Birner 1994). In (40), for instance, the normal 'flow of information' from given to new is turned around, with the brand-new value subject *French farmers* preceding the new-anchored variable *one main reason* etc. The presentation of new information in initial position – where normally coherence is signalled (e.g. Bolinger 1952; Halliday 1967) – results in a clause-initial 'peak' of information (i.e. so-called 'thematic' prominence). This is expected to require more 'processing effort' (Wilson and Sperber 2006), since the initial new information provides no direct clues as to how the clause relates to the prior discourse. The new value thereby grasps the hearer's attention, for instance to indicate a (sudden) shift towards a new (sub)topic.

(40) The single factor most likely to stop the talks [about rich countries helping poorer ones] dead is the European Union's stance that the "environment" must be on the table, too. This is code for wanting the right to ban imports […] <u>French farmers are one main reason why the EU has taken this stance</u>. Offering French forces to the coalition will no doubt continue to boost the French President in the polls.   (WB)

The high frequency of this 'marked' presentation of information suggests that givenness is not a decisive factor – or at least not the only deciding factor – for making the value the subject of the clause. Instead, the option for a reversed or

a non-reversed construal is a meaningful choice, based on the different 'perspectives' that the choice of subject offers on the specificational relation. As Halliday (1985, 76) argues, the subject is the 'modally responsible' element, i.e. the element "on which the validity of the information is made to rest" and in which "is vested the success or failure of the proposition". In that light, specificational clauses allow for two kinds of proposition and can make two kinds of claim, depending on whether the variable or value is subject. In non-reversed clauses (where the variable is subject), the validity of the proposition is assessed as follows: 'are the criteria stipulated by the variable such that they are properly exemplified by the value' (Van Praet 2020, 117)? In reversed clauses, the value is responsible for the validity of the proposition, as follows: 'are the qualities of the value such that they meet the criteria stipulated by the variable' (ib.)? Because the subject is the modally responsible element in the clause, the subject referent is likely to be a familiar entity. Therefore, the use of 'new' value subjects is marked not only because it (often) results in the presentation of new information in initial position, but also because new referents are normally not assigned modal responsibility.

An effect of this marked pattern, I propose, is that the reversed specificational clause can thereby acquire mirative overtones. Mirativity is commonly associated with the expression of 'surprise' or 'sudden realisation' (Delancy 1997, 33), but it has more recently been interpreted as also including meanings of 'unsurprisingness', 'self-explanatoriness', or 'obviousness' (Aikhenvald 2012, 432; Adelaar 2013). These meanings of 'sudden realisation' and, more importantly, of 'obviousness' and 'self-explanatoriness' seem to arise when reversed specificational clauses take new information as value subject. By vesting the validity of the proposition in a new value, the value is presented as information that can be readily accommodated by both speaker and hearer. In (41), for instance, the value subject *the conspicuous consumption of Donatella Versace* is discourse-new information, but the suggestion that it qualifies as something we can rely on 'pretty safely' is presented as plain and clear.

(41)   What can we rely on in these uncertain, straitened times? <u>The conspicuous consumption of Donatella Versace would be a pretty safe option</u>. Holed up in a suite at the Paris Ritz, Donatella window-dressed the place with huge bunches of white lilies and roses, Diptyque candles, iced buckets of Cristal and Melba toast piled high with caviar. She never touched any of it.   (WB)

Mirative meanings, such as 'obviousness' or 'self-explanatoriness', need not be coded explicitly, but – as acknowledged in other studies (e.g. Beltrama and Hanink 2019) – they can also arise by implication via linguistic expressions that normally express other meanings than mirativity. That seems to be the case in the clauses discussed here: in reversed specificational clauses, it is the combination of

the 'thematic prominence' of the initial new information and the modal responsibility of the subject that gives rise to the implication of mirativity.

This mirative meaning is particularly common when the value is realised by what can be considered 'unused' information in Prince's (1981) model. Unused information is brand-new to the discourse but assumed to be familiar to the hearer, e.g. *Saddam Hussein's Iraq* in (42). Such unused discourse referents have homophoric reference: they are retrievable not from the prior discourse but from cultural knowledge that is assumed to be shared by interlocutors of the same community (Martin 1992, 121). The information expressed by such unused values is, therefore, readily acceptable, so that its discourse-newness poses no problem for assessing the validity of the proposition. This makes unused values more likely to be given modal responsibility, and their role in the specificational clause is more likely to be presented as 'self-explanatory' or 'obvious'. This is evidenced, for instance, by the use of *obviously* in (42).

(42)   If the war on terrorism is to succeed, it must target not just terrorist groups but the states that support them and might equip them with the means to commit mass murder. <u>Saddam Hussein's Iraq was obviously one such state</u>, a rogue regime if there ever was one.   (WB)

The mirative effect comes out even more strongly in dialogue. In (43), for instance, Speaker A specifies *Brian Statham* as an example of a 'Lord's performer', which Speaker B confirms. (The use of level tones itself – i.e. by Speaker B when saying *yes* – has been associated with signalling 'obviousness' as well, e.g. Halliday 1970.)

(43)   A.   ^he's [Graham McKenzie] probably one of these :Lords perf/\ormers#
               ((1 syll)) ^<u>Brian !St/\atham was {\one {^w\asn't *he#}#}#* who
               ^/\always ^did well at L\ords#}#
         B.   **^y=es#** *^y=es#* (LLC)

In total, 59 examples, or 24% of the value subjects, qualify as presenting 'unused' information. Other cases are more ambiguous. In (44), for instance, the value subject *the La Cage district of Marseille* is new to the discourse. The suggested 'obviousness' of the exemplification implies, however, that *the La Cage district of Marseille* is assumed to be shared knowledge. The writer, not knowing who their actual audience will be, seems to exploit the potential mirative effect of the reversed clause to present the 'La Cage' district *as if* it is an 'obvious' example. While the reader may or may not be familiar with 'La Cage', they are thereby informed that the reputation of the district makes it self-explanatory that 'La Cage' qualifies as a 'ghetto'.

(44) There has been strong criticism of any European adoption of the reformulated American concept of the 'ghetto'. The word 'ghetto' has been commonly used in France in academic as well as popular literature […] Certainly French cities […] have witnessed the emergence of areas of very strong minority group concentration. <u>The La Cage district of Marseille is an obvious example</u>.　　(WB)

To conclude this section, the findings from the discourse-embedding of predicative and specificational clauses challenge existing assumptions that the two clause types can be described, or even defined, in terms of a given – new contrast. Instead, while the semantic roles of describee and description and of variable and value set preferences for different discourse statuses in specific contexts of use, these two dimensions of linguistic structure are not determined by one another.

The corpus study revealed significant differences between the discourse-familiarity of the NPs in the four clause types. These differences can be analysed in terms of three main contrasts. First, predicative and specificational clauses are integrated in the discourse in typically different ways. Predicative clauses often comment on a given subject, of which a new description is given. Non-reversed specificational clauses, by contrast, typically shift the discourse in a new direction by raising a new (or inferrable) (sub)topic with the variable/subject, on which the (equally new) value then elaborates.

Second, less substantial differences are found between non-reversed specificational clauses with indefinite vs definite variable NP. The first are typically new-anchored, while the second have a higher preference for inferrable information. However, no strong contrast in discourse-embedding was found between these two clause types. The difference between the two kinds of variable is motivated primarily by the different implicatures they trigger, rather than by discourse-embedding. This was put forward as evidence that indefinite variable NPs are not something semi-aberrant for which conditions making them acceptable need to be specified. Instead, the indefinite variable NP offers a contrastive alternative to its definite counterpart, motivated by the implicature of non-exhaustiveness triggered by the indefinite NP.

Finally, more substantial differences were found between non-reversed and reversed specificational clauses. Reversed specificationals are split between two preferred patterns of discourse-embedding: one in which the value subject expresses brand-new information and another in which it picks up on textually given information. In the first pattern, the value subject often introduces 'unused' information, i.e. information that is new to the discourse but assumed to be familiar to the hearer. The use of such unused entities as subject was argued above to invite a special kind of mirative meaning, viz. one of 'self-explanatoriness'. In the second pattern, the value subject, often a demonstrative, was analysed as having 'extended text reference', i.e. summarising propositions and longer stretches

of discourse and presenting them as one abstract piece of information. Like non-reversed specificational clauses, this second pattern can have a 'discourse-organising' function, not to shift the discourse to a new topic but to summarise the previous discourse as a conclusion to the topic identified by the variable.

5.   Conclusion

This paper studied the discourse-embedding of specificational clauses to examine existing assumptions that the specificational variable – value relation entails an intrinsic given – new contrast. Three points of comparison were made, viz. (i) between specificational and predicative clauses, (ii) between specificational clauses with indefinite vs definite variable NP, and (iii) between non-reversed and reversed specificational clauses.

The findings challenged claims in previous studies that, while no specific discourse conditions apply to predicative clauses, specificational clauses are characterised by an intrinsic contrast between given 'presupposed' variables and new 'focal' values (e.g. Chomsky 1969; Akmajian 1970; Declerck 1988). These assumptions were contradicted by evidence from the corpus study, where both the variable and the value showed much more variation in their actual discourse-familiarity.

Against previous assumptions, this study argued that the semantics of predicative and specificational clauses form a separate layer of coded meaning and cannot be conflated with the dimension of discourse structure. While the semantic roles of the describee and description and of the variable and value may set different preferences for specific discourse statuses, these semantics do not determine how the different roles are integrated in specific contexts of use, or vice versa. Since specificational clauses exhibit variation in their prosodic focus marking (Van Praet and O'Grady 2018; Van Praet 2019) as well as in their discourse-embedding, previous accounts of specificational clauses and of information structure are in need of revision, so that the different layers of meaning that have been conflated are separated out.

Finally, by interpreting discursive organisation as influenced by multiple factors, this study showed that not only the semantics of predication and specification, but also the grammatical functions (viz. subject and complement) of the variable/value as well as the (in)definiteness of the variable NP interact with the familiarity of information in the larger discourse. None of these factors, however, exclusively determines the actual discourse-embedding of the copular clause. By looking in detail at the specific discourse patterns for the various clause types, this study has contributed to a better insight into the interaction between the clause

types' semantic functions and the actual discursive functions they serve in specific contexts of use.

Funding

Acknowledgements

References

Adelaar, Willem F. H. 2013. "A Quechuan Mirative?" In *Perception and Cognition in Language and Culture*, ed. by Alexandra Y. Aikhenvald and Anne Storch, 95–109. Leiden: Brill. https://doi.org/10.1163/9789004210127_005

Aikhenvald, Alexandra Y. 2012. "The Essence of Mirativity." *Linguistic Typology* 16(3), 435–485. https://doi.org/10.1515/lity-2012-0017

Akmajian, Adrian. 1970. Aspects of the Grammar of Focus in English. MIT Ph.D. dissertation.

Beltrama, Andrea, and Emily A. Hanink. 2019. "Marking Imprecision, Conveying Surprise: *Like* Between Hedging and Mirativity." *Journal of Linguistics* 55(1), 1–34. https://doi.org/10.1017/S0022226718000270

Birner, Betty J. 1994. "Information Status and Word Order: An Analysis of English Inversion." *Language* 70(2), 233–259. https://doi.org/10.2307/415828

Birner, Betty J. 1996. *The Discourse Function of Inversion in English*. New York, NY: Garland.

Bolinger, Dwight. 1952. "Linear Modification." *PMLA* 67(7), 1117–1144. https://doi.org/10.2307/459963

Chomsky, Noam. 1969. *Deep Structure, Surface Structure, and Semantic Interpretation*. Bloomington, IN: Indiana University Linguistics Club.

Davidse, Kristin, and Wout Van Praet. 2019. "Rethinking Predicative Clauses with Indefinite Predicate and Specificational Clauses with Indefinite Variable: A Cognitive-Functional Account." *Leuven Working Papers in Linguistics* 38: 1–36.

Declerck, Renaat. 1988. *Studies on Copular Sentences, Clefts, and Pseudo-Clefts*. Dordrecht: Foris. https://doi.org/10.1515/9783110869330

Delancey, Scott. 1997. "Mirativity: The Grammatical Marking of Unexpected Information." *Linguistic Typology* 1(1), 33–52. https://doi.org/10.1515/lity.1997.1.1.33

Firbas, Jan. 1992. *Functional Sentence Perspective in Written and Spoken Communication*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511597817

Gentens, Caroline. 2016. The Factive–Reported Distinction in English: Representational and Interpersonal Semantics. KU Leuven Ph.D. dissertation.

Halliday, M.A.K. 1967. "Notes on Transitivity and Theme in English: Part 2." *Journal of Linguistics* 3(2), 199–244. https://doi.org/10.1017/S0022226700016613

Halliday, M.A.K. 1970. *A Course in Spoken English: Intonation*. Oxford: OUP.

Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. 1st ed. London: Edward Arnold.

Halliday, M.A.K., and Ruqaiya Hasan. 1976. *Cohesion in English*. London: Longman.

Hawkins, John. 1978. *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction*. London: Croom Helm.

Heycock, Caroline. 2012. "Specification, Equation, and Agreement in Copular Sentences." *Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 57(2), 209–240. https://doi.org/10.1017/S0008413100004758

Higgins, Francis R. 1979. *The Pseudo-Cleft Construction in English*. London: Routledge.

Kaltenböck, Gunther. 2005. "*It*-Extraposition in English: A Functional View." *International Journal of Corpus Linguistics* 10(2), 119–159. https://doi.org/10.1075/ijcl.10.2.02kal

Lambrecht, Knud. 1994. *Information Structure and Sentence Form: Topic, Focus, and the Mental Representations of Discourse Referents*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511620607

Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar, Vol. 2: Descriptive Application*. Stanford, CA: Stanford University Press.

Lehmann, Christian. 2008. "Information Structure and Grammaticalization." In *Theoretical and Empirical Issues in Grammaticalization*, ed. by Elena Seoane and María José López-Couso, 207–229. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/tsl.77.12leh

Martin, James. 1992. *English Text: System and Structure*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/z.59

Mikkelsen, Line. 2005. *Copular Clauses: Specification, Predication and Equation*. Amsterdam/Philadelphia: John Benjamins. https://doi.org/10.1075/la.85

Moro, Andrea. 1997. *The Raising of Predicates: Predicative Noun Phrases and the Theory of Clause Structure*. Cambridge: CUP. https://doi.org/10.1017/CBO9780511519956

Patten, Amanda L. 2012. *The English It-Cleft: A Constructional Account and a Diachronic Investigation*. Berlin/New York: De Gruyter Mouton. https://doi.org/10.1515/9783110279528

Prince, Ellen F. 1981. "Towards a Taxonomy of Given–New Information." In *Radical Pragmatics*, ed. by Peter Cole, 281–297. New York, NY: Academic Press.

Prince, Ellen F. 1992. "The ZPG Letter: Subjects, Definiteness and Information Status." In *Discourse Description: Diverse Linguistic Analyses of a Fund-Raising Text*, ed. by William C. Mann and Sandra A. Thompson, 295–325. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.16.12pri

Stalnaker, Robert. 1973. "Presuppositions." *Journal of Philosophical Logic* 2(4), 447–457. https://doi.org/10.1007/BF00262951

Vallduvi, Enric. 1992. *The Informational Component*. New York/London: Garland.

Van Praet, Wout. 2019. "Focus Assignment in English Specificational and Predicative Clauses: Intonation as a Cue to Information Structure?" *Acta Linguistica Hafniensia* 51(2): 222–241. https://doi.org/10.1080/03740463.2019.1641318

Van Praet, Wout. 2020. English Specificational and Predicative Clauses: A Functional-Cognitive Account of their Representational and Textual Organisation. KU Leuven and UNamur Ph.D. dissertation.

Van Praet, Wout, and Gerard O'Grady. 2018. "The Prosody of Specification: Discourse Intonational Cues to Setting up a Variable." *Journal of Pragmatics* 135: 87–100.
https://doi.org/10.1016/j.pragma.2018.07.013

Wilson, Deirdre, and Dan Sperber. 2006. "Relevance Theory." In *The Handbook of Pragmatics*, ed. by Laurence R. Horn and Gregory Ward, 607–632. Malden, MA: Blackwell.
https://doi.org/10.1002/9780470756959.ch27

Publication history