

RESEARCH ARTICLE

A discourse-functional approach to theticity, subtypes of integrated relative clause and extraposition in English

Francis Cornish 

University of Toulouse-Jean Jaurès 2, France
Email: francis.cornish@univ-tlse2.fr

(Received 1 November 2024; revised 29 October 2025; accepted 11 November 2025)

Abstract

The article is an attempt to develop Francis and Michaelis' (F&M) (2014, 2017) account of 'relative clause extraposition' (RCE) in English, in terms of a more discourse-oriented dimension. On the basis of a corpus study, these authors select certain constituent types, enabling a comparison between configurations with and without RCE orderings. The result is a 'prototypical' sequence of constituent types that is claimed to predict whether RCE is felicitous or not.

To further develop this analysis, the present article puts forward a three-way distinction, in terms of their degree of communicative dynamism, amongst presupposed (i.e. 'grounded') restrictive RCs, non-presupposed RRCs and 'a-restrictive' RCs (neither restrictive nor (strictly) non-restrictive). Only the non-grounded RCs result in a felicitous utterance when extraposed, since it is only such RCs that may realise a presentational function via RCE ordering. More generally, it is shown that the three main sentence-internal factors claimed by F&M to favour RCE derive from the thetic ('all-new') information-structure status of RCE-containing utterances: thus the key features highlighted are the expression-level reflection of the more basic Information Structure articulation involved in each case.

Keywords: integrated relative clauses; extraposed relative clauses; thetic vs topic–comment articulations; predication; focus

1. Introduction

The article aims to show that, though it is indeed valid on its own terms, Francis & Michaelis' (henceforth F&M) (2014, 2017) lucid demonstration of the multiple factors that combine to favour the extraposition in production of integrated (i.e. NP-internal) relative clauses in English, lacks a vital dimension: namely, the full discourse contribution in terms of information structure of the relative clause and its immediate co-text that is assessed for this property: see [section 4.1](#) for a presentation. It argues that for the demonstration to be complete, central account needs to be taken of the latter dimension, as well as, more generally, of the utterance context presiding over the possible use of an NP from which a relative clause may be extraposed. For as will be shown, each subtype of NP-integrated relative clause has specific semantic-pragmatic properties, which are harnessed by speakers and writers in order to realise particular discourse functions.

The structure of the article is as follows. [Section 2](#) is a presentation of English relative clauses, whether integrated *in situ*, extraposed or non-restrictive, as well as of the extraposition

of other types of constituent more generally, foreshadowing the argumentation to come. There follows a characterisation of three subtypes of NP-integrated relative clauses, developing Cornish's (2018) discourse-functional analysis (section 3). The purpose of this three-way distinction is to prepare the later demonstration, presented in section 5, that only two of these subtypes, the non-presupposed ones (see below), may give rise to extraposition.

Section 4 is a synthesis of F&M's (2014, 2017) claims, made largely on the basis of expression-level factors, regarding an integrated relative's extraposability in a clause realisation. Section 5 (section 5.1) then compares the versions of the integrated RCs (henceforth 'IRCs') from section 3 with a possible extraposed variant of each, and re-examines in this light the examples presented by F&M (2014), as well as their claims. Section 5.2 then places the RCE data presented in section 5.1 within an information-structure context, characterised briefly in sections 2 and 4.1.

It appears that it is only one of the three IRC subtypes that is assumed by F&M to give rise to RELATIVE CLAUSE EXTRAPOSITION (RCE), namely 'A-RESTRICTIVE' RCs (neutral as between restrictive and strictly non-restrictive RCs) – though, as will be shown, non-presupposed RRCs may also be extraposed. Both of these subtypes of IRC function to highlight via extraposition the key information load of their containing utterance, a load which outweighs the contribution of the predicative component.

The article claims that all the characteristic sentence-internal features of RCE isolated by F&M (2014, 2017) may be more revealingly analysed as centrally flowing from the thetic character of utterances involving this type of NP discontinuity. Briefly, a 'thetic' act of utterance consists in asserting the existence of a situation presented as all-new information for the addressee or reader, as in *There's a badger eating your runner beans!* See also the attested example (i), note 8, in section 4.1.

2. Situating IRCs with respect to non-restrictive relative clauses, and common characteristics of extraposition

Let us look, first of all, at some of the distinctive features of RCs in English.

A restrictive (and hence integrated) relative clause acts as a postmodifier of its head noun within an NP. Its main characteristic can be revealed by comparing it to a superficially similar construction, where the introducing morpheme in each NP-integrated construction is *that* or zero:

- (1) (a) The worry *that we might not get to the airport on time* was debilitating.
- (b) The accommodation *that we stopped at on our last journey* was already fully booked.

One test of the distinction consists in deleting the introducing morpheme in each case, and judging whether the resulting clause structure constitutes a complete sentence:

- (1) (a') We might not get to the airport on time.
- (b') *We stopped at on our last journey.

Clearly, the complement (or content) clause introduced by *that* in (1a) does correspond to a complete sentence, expressing a full proposition, its semantic function being to convey the nature of the 'worry' denoted by the head noun. However, the relative clause in (1b) does not: for there is clear evidence of a gap between *at* and *on*, corresponding to what Dik (1997: 3) calls the 'RELATIVISED VARIABLE'. This gap occurs in a particular functional position in the clause, where its functional-syntactic role is reflected by those RC-marking *wh*-pronouns that show case information: i.e. *who*, *whom* and *whose*; see also Chierchia & McConnell-Ginet (1990: 331). Again following Dik (1997), we may call the various types of IRC introducers (*wh*-expressions, subordinator *that* and zero) 'RELATIVE MARKERS', or alternatively 'RELATIVISERS'.

Semantically, the IRC in (1b) serves to narrow down the denotation of its head noun, and to actually co-create discursively the referent of the full NP which contains it.

Regarding strictly non-restrictive or ‘appositive’ relatives (henceforth NRRCs: see (2) below for an example), their differences from IRCs are as follows. NRRCs carry a separate intonation contour with respect to the clause to which they are related, whereas IRCs in relation to the matrix sentence do not: it is the complex NP of which they are an integral part that determines their (dependent) contour. Furthermore, NRRCs are interpreted as anchored to a given phrasal constituent (NP, AP, VP, PP or clause) in their immediate environment, whereas IRCs depend on a lexical constituent, the head noun of the containing NP. This reflects the fact that NRRCs relate to an already established (i.e. ‘given’) discourse entity, whereas restrictive IRCs serve to create the referent targeted by their containing NP as a whole. Thus NP-internal RCs such as RRCs are ‘integrated’ ones, while NRRCs are non-integrated, i.e. NP-external.

Moreover, whereas an NRRC carries an independent illocution in relation to the preceding clause, and hence constitutes a distinct discourse unit, an IRC does not. Two further distinguishing features are the fact that quantified and non-referring head nouns are acceptable with IRCs, but such-headed NPs are not for NRRCs; in addition, finite IRCs may occur with subordinator *that* or zero as well as *wh*-expressions, while NRRCs show a clear preference solely for the last-mentioned subtype of relativiser: see table 1 in section 3.3 and de Haan’s (1989: 126) results from his corpus survey.

An attested written extract forcing us to distinguish an NRRC from an IRC follows:

- (2) # Reginald decides he will do anything for love *which* [sic] *leads to some unexpected consequences*. (*Radio Times* 03–09.08.24, p. 98)

Here, there is no comma between the words *love* and *which* (an evident typo, for the comma would correctly signal that the RC at issue is a non-restrictive relative). In fact, the content of the RC introduced by *which* does not restrict in any way the denotation of the noun *love*; rather, the RC as a whole continues the event evoked by the initial clause, the *wh*-pronoun picking this eventuality up in order to advance the narration. As such, it requires a preceding comma in the written form and a disjuncture in the spoken, in order to achieve coherence: cf. also Huddleston & Pullum *et al.* (2002: 1035). The RC, therefore, must correspond to an NRRC, but not to an IRC.

Concerning possible extraposed RCs, these can only be produced by extraposition from an originally NP-internal position, whether subject or non-subject. Hence they are necessarily ex-IRC, their extraposed status being motivated by the intention to highlight what the speaker or writer considers to be the utterance’s most significant information load in context. An attested example of an RCE is given in (3), where the RC is separated from its head noun by the adverb *tonight*:

- (3) I’ve been asked out by this actor/journalist tonight *that I went to see last week*. [...]. (ICE-GB:W1B–010 #064:2; written extract)

NRRCs, however, are not candidates for RCE, since by definition, they cannot be analysed as originating from such a position – even though, like RCEs, as in (3), they are detached from their contextual anchor. In terms of discourse structure, apart from ‘continuative’ NRRCs, as in (2) (cf. Looock 2010), they are in effect parenthetical constructions, often following a topic-introducing clause, providing context, further-identifying or explanatory information.

Let us now briefly characterise extraposition more broadly, in discourse terms. Other types of extraposition include that of clausal subjects (finite *that*-clauses, infinitival or gerundive constructions: cf. Miller, 2001) or of prepositional phrases, as in (4):

- (4) [Out] of 1m rides, just 390 crimes were committed *in which hitchhikers were the victims*. (*Observer Online*, 04.09.25)

See also example (24) in [section 5.2](#). As for clausal subjects, Miller claims that where they are ‘discourse-new’, they must be extraposed. This is consonant with our claims vis-à-vis RCE in English (see [section 5](#)), and is clearly what is implied in F&M (2014). However, if they are ‘discourse-old’, Miller claims they can either be extraposed or remain *in situ*. If they remain *in situ*, however, they must represent ‘discourse-old’ information.¹

The concepts ‘discourse new/old’ assumed by Miller relate, respectively, to whether the information conveyed via a given unit has not already been mentioned in prior discourse, or has been. As we shall see in [section 4.2](#), F&M (2014) also use this criterion in characterising their concept of ‘superset mention’. However, it seems that in fact this distinction corresponds, rather, to what is usually termed ‘addressee-new/old’ information, distinct from ‘discourse-new/old’ information *stricto sensu*. For a speaker is not totally constrained by whether or not some entity or situation has been mentioned or is inferable from context; they may well produce a focus-oriented utterance involving LOCATIVE INVERSION, for example, in part on the basis of ‘addressee-old’ constituents: here, the preposed locative phrase links up with the prior context, whereas the main clause presenting the subject–verb inversion conveys discourse-new information: see example (23) in [section 5.2](#) below. As we shall see (in [section 3.1](#) in particular), the notion ‘discourse-old’ adopted in the present article depends, rather, on the fact that the information conveyed is pragmatically presupposed. It is noteworthy that the construction created by extraposing a clausal subject is an impersonal one (see (5) below), which as such is one possible realisation of athetic (‘all-new’) articulation, whose principal property is to assert the existence of a situation: see [section 4.1](#) below. Its subject is expletive *it*, which is non-referential, so cannot participate in a ‘Topic–Comment’ articulation.

- (5) It’s sad *that the number of votes in the last election was so low*.

Where the clausal constituent remains in subject position, then we may indeed speak of a Topic–Comment articulation, the ‘Comment’ representing a predication applied to the topical subject entity. See [section 4.1](#) below for further considerations on thetic and topic–comment articulations, and their relevance to IRC extraposition.

Let us look now more closely at the IRC subcategory of relative clauses.

3. Three subtypes of integrated relative clause

There are in fact several subtypes of IRCs in English, viewed in terms of their discourse functionality.² First, pragmatically presupposed restrictive relative clauses ([section 3.1](#)),

¹ In support of this claim, Miller (2001: 687) gives an attested written example where the clausal subject NP is extremely long (example (7), some 36 words long); but due to the fact that it is presupposed of the NP’s referent (it summarises the preceding 40 lines of the containing text, converting the passage into a discourse referent: Miller 2001: 687), it serves to ‘ground’ the Topic–Comment utterance at issue to that extent; hence it is nonetheless felicitous. As such, it is not extraposable. See [section 5](#) for further evidence of this constraint in connection with RCE, which runs counter to F&M’s (2014) stance on very long IRCs: namely, that for the authors, this status renders them *ipso facto* extraposable (see [section 4.2](#) below).

² Here we follow Cornish (2018: 452, fig. 1), who characterises the various RC subtypes the author discusses in terms of their respective degrees of ‘communicative dynamism’; that is, the extent to which each subtype advances the communication being established: presupposed restrictive RCs are the least ‘dynamic’ in this respect, and ‘continuative NRRCs’ (see example (2) above) the most. The intermediate levels are represented by non-presupposed RRCs, ‘a-restrictive’ IRCs, ‘Relevance’-indicating NRRCs and ‘subjective’ NRRCs, in that order.

second, non-presupposed RRCs (section 3.2), and third, what Cornish (2018) terms ‘a-restrictive’ relative clauses (ARRCs: section 3.3). Our purpose in presenting this three-way distinction, as will become apparent in section 5, is to show in this later section that, on the basis of their semantic-pragmatic properties, only the second and third subtypes are capable of being felicitously extraposed in context.

Although F&M (2014) restrict their discussion of extraposed RCs to subject-contained IRCs, it seems from our RCE reconstructions in section 5.1 of at least the non-presupposed RRCs and ARRCs, that non-subject NPs may also permit RCE quite naturally (cf. also Levy *et al.* 2012), and may also be analysable asthetic constructions: see sections 4.1 and 5.2. Let us look at each subtype in turn.

3.1. Presupposed restrictive relative clauses

Here, the relative clause serves to restrict the denotation of the head noun of its containing NP, co-determining its intended referent; but its content is also pragmatically presupposed by the speaker to be true of that NP’s referent. Two attested illustrations follow:

- (6) ... The dish *that really blows Fred away* is a simple one of fresh peas with *sobrassada*. (*Radio Times* 12–18.08.23 p. 56)
- (7) ... in fact the last interview *ø Dad did* was with Pru [actress Prunella Scales], (*Radio Times* 12–18.08.23 p. 10)

In (6), the writer is pragmatically presupposing that of all the newly adapted traditional Majorcan recipes at issue, it is just one that Fred (Sirieix) finds really extraordinary. The RRC serves to create a unique subcategory of Majorcan dishes. In (7), it is the subcategory of ‘interviews that Lucie Briar’s actor-father Richard did’ that is created by means of the RRC.

One test of the restrictive status consists in eliding the relative clause as a whole, and judging whether the referent of the NP that results is the same as the one evoked when the RC is present, or whether this manipulation results in a coherent discourse segment. If the NP’s referent is not the same, or if the RC’s elision does not achieve coherence, then that RC plays a role in identifying the referent at issue; but if it is or does, then such a role is irrelevant.

- (6) (a) The dish is a simple one of fresh peas with *sobrassada*.

Here, the interpretation of the definite residual NP *the dish* would seem to be restricted to an anaphoric one. However, the co-text upstream of this potential utterance in the source text has not evoked any one particular dish, so this formulation would result in incoherence in context. In any case, its potential interpretation would not include the particular Majorcan dish preferred by FS.

- (7) (a) ... in fact the last interview was with Pru.

Here, the residual definite NP *the last interview* does not necessarily imply that it was with Richard Briar, either as interviewee or interviewer. As with (6a), no mention of ‘interviews’ has preceded the occurrence of the relevant NP in extract (7), so the remnant NP in (7a) would be uninterpretable in context.

For the presupposed vs non-presupposed status of these RRCs, following Cornish (2018), we will exploit the ‘lie test’ procedure proposed by Erteschik-Shir (2007: 39, 164). This test is a heuristic for determining what is in focus in a given utterance. If it is felicitous to

contradict the content of a given segment in context, then this is indeed in focus, i.e. discursively at issue; but if it is not, then it is by default part of the presupposition of the utterance concerned. This test should be construed in relative, not absolute terms: it is the relative degree of ease with which the contradiction may be realised that reflects the degree of ‘foregroundedness’ of the information unit at issue. Let us apply it to both the matrix and the relative clauses presented earlier in this subsection.

- (6) (b) A to B: The dish *that really blows Fred away* is a simple one of fresh peas with *sobrassada*.
 B: #That’s not true, it doesn’t!
 (c) B: That’s not true, it isn’t!
- (7) (b) A to B: ... in fact the last interview *ø Dad did* was with Pru.
 B: #That’s not true, he didn’t!
 (c) B: That’s not true, it wasn’t!

In both cases, the contradiction of the matrix clauses is felicitous, since they evoke ‘at-issue’ information; whereas that of the relative clauses is not fully so, since the information concerned is background, not foreground in status, hence not in focus discourse-pragmatically – i.e. not at issue in this context. The negation test would see the scope of negation limited to what is in focus in each utterance, leaving the content of the relative clauses unaffected:

- (6) (d) The dish *that really blows Fred away* **isn’t** a simple one of fresh peas with *sobrassada*.
 (7) (d) ... in fact the last interview *ø Dad did* **wasn’t** with Pru.

3.2. Non-presupposed restrictive relative clauses

However, there are also IRCs that are pragmatically non-presupposed, though still restrictive in function. The profile of non-presupposed RRCs corresponds to the following outcomes of the two heuristics: the fact that a distinct, or otherwise contextually problematic, referent is evoked following elision of the RC; but felicitous contradiction of the RC. See (8) and (9) below:

- (8) The Olympic programme includes sports *that are better represented elsewhere*. (*Radio Times* 03–09.08.24, p. 18)

Deletion test (contextually negative):

- (8) (a) # The Olympic programme includes sports.

This sentence, although grammatical, would be singularly uninformative as well as redundant, hence discursively incoherent.

Lie test (positive):

- (8) (b) A, to B: The Olympic programme includes sports *that are better represented elsewhere*.
 B: That’s not true, they aren’t!
 (c) B: That’s not true, it doesn’t!

A further example is (9):

- (9) ... The programme explores the factors *that might have led to the disaster* ... (Radio Times 12–18.08.23 p. 68)

Here, the relative clause restricts the denotation of the head noun *factors*. This restrictive status is confirmed by the deletion test, since it would be most unclear what the residual NP *the factors* refers to. Thus this short programme note would become incoherent:

- (9) (a) #The programme explores the factors.

The lie test, however, shows that the content of the RC is to an extent focal, and not presupposed:

- (9) (b) A to B: The programme explores the factors *that might have led to the disaster*.
B: That's not true, they wouldn't have!
(c) B: That's not true, it doesn't!

3.3. 'A-restrictive' relative clauses

The third subtype of IRCs, ARRCs, is neutral as between the restrictive vs non-restrictive subtypes of relative clause. Denison & Hundt (2013: 142) call these '*aspective*' relatives. These are said to be

essential nonrestrictive clauses that bear a formal similarity to restrictive clauses in that they (a) occur in the same intonation contour as the matrix clause and (b) allow for relativisers other than *wh*-pronouns. ... [They] are thus not set-delimiting but add information to the noun in the matrix clause that is essential to the discourse.

Such relatives simply serve to elaborate the head noun's content, and indirectly the containing NP as a whole, with information that does not restrict the head noun's denotation.³ However, unlike canonical non-restrictive RCs (NRRCs), ARRCs do not constitute self-contained discourse units, carrying their own illocution as well as intonation contour, separate from those of the NP which contains them. Example (10) illustrates:

- (10) ... There's also a lighter pack *that uses less paper and no plastic*. (Advertisement, Radio Times 12–18.08.23, p. 14)

Here, the IRC simply elaborates on 'the lighter pack (of items constituting a full fundraising kit)' just evoked. It does not restrict the category of entity at issue at all, enabling the reader to identify the referent involved; rather, it simply contributes additional features of the particular 'pack' evoked by the advertiser.

In fact, it would be perfectly possible to delete the RC altogether, leaving an autonomous indefinite residual NP in its place:

- (10) (a) There's also a lighter pack.

The lie test confirms the ARRC characterisation:

³ Cf. also Depraetere (1996). Rydén (1974: 544–5) similarly distinguishes a 'descriptive or characterizational' IRC type. See also de Haan's (1989: 54) third category of relative clause uses in context, namely 'describing'.

- (10) (b) A to B: There's also a lighter pack *that uses less paper and no plastic*.
 B: That's not true, it doesn't!
 (c) B: That's not true, there isn't!

Thus the ARRC in (10) is neither presupposed, nor restrictive, but at the same time occurs NP-internally.

A further example is (11):

- (11) ... I would draw underground escape pods *that my family could go and live in ...* (*Radio Times* 12–18.08.23, p. 6)

Here, the relative clause is non-identifying, hence non-reference-restricting. It merely elaborates the 'underground escape pods' in question, in terms of their purpose according to the writer. Furthermore, the deletion test is positive:

- (11) (a) I would draw underground escape pods.

The negation test confirms the potentially focal status of the RC: 'You **didn't** draw underground escape pods that your family could go and live in': here, the negation may deny the drawing of a set of underground escape pods by the writer, or that their purpose was for her family to live in them. See also the positive outcome of the lie test in (11b.B) below.

- (11) (b) A, to B: I would draw underground escape pods *that my family could go and live in*.
 B: That's not true, they couldn't!
 (c) B: That's not true, you wouldn't!

So the relative clause in (11) is again an ARRC.⁴

It seems clear then that, although syntactically the three types of IRC are all attached to the head noun of their containing NP, semantically they differ in terms of closeness of connection: presupposed RRCs manifest the tightest connection, since they serve to carve out a subset of the class of entities denoted by the head noun, in virtue of their restrictive force, and also pragmatically presuppose the referent of the NP as a whole. Non-presupposed RRCs, however, simply restrict the entity denoted by the head noun to the category specified by the RC, without presupposing the existence of the entire NP's referent. ARRCs show the loosest connection, since they neither restrict their head noun's denotation, nor presuppose the existence of the broader NP's referent. Their sole function is to elaborate that head noun with information that is relevant to the discourse at issue downstream of the occurrence, but is not needed for the establishment of the NP's referent.

Indeed, each of these three subtypes of IRC makes a distinctive contribution to the discourse being created in a communicative event: presupposed RRCs by grounding the referent of the containing NP via their presupposed status, and also providing the addressee/reader with the further means of identifying it, via their restrictive character; and the non-presupposed RRCs and ARRCs, via their semantic-pragmatic detachment from the head noun, in the case of the latter subtype, and via the pragmatic aspect (i.e. their non-presupposed status) in that of the former one. These properties open up the possibility for these two subtypes of being extraposed in order for the speaker/writer to assert and thus highlight their content, as will be shown in [section 5](#).

The corpus of RCs which [table 1](#) reflects is a combination, first of manual extracts from two editions of the UK Radio and TV magazine *Radio Times* (12–18.08.23 and 03–09.08.24),

⁴ See Cornish (2018: 439–46) for further examples of each integrated RC subtype, as well as justification of the tests adopted for their statuses.

Table 1. Preferred relativisers for four subtypes of relative clause in English (raw data)

Preferred relativiser	Presupposed restrictive RC	Non-presupposed restrictive RC	A-restrictive RC	Non-restrictive RC
<i>that</i>	16.9% n. 10	42.4% n. 147	41.5% n. 71	–
<i>zero</i>	62.7% n. 37	17.6% n. 61	4.7% n. 8	–
<i>wh-expression</i>	20.3% n. 12	40% n. 139	53.8% n. 92	100% n. 81
Total %	100%	100%	100%	100%

including expository written examples and direct speech extracts from interviews. The remaining extracts were drawn from the ICE-GB Sample corpus, and include both written expository and informal spoken data.

A chi-square test was performed to assess the statistical significance of the relationship between the type of relative clause and the relativiser used, with the following results: $\chi^2(6df) = 207.67$, $p < .001$; Cramer's $V = 0.397$.

Thus there is indeed a statistically highly significant relationship between relative clause subtype and relativiser overall, there being a probability of less than 1 in 1,000 of obtaining the observed distribution under the null hypothesis of no relationship between the variables.⁵ Observation of the standardised residuals allows us to determine which cells in the table contribute significantly to the overall value of the statistic.⁶

The relevant conclusions are:

- *A-restrictive RCs* have a significantly higher frequency of occurring with *that* relativisers than would be expected on the basis of a null relationship between the variables.
- *Presupposed restrictive RCs* have a significantly higher frequency of *zero* relativisers than expected, whereas *a-restrictive RCs* have a significantly lower frequency than expected.
- *Presupposed restrictive RCs* have a significantly lower frequency of both *that* and *wh-relativisers* than expected.
- *Non-restrictive RCs* have a significantly higher frequency of occurring with *wh-relativisers* than expected.

These statistically significant correlations between subtypes of relativiser and subtypes of IRCs provide indirect evidence in favour of two of the claimed three-way distinctions amongst the latter (PRRCs and ARRCs).

Having established this three-way distinction amongst IRCs, on the basis of their semantic and pragmatic properties, we are in a position to examine F&M's (2014, 2017) multifactorial account of English IRC extraposition in sections 4.2 and 4.3.

⁵ The p value tells us only about the degree of confidence we can have in claiming significance; it says nothing about the size of the effect. This is measured by Cramer's V . A value between 0.2 and 0.5 indicates a small to medium effect.

⁶ Values of the standardised residual of 2 or greater, or of -2 or less, indicate significance at the .05 level. Positive values indicate that the frequency is higher than expected on the basis of no relationship between the two variables, while negative values mean that the frequency is lower than expected.

4. Francis & Michaelis' (2014, 2017) multifactorial account of relative clause extraposition in English

4.1. Some preliminary information-structure considerations: *theticity and topicality*

As a prerequisite to the presentation of F&M's (2014, 2017) conception of English relative clause extraposition (RCE: see sections 4.2 and 4.3, as well as examples (3) in section 2 and (12) below for initial illustrations), it is worth considering, as background to the discussion, certain aspects of Information Structure (IS): in particular *THETIC* and *TOPICAL* articulations. Information structure has to do with the following types of consideration: in general terms, it is concerned with the ways in which the form of a language expression is closely linked to contextual characteristics of the situation of utterance. That is, at every moment in the ongoing discourse co-construction process, decisions need to be made – such as what the speaker is attending to at any given moment, what s/he wishes the addressee(s) to focus specifically on, what s/he is assuming is already known information, and what can be considered as background to the ongoing discourse. A corresponding textual formulation can then be prepared in consequence.

In connection with the RCE construction type, Schultze-Berndt (2022) and Kirkwood (1977) characterise discontinuous NPs such as those occurring with an extraposed relative clause or prepositional phrase, at least on one possible interpretation, as *thetic* constructions.

Thetic constructions comprise 'all-new' (sentence focus) clauses, presentational constructions of various kinds, impersonal clauses and the so-called Locative Inversion one. All these share the following properties: there is no bipartite division, as there is in *TOPIC–COMMENT* articulations, between a topical NP and a predicator which ascribes some property to its referent, normally presupposed by the speaker/writer. Instead, the entire clause is fully integrated, i.e. 'monolithic'. At the discourse-pragmatic level,⁷ their function is to present an entity, a proposition or a state of affairs as a piece of new information for the discourse at issue. As such, they assume only a minimal discourse context; thus they may initiate a given discourse, or a discourse unit.⁸

They achieve this by resetting certain of the parameters of the new discourse unit: the spatio-temporal frame, the cognitive-referential domain, etc. Moreover, since they carry no presupposition, the proposition evoked as a whole, including the subject argument, falls within the scope of the assertion. The subject is normally prominent prosodically (i.e. pitch-accented): if it corresponds to an argument, it is treated as denoting an integral part of the situation evoked by the clause as a whole, and does not refer independently. It is for this reason that indefinite existential NPs may fulfil the subject function in such clauses: cf. Dobrovie-Sorin (1997). In addition, the verbal element involved is non-predicating: it is either a copula, or an existential verb of some kind, and as a result is de-accented, hence backgrounded.

⁷ Belligh & Crocco (2022: 1288–9) conclude, after an in-depth examination of eight types of Italian constructions expressing a *thetic* articulation, that there are no dedicated constructions in Italian that grammatically encode *theticity* (i.e. are in a one-to-one relationship): for the constructions at issue may also express *Topic–Comment*, *Argument Focus* and *Predicate Focus* functions. Hence it is claimed that *theticity* (along with other IS articulations) is not part of the grammar of a language, but rather a function of language use; that is, it operates at the level of discourse. Cf. also Matic' & Wedgwood (2013) regarding the category 'focus'.

⁸ Here is an attested example, complete with an extraposed infinitival purpose-clause modifier of the subject head noun:

- (i) A campaign has been launched to boost the incomes of tens of thousands of older people whose pension is paid by the Pension Protection Fund. [...] (Initial sentence of article 'Pensioners are losing out', *Radio Times* 12–18.08.23, p. 132)

See Schultze-Berndt (2022: 874–9) and Sasse (2006: 264–71) for a number of specific expression-level features of thetic constructions in a variety of the world’s languages. Schultze-Berndt (2022: 868) claims that ‘discontinuity [in the world’s languages] can serve as one of the strategies for marking “all-new” utterances of the type described as thetic’. This is the case, she argues, since by framing or enclosing the predicator – i.e. by keeping the residual NP in its earlier position in the clause, and placing the ‘extraposed’ unit [relative clause or PP] to the right of the predicator –, with both elements accented, it is not possible to interpret the subject NP as a topic, and the predicator as ascribing a property to it, as in Topic–Comment constructions. Thus the entire sentence is in focus, not just the extraposed constituent. An example from Schultze-Berndt (2022: 187) is given as (12):

- (12) At last, a doctor arrived who actually knew what to do. (BNC Written Books and Periodicals; CA8/95687980; 1987; Schultze-Berndt 2022: 187, ex. (16))

In (12), the reason for the doctor’s implied competence is asserted as being the case. The ‘contiguous’ realisation would also be possible as an alternative (*A doctor who actually knew what to do arrived*). In this case, the IS value would be that of a Topic–Comment articulation: the RC would identify the kind of doctor which is at issue, and it would be his/her arrival at the place assumed that is predicated of him/her. Note also that the emphatic construction in (12), more strongly than the alternative contiguous variant, would imply that other doctors present on the scene would *not* ‘know what to do’ in the circumstances at issue.⁹

4.2. Francis & Michaelis (2014): corpus survey and analysis

F&M (2014) consider, along with a variety of other authors whom they cite in this respect, that RCE is a means used by speakers and writers to highlight the content of the RC in question (see (3) and (12) above). They present the general consensus as being ‘that RCE is used to highlight new, contrastive or important information conveyed by the subject NP while backgrounding the information contained in the main clause predicate’ (F&M 2014: 71). That is, it is a type of presentational device, in information-structure terms. As the authors point out, the RCE construction is marked in relation to the non-RCE (canonical) one, and thus is relatively infrequent in usage compared with the latter.¹⁰

F&M (2014, 2017) do not frame their presentation primarily in terms of IS (Thetic vs Topic–Comment articulations), but rather assume its relevance, without always making it explicit or harnessing it to motivate their analyses. However, they do state (2017: 334) that ‘With respect to discourse function, RCE is preferred when the subject NP is focal and/or the VP is backgrounded’: cf. also the point made by F&M (2014: 71). By ‘focal’, F&M mean ‘new’, ‘important’ or ‘contrastive’ (2014: 71) – yet the example of the supposedly contrastive character of the (definite) subject NP presented in their example (3) (p. 71), does not in fact seem to be so.¹¹ See section 4.3.1 below for evidence of the relevance of these and other concepts for RCE in English, as well as section 5.2.

⁹ This analysis is my own.

¹⁰ My own corpus survey confirms this situation: there was only one ostensible instance of RCE ordering amongst the 400 originally IRC tokens extracted from *Radio Times*, and only 5 within the 177 tokens of IRCs from the ICE-GB Sample corpus. This is also reflected in the results of Levy *et al.*’s (2012) experiments 1 and 2, involving RCE vs the lack of it from subject and object NPs, respectively. Here, longer reading times were systematically observed for extraposed RCs vs non-extraposed ones for the stimuli used.

¹¹ (i) The guy just came in that I met at TRENO’s yesterday. (F&M 2014, ex. (3), p. 71)

In fact, this example would seem, rather, to be an instance of a Topic–Comment articulation. The extraposed RC is arguably not focal here, but identifying. As such it is grounded, since as the authors note, two ‘guys’ had been introduced prior to this utterance. In addition, the remnant NP is definite. So ‘the guy the speaker met at TRENO’s’ is

F&M approach the construction in essentially constituent-type terms, concentrating as they do on characterising the component elements that prototypically make it up, as well as on those comprised by the matrix clause. This is particularly in evidence in the presentations and analyses of the two experiments they conducted in F&M (2017) (see section 4.3).

The authors apply a range of heuristics, both grammatical and discursive, to a selection of utterances available within the ICE-GB corpus (*International Corpus of English – Great Britain*), in order to determine what factors favour or inhibit relative clause extraposition. The principal one is that of the length and complexity of the relative clause: the longer and more complex an RC is, particularly in relation to the matrix VP, the more likely it is that it will be extraposed. Of course, the longer a given RC is in relation to the matrix VP, the more it is likely to be part of the focus, in information-structure terms, since its content will be envisaged by the speaker/writer as relatively more important for the discourse at issue: however, see note 1 in section 2 above, as well as sections 4.3.1 and 5 below for a significant restriction on IRC extraposition: namely, an IRC's potentially presupposed status. Moreover, extraposing a long and complex sequence to the end of a discourse unit makes the whole unit easier to process, since the matrix predication can already be established as a base, without the interpreter having to wait until the otherwise heavy subject, or indeed non-subject, NP is processed: cf. F&M (2014: 72).

Further factors claimed to predispose an integrated RC to extraposition are the nature of the matrix verb (intransitive or passive, copula, denoting existence or disappearance, etc.), the indefinite status of the subject NP containing the IRC, and the degree of accessibility of the subject NP's referent and the main predicate, as derived from the preceding co-text.

The latter factor relates to F&M's (2014) notion of 'superset mention' for both subject NPs and main predicators. This assumes that the category of entity or state of affairs evoked by these two elements will have already been mentioned or have otherwise been at issue in the discourse upstream – or not, as the case may be. Its essential role is to increase the predictability of both the remnant subject NP and the predicate used in the matrix clause, and thus to ground their content to that extent.

4.3. Experimental verifications (comprehension and production)

F&M (2017) present and discuss two experiments designed to verify the conclusions drawn from the corpus survey in F&M (2014) (see section 4.2): a reading comprehension plus assessment one involving RCE and non-RCE examples, and an elicited spoken production one designed to assess how often subjects used sentences containing RCE, and what factors determine the choice of such formulations where this construction was chosen.

4.3.1. Experiment 1

The first experiment presented subjects with pairs of sentences with different constituent ordering (either RCE or non-RCE), and required them to indicate which sentence of each pair seemed more natural. All the stimuli used passive matrix predicates, in order to ensure both orderings were acceptable. The four types of possible combinations of the features at issue were as follows:

given information. The extraposition would then be motivated via the need for clarity. The authors in fact acknowledge (2014: 86) that 'a significant minority of RCE tokens (about 20%) [in their corpus] appeared to have topical subjects and focal predicates.' Since their example (3) is evidently a spoken one, it would be relevant to indicate the prosody with which it was realised, as evidence one way or the other.

- (13) *Indefinite subject NP, short VP, long RC*
- (a) Some research was conducted that refutes the existing theories with very clear and convincing new evidence. (RCE) (Ex. (2a), F&M 2017: 345)
 - (b) Some research that refutes the existing theories with very clear and convincing new evidence was conducted. (non-RCE) Ex. (2b), F&M 2017: 345)
- (14) *Definite subject NP, long VP, short RC*
- (a) The research has been conducted fairly recently that refutes the existing theories. (RCE) (Ex. (3a), F&M 2017: 345)
 - (b) The research that refutes the existing theories has been conducted fairly recently. (non-RCE) (Ex. (3b), F&M 2017: 345)

The predicted preferred choices amongst these four examples were claimed to be (13a), where the residual subject NP is indefinite, the preceding VP short and the subsequent RC long, over (13b), in which the long RC precedes the short VP; and (14b), where the subject NP is definite, and where the RC precedes the (same-length) VP, over (14a).

The results showed, confirming the authors' predictions, that RCE was the preferred choice (71.9%) when the VP was relatively light (i.e. short), but the RC relatively heavy (long), and also when the subject NP was indefinite rather than definite. Non-RCE ordering, however, was preferred (34.1%) when the VP was relatively heavy, but the RC light, and when the subject NP was definite.

Looking back at the above four examples, it seems clear that, considered in terms of their potential uses in context, there is in fact more to the issue than simply a choice of one or another structured alignment of designated constituent types in a clause realisation *per se*. For each of the two predicted preferred combinations (13a) and (14b) would in reality be used for different discourse purposes: in (13a), the indefinite subject NP *some research* would be pronounced with a high-fall intonation contour, signalling its discourse-new status, and the verb *conducted* would be unaccented. The extraposed RC, constituting the main point of this utterance, would be asserted: the verb *refutes* would be pronounced with a high-fall contour, and both the adjectives *clear* and *convincing* would be accented.

By contrast, in (14b) the head noun *research* within the definite subject NP *the research ... theories* would be pronounced with a low-rise contour, and the contiguous RC would be presented as a presupposed RRC (in our terms), thereby grounding the complex NP's referent. The VP occupying final position in the matrix clause would then constitute the essential point of the utterance, predicating the recent implementation of the research in question. So while (14b) is a clear instance of a Topic–Comment articulation, (13a) is an equally clear case of a Thetic ('all-new') one.

Finally, both (13b) and (14a) appear to manifest a degree of information-structure dissonance. In (13b), the essential point conveyed is that the research in question 'refutes very convincingly' the then existing theories', rather than the fact that it 'has been conducted'. But the former point is expressed in a subject-contained IRC, and yet it should normally have been highlighted via RCE.¹² The reverse applies to (14a): here the RCE should rather have been retained within the subject NP, which is definite. As such, it would function to complete the grounding of the utterance as a whole, by restricting the NP's referent to the type of research at issue here. Hence this referent would be topical in status, so there would be no motivation for extraposing the IRC to create a thetic articulation, however long the IRC might be.

¹² See also Huddleston & Pullum *et al.* (2002: 1066) on the key factor whereby an IRC's greater informational load in relation to the matrix predication is the crucial determinant of the choice of its extraposed ('postposed' in their term) ordering.

It would seem, therefore, that it was the subconscious realisation of these differing information-structure statuses by the subjects tested that determined the results obtained, rather than the various different combinations of the three expression-level features at issue *per se*. The latter are simply the perceptible reflection of the more basic IS articulations involved.

4.3.2. Experiment 2

This elicited spoken production experiment sought to obtain RCE and non-RCE sentences varying according to the length and definiteness of the subject NP. The participants tested had to construct a sentence made up of three phrases presented in random order: a potential subject NP, a VP and a relative clause, and then to pronounce the sentence created. The results showed, as predicted, that participants used RCE more frequently when the subject NP was indefinite rather than definite, by a ratio of 50 percent indefinite with RCE constructions to 24 percent definite. Crucially, whenever the subject NP was definite but where the RCE option was chosen, subjects were slower to initiate their oral responses than when the subject NP was indefinite. The sentences constructed with canonical, host-attached RCs, by contrast, showed no such difference between the use of definite and indefinite subject NPs.

The authors explain this contrast in terms of the production and comprehension strategy of placing shorter and simpler constituents early in a clause realisation, and longer and more complex ones later, coupled with one favouring speedy retrieval from memory of ‘the most frequently used subtypes of a construction’ (F&M 2017: 333). However, they also note (2017: 332) that the discontinuous NP-cum-RCE construction is marked in relation to the ‘contiguous’ RC one: it is not only relatively infrequent in usage, but more complex compared to its non-RCE counterpart. So it is hard to see that, although ‘marked’ and more complex as a construction, it is nonetheless supposedly ‘easier’ to retrieve from long-term memory.¹³

And yet, according to the authors, this contrast reflects a sensitivity on the part of speakers solely to the ‘definiteness’ factor relating to the use of RCE. The presence of an indefinite subject NP is viewed as a ‘cue’ that speakers of English are said to recognise as a potential trigger for RCE ordering.

These two experiments were based on written sentences (experiment 1) or sentence fragments (experiment 2) as the stimuli. No supporting context was given for the participants to choose whether one sentence variant or combination of the phrases provided would correspond to one or another discursive intention on the part of the writer (experiment 1) or potential speaker (experiment 2). The appeal to context was therefore limited (intentionally so) to the conventional sense of given language forms: the determiners *a* or *some*, on the one hand, and *the* on the other, as well as the structural difference between canonical RC-containing NPs and discontinuous ones with RCE.¹⁴

We are now in a position to compare as well as integrate the two approaches outlined in section 3 (the communicative-dynamic one) and section 4 (the multifactorial one), the purpose of section 5.

¹³ This claim is in fact contradicted by the results of the extraposed RC stimuli in Levy *et al.*’s (2012) reading-time experiments, which involved systematically longer reading times than did the non-RCE stimuli (see note 11 in section 4.2 above).

¹⁴ In fact, the prosody with which the subjects pronounced the sentences they constructed in the second experiment could be put to use most effectively as a possible contextualising device, reflecting the intended information structure involved.

5. RCE: comparing the communicative-dynamic (discursive) and the multifactorial accounts

5.1. Observation in vivo and creation in vitro of RCE (re-)structuring

Let us examine first of all the corpus examples presented by F&M (2014) in the light of the three-way distinction in terms of communicative dynamism (see note 2 in section 3) presented in section 3. Their example (1a) is reproduced as (15):

- (15) Further research has been conducted on this that indicates this criticism may not be just. (F&M 2014: 70, ex. (1a))

The extraposed RC here is clearly an ARRC. The subject is indefinite (conveying discourse-new information), and the extraposed RC may be deleted, leaving the residual NP to function autonomously, as it already does in (15). The RC serves to convey the upshot of the ‘further research’ at issue. It is clearly not pragmatically presupposed: see the possible felicitous contradiction ‘That’s not true, it doesn’t!’.

- (16) (a) Certain conditions existed which cannot be applied to all other countries at all times. (F&M 2014: 72, ex. (5a))

As in (15), the RC here is an ARRC. The non-RCE version of this example which the authors present as their (5b) seems somewhat infelicitous:

- (16) (b) Certain conditions which cannot be applied to all other countries at all times existed.

This is parallel, structurally, to the authors’ illustrative example (13b) in section 4.3.1 above. Its relative infelicity is due, as the authors point out, to the extreme discrepancy between the length and complexity of the RC and that of the main clause predicate (a single lexical item), reflecting the hiatus between the higher message-level importance of the IRC over what is predicated of its referent in the main clause; and as a result, adding to the processing burden this variant would place on the reader. Hence extraposing the IRC would be virtually mandatory here.

As we have seen, these illustrations presented by F&M as examples corresponding to their prototype of the RCE construction are in reality precisely what we are calling ARRCs. For indeed, the ARRC subtype of IRCs is the one best suited to extraposition, since it manifests the loosest connection to its head N of all three subtypes (see section 3.3). Hence it is no coincidence that the authors should be selecting examples of this subtype for their demonstration – though they do not in fact recognise it as such.

Let us attempt now to apply RCE to the three subtypes of IRCs characterised in section 3. In order to assess the results, these should be compared in terms of degree of naturalness with the contiguous variant, as in the original examples presented in section 3, as well as with the various other possibly extraposed RC subtypes involved. First, then, the presupposed RRCs.

- (17) The dish is a simple one of fresh peas with *sobrassada that really blows Fred away*. (Extraposed variant of (6))

- (18) # ... in fact the last interview was with Pru *ø Dad did*. (Extraposed version of (7))

The RCE version of (7) ((18)) is clearly infelicitous, as compared to its contiguous counterpart.

As for (17), if it is acceptable, then its interpretation would involve major differences with regard to the originally integrated presupposed RRC in (6). First, the residual subject NP is

definite, signalling by default an anaphoric interpretation. However, no particular dish has been evoked in the discourse upstream in the short containing text for it to maintain. Hence this variant would immediately result, in context, in incoherence. In fact, it is the RC itself in (6) that has made the NP as a whole definite (*viz. The dish that really blows Fred away...*),¹⁵ since its interpretation would seem to correspond to a ‘uniquely identifiable’ reference; but this would be impossible without the originally contiguous RC. Furthermore, the RC in (6) is both presupposing and restrictive, as we saw in section 3.1. But the ‘RCE’ version in (17) would have neither of these properties. In this ‘extraposed’ variant, the RC would have to be a-restrictive in function, since it simply adds an extra, incidental, proposition to the NP as a whole: the lie-test and scope of negation heuristic would both be positive here. So extraposing, if this is the correct analysis, the initially integrated RC would negate its originally specific functional properties. Thus, even if the utterance as a whole is considered felicitous, it would not constitute a counterexample to what I am claiming – rather the reverse, in fact. Compare the extraposed versions of the relevant IRC ones in sections 3.2 and 3.3 in (19)–(22) below, which involve non-presupposed RRCs and ARRCs: all are perfect as RCEs, and the RCs in question all maintain their *in situ* properties, unlike the presupposed RRC ‘extraposed’ in (17).

Now to the second subtype of IRCs presented in section 3.2, namely non-presupposed RRCs.

- (19) The Olympic programme includes sports, of a varied range of types, *that are better represented elsewhere*. (Ex. (8) modified)¹⁶

The result is felicitous.

- (20) The programme explores the factors – there were many – *that might have led to the disaster*. (Ex. (9) modified)

Again, the RCE version of example (9) is felicitous.

Finally, the ARRC examples from section 3.3:

- (21) There’s also a lighter pack, very convenient for users, *that uses less paper and no plastic*. (Ex. (10) modified)

The result is perfect.

- (22) I would draw underground escape pods – I’m sure you know the sort I mean – *that my family could go and live in*. (Ex. (11) modified)

Yet again, the result is perfectly natural.

¹⁵ This factor no doubt explains why removing the IRC from the subject NP causes its definite status to require an alternative motivation (e.g. a potentially anaphoric one).

¹⁶ Unlike (17) and (18), where the RCs have been extraposed from their subject NP in examples (6) and (7), in (19)–(22) below, this is impossible, since the RCs concerned were all initially object- or predicate nominal-contained ones: as such, they are already placed within the VP of the original examples (8)–(11), hence already occurring to the right of the main verb. Thus the only way to detach them from their respective head nouns was to insert a parenthetical element between these two constituents. The aim was to better motivate the extrapositions in question, rendering the result more natural, and also to ensure that the head noun and the RC are no longer contiguous. Francis (2022) proceeds likewise, both in her own examples and in citing those of other authors. As already noted more generally regarding RCE constructions, the longer the latter (particularly as compared to their matrix VP), the easier it is to interpret them as focal rather than strictly presupposed. This is predominantly the case for the utterance-final position, but less so for an utterance-initial or utterance-medial one, as in (6) (see (17)) and (7) (cf. (18)).

5.2. Discussion: invoking an information-structure framework

Clearly, then, the non-presupposed RRC examples (19) and (20), and both the ARRC ones (21) and (22) are perfect with RCE ordering.

Now amongst these felicitous RCE variants of the examples from section 3, one contains a definite object NP and a transitive verb ((20)), two an indefinite object NP and a transitive verb ((19) and (22)) and one a copular verb with an indefinite predicate-nominal NP ((21)). Apart from the three examples with indefinite (non-subject) NPs and the universally longer RCEs than their respective VPs, these extracts do not share all the characteristic features of potential utterances permitting RCE from within a containing NP, according to F&M (2014) – and yet these examples seem acceptable nonetheless. Let us look more closely at this subset.

First, (19) (examples (19)–(22) repeated below):

- (19) The Olympic programme includes sports, of a varied range of types, *that are better represented elsewhere.*

Here, both the fact that the head noun is indefinite as well as plural (*sports*), and the choice of the ‘light’ matrix verb *includes* (replaceable in context by *has*) serve to displace the central burden of the message at issue onto the extraposed RC, thereby motivating this potential discourse-pragmatic realisation to that extent. The non-presupposed status of the RRC further increases this property. All these features combine to make the RC in question a good candidate for extraposition.

As for (20),

- (20) The programme explores the factors – there were many – *that might have led to the disaster.*

the residual NP *the factors*, though definite, is not topical, since it is an integral part of the rheme or focus of the corresponding utterance.¹⁷ In addition, the matrix verb *explores* is analysable as quasi-existential in character: viz. ‘bring under the analytic spotlight’. Moreover, the RC is ‘heavier’ than the matrix VP. Again, the RC represents the essential point of the sentence as a whole, increasing its propensity for extraposition.

In (21) (‘There’s also a lighter pack, very convenient for users, *that uses less paper and no plastic*’), the residual NP *a lighter pack*, being indefinite and part of the rheme, is non-topical. Furthermore, its ARRC status already gives the RC the highest degree of independence in relation to its head noun when in a contiguous position, as compared with the other two subtypes of integrated RCs. The preponderance of these factors conspires to motivate its candidacy for extraposition in this instance.¹⁸

Finally, in (22),

- (22) I would draw underground escape pods – I’m sure you know the sort I mean – *that my family could go and live in.*

although the residual NP *underground escape pods* is indefinite, the main verb is transitive. Nonetheless, it may also be construed as existential, in that its sense in context corresponds

¹⁷ With the discontinuous variant with RCE, this remnant NP could in no way be construed as a potential topic: # ‘About the factors, the programme explores them.’ For the ‘factors’ in question have not yet been introduced (this is precisely the function of the RCE here). Dalrymple & Nikolaeva (2011: 165) claim that focal and topical objects are equally appropriate cross-linguistically.

¹⁸ Indeed, the original non-RCE example (10) was already presentational in structure (*there be* + expanded predicate nominal).

to the coming into being of ‘drawings of underground escape pods’. In addition, the indefinite residual NP fulfils the direct object function, which, especially if it is indefinite, as here, marks its content as being part of the rheme of the utterance, but not the topic: this referent is presented as constituting new information, and there is no ‘aboutness’ relation motivating a potential topic function: # ‘About underground escape pods, I would draw them’. The RC is also heavier than the matrix VP; moreover, its ARRC status, as in (20), increases the likelihood of its being extraposed.

Finally, all the matrix verbs in (19)–(22), respectively *include*, *explore*, *be* and *draw*, can be viewed as corresponding to Schultze-Berndt’s (2022: 882) claim that ‘any verb that can be interpreted as introducing an entity into a discourse world is permitted [in RCE constructions]’.

More generally, why is it that only non-presupposed RRCs and ARRCs amongst the three subtypes of IRCs may felicitously permit RCE? Well, since RCE is a type of presentational construction, it would be contradictory to attempt to extrapose structures which are pragmatically presupposed of the referent of the NP whose head noun’s denotation they restrict (i.e. presupposed RRCs).¹⁹ A given constituent cannot at the same time be conceived felicitously as backgrounded *and* foregrounded, as is shown in section 3.1 by the relatively infelicitous attempts to contradict the proposition these RCs express.

As we saw in section 4.2, the relative VP-RC weight factor can be explained analogously: an IRC’s greater ‘weight’ in terms of relative numbers of words in relation to the main predicator reflects the fact that the former will thereby assume a greater information load than the latter. This factor may justify its surface ordering further towards the end of the sentence as a whole, since it would thereby assume a more clearly rhematic or focal status. However, if the RC, even if it is long in relation to the VP, is a presupposed RRC, then it will not be susceptible to extraposition: see note 1 in section 2 in relation to clausal subjects, as well as sections 4.3.1 and 5.1.

So it is discourse-related factors and principles that hold greater sway than the selection *per se* of specific sentence-internal factors highlighted by F&M (2014, 2017): indefiniteness of the subject NP, types of matrix verbs with ‘presentational’ semantics, relative grammatical weights of the RC vs the VP – but also, the degree of discourse accessibility of the referent of the subject NP and of the main predicator’s contribution. Although all these expression-level factors are correctly specified and relevant, what they all have in common, clearly, is their overall predisposition to realising a thetic IS articulation in context.

One crucial factor in the marked occurrence of RCE is the generalisation to the effect that the predicative function otherwise associated with non-copular main verbs of clauses containing an extraposed RC is lacking. This is generally the case with all thetic or other

¹⁹ Cf. also Goldberg’s (2013) more general restriction on the ‘extraction’ of constituents, applying only to those which are *not* backgrounded. Adopting an information–structure approach, Goldberg re-examines a number of the supposedly ungrammatical instances of ‘extraction’ of constituents in terms of long-distance dependency relations from the formal-syntactic literature – in particular, so-called ‘island’ constraints. An example involving a *wh*-constituent is **Who did she see the report that was about* (Goldberg 2013: 222, table 10.1). This is standardly claimed to be ungrammatical due to the ‘Complex Noun Phrase constraint’, inhibiting extractions from within such a constituent.

However, the author argues, analogously to the present article, that it is due to the fact that the relative clause embedded in the NP *the report that was about X* is presupposed, hence backgrounded, that the extraction in question is inhibited. Analogously to what we are arguing in the case of the difficulty of extraposing a presupposed RRC, the contradiction involved here corresponds to the ‘extraction’ of a would-be focal, foregrounded element (the interrogative pronoun *who*) from within a presupposed, thus backgrounded, segment of an utterance. That is, it is for information-structural, utterance-level reasons, rather than strictly grammatical ones, that this and other such constructions are impossible.

presentational constructions,²⁰ such as the one known as Locative Inversion. Locative inversion has several of the key features of the RCE construction type: bleached or copular main verb (i.e. where the latter is non-predicating), and the postposing of the entire subject NP to the end of the clause, rather than just the RC it contains, in the case of complex NPs with an IRC: this displacement thereby highlights its content, as in the case of an RCE. Its specific characteristic is to be prefixed by a locative or temporal phrase, marking the grounding of the utterance in the preceding discourse context. Cf. Birner & Ward's (1998: 156–94) IS-oriented constraint to the effect that, for the inversion to be felicitous, the postverbal constituent (here the subject NP) must not represent information which is more 'discourse-old' than that conveyed by the preverbal one. Like canonical RCE constituents involving extraposed ARRCs, Locative inversion constructions serve to set the scene for what is to come in the succeeding discourse, as illustrated in examples (i) in note 21 and (23) below.

In (23), the verb is locative and existential in value, and, as with the use of any non-locative or existential verb in athetic IS articulation, is semantically 'bleached', hence backgrounded in relation to the more prominent postposed subject NP. This generalisation would cover the set of possible verbs occurring within this construction as specified by F&M (2014, 2017) and Francis (2022): locative, passive or intransitive verbs, copulas, and verbs denoting the existence, appearance or disappearance of some entity or state of affairs. In main-clause presentational constructions, the subject NP is often postposed to a position to the right of the verb (in SVO languages), where it can receive a nuclear accent, as in (23):²¹

- (23) On the veranda, upon a wooden table with four legs, stood an enormous brass telescope. With it, you spied out the steamers passing ... (Extract from Jean Rhys, 'Mixing cocktails' (1927), *The Collected Short Stories*, London: Penguin, 2017, p. 34)

Here, the verb *stood* could be replaced quite naturally in this construction by a copula (*was*, or *there was*). However, it could not be modified naturally by an adverb (e.g. *proudly*). Such an adverb would, by contrast, be perfectly acceptable as a modifier of this verb when used predicatively, as in the canonical construction *An enormous brass telescope stood proudly on the veranda upon a wooden table with four legs*. Evidently, replacing the verb here by a copula would be impossible.

As with the RCE construction, an entity focally introduced into a discourse via a preceding Locative Inversion one may be reformatted anaphorically as topic in a following Topic–Comment articulation, as is the case in the attested textual example (23) above.

From an IS perspective, as with presentational constructions generally, the choice of athetic utterance involving RCE assumes no particular prior context; thus, as we saw in section 4.1, it may mark the start of a new discourse unit, thereby adding a prominent discourse-new item of information to the addressee's/reader's pragmatic knowledge base: see example (i) in note 8 in section 4.1 above. This factor evidently motivates the indefinite subject NP status claimed by F&M as the preferential choice for RCE occurrences, and the

²⁰ See example (23) below as well as Hannay (1991) for a typology of presentational as well as topical constructions in English within a Functional Grammar context.

²¹ Here is another attested example of the use of this construction:

- (i) On a table was spread a white cloth and there were blue cups and saucers and a plate of gingerbread cut into slices and thickly buttered ... (Jean Rhys, 'Tea with an artist' (1927), *The Collected Short Stories*, London: Penguin, 2017, p. 29)

The locative inversion construction here, together with the *there*-presentational one, serve to set the scene for the tea about to be taken with the artist in question.

downgrading or backgrounding of the matrix predicator – since all aspects of the containing sentence should be subservient to, and not compete with, the foregrounding of the content of the extraposed RC. BBC radio news bulletins, in particular, use this format regularly to present new items of information, both events and individuals.²² Indeed, RCE usage can be characterised as a hallmark of this particular genre. An attested example, albeit from a different source, is given in (24), here illustrating PP extraposition from within a subject NP – although the PP does contain a relatively heavy RC:

- (24) A hunt's begun for two gunmen who burst into a pub in South London and opened fire on drinkers killing two (ICE-GB: S2B-016 #003:1:A; spoken extract)

This is a type of thetic construction, since the predicator 'has begun' was no doubt de-accented, hence non-prominent and non-predicating, simply serving to evoke the start, or 'coming into being', of this particular event; whereas the emphasis achieved via PP extraposition is laid upon the search for the two gunmen being introduced into the world of discourse in terms of the situation involved. I would assume that the head noun *hunt* within the subject NP would be pitch-accented with a high-fall intonation contour,²³ and that the expressions *burst* and *pub* would also be accented.

6. Conclusions

F&M (2014: 85) claim that the various formal and discourse-pragmatic factors that favour RCE which they put forward as predictors of this ordering, taken together, serve to form an RCE prototype. These factors are, as we have seen, 'indefinite discourse-new subject', 'presentative verb', and '[proportionately] heavy relative clause'.

While the authors' claimed prototypical expression-level factors are certainly valid in their own right, they may be seen as choices made by a speaker or writer in order to *express* his or her discourse intentions in language use. As such, they should be viewed rather as symptoms, not 'causes' of the essentially discourse-level phenomenon at issue – the authors indeed characterise them as 'soft constraints', rather than as strict grammatical ones. What is lacking is a consideration of the utterance-level conditions underlying the use of expressions containing RCE ordering: a speaker/writer, an addressee, a discourse intention as well as occasion on the part of the former with regard to the latter, a context, and so on. The assumption is, broadly speaking, that simply taking account of the relevant intra-sentential constituents on their own is sufficient to motivate an RCE vs a non-RCE usage; and this has been shown not always to be the case.²⁴

Concerning the possible occurrence of definite NP [non]subjects as well as transitive verbs in RCE instances, Schultze-Berndt (2022) argues that, in the former case, they will not be marked as potential topics (e.g. via accenting or expression as non-subjects: see *the factors* in (20)), and in the latter one, will be 'de-predicativised' via de-accenting.

²² This utterance type is illustrated in (i) below, where there is no disjuncture between *street* and *which*, and no change of key at that point:

(i) A bomb exploded yesterday in Wykeham high street which caused considerable damage.

²³ Cf. Cruttenden (2006: 311) on the accenting of discourse-prominent referents.

²⁴ One problem for this approach is the fact, noted by the authors themselves (2014: 86) (see also [note 11](#) in [section 4.2](#)), that about 20 per cent of the tokens of RCE they extracted from their corpus 'had topical subjects and focal predicates'. For the spoken extracts among these, it should be possible to characterise the prosody with which they were pronounced, thereby providing solid evidence for one IS articulation or the other. For clearly, in these cases, the particular configuration of relevant constituents *per se* would not in itself reveal an intended 'thetic' interpretation of the RCE examples at issue – rather, a 'topic-comment' one.

Our corpus-based study has shown (indirectly) that RCE may potentially occur with other than solely subject-contained RCs: cf. also Levy *et al.* (2012); so F&M's (2014) restriction of RCE only to this subtype does not in fact generalise. It has also shown that it is not only ARRCs that can undergo RCE to achieve a potentiallythetic utterance (recall the point made in section 5.1 that the RCE examples the authors present are in fact instances of this subtype), but also non-presupposed RRCs. These subtypes have in common that they potentially present new, non-grounded information, unlike presupposed RRCs, which are relatively infelicitous or otherwise altered when extraposed.

We have also seen that when sentences with RCE are conceived asthetic utterances, there is a potentially wider range of constituent types available for their expression than the three highlighted by the authors – though of course this is perfectly compatible with the authors' characterisation of their selection as a 'prototype'; nonetheless, the modified RCE examples (19)–(22) in section 5.1 all seem equally as acceptable as any 'prototypical' one. These 'non-prototypical' constituent types are, for example, non-topical, accented remnant host NPs, whether subject or non-subject in function, definite or indefinite, a potentially bleached 'de-predicativised' verb, whether transitive or intransitive, and a relatively heavy non-presupposed RC, thereby warranting focal treatment when extraposed.²⁵ It is also important to set aside NRRCs, which cannot be considered as having been 'extraposed' from an NP-internal position: a potential misanalysis is particularly likely whenever the relativiser introducing a post-verbal RC is a *wh*-expression, rather than *that* or zero. The distinction betweenthetic and topic–comment articulations, when considering possible NP-*cum*-RCE constructions, is also highly relevant.²⁶

On the basis of the data and analyses in sections 3 and 5, then, it would seem that it is indeed the full discourse motivation lying behind these choices that is paramount, and that the latter flow from the former. Interestingly, Francis (2022: 151) (cf. also Levy *et al.* 2012: 30), actually mentions this determining factor, in suggesting that

another option [for characterising RCE in English] would be to say that these constraints are not really specific to the grammar of RCE at all, but perhaps can be understood in terms of general principles of information flow in discourse.

This is, of course, exactly the approach adopted here towards relative clause extraposition in English.

Acknowledgements. I would like to thank for their helpful comments on earlier versions of this article Sarah Blackwell, Christopher Butler, Anita Fetzer, Matthias Klumm, Jan Rijkhoff, Eva Schultze-Berndt, my handling editor Bernd Kortmann and the two anonymous reviewers for the journal.

References

- Belligh, Thomas & Claudia Crocco. 2022. Theticity and sentence-focus in Italian: Grammatically encoded categories or categories of language use? *Linguistics* 60(4), 1241–93.
- Birner, Betty J. & Gregory Ward. 1998. *Information status and noncanonical word order in English*. Amsterdam and Philadelphia: John Benjamins.
- Chierchia, Gennaro & Sally McConnell-Ginet. 1990. *Meaning and grammar*. Cambridge, MA: MIT Press.

²⁵ The latter feature, of course, apart from its specifically non-presupposed character, was indeed retained by the authors, as we have seen. However, they did not motivate the choice of RCE from an IS perspective. As we saw above (cf. section 4.3.1), however proportionately longer a given RC is, if it is a presupposed RRC, then its contribution to the grounding of the containing utterance will override this factor, causing it to remain in contiguity with the head noun inside the definite NP (non-)subject.

²⁶ Interestingly, according to Komen (2014: 122–3), in written texts in Chechen (a Caucasian SOV language), extraposed restrictive relative clauses systematically relate to a head noun in focus position (the immediately preverbal position). Non-restrictive relatives, on the other hand, are not so restricted, and may occur in a variety of positions in relation to their modified NP or clause.

- Cornish, Francis. 2018. Revisiting the system of English relative clauses: Structure, semantics, discourse functionality. *English Language and Linguistics* 22(3), 431–56. doi:10.1017/S136067431700003X.
- Cruttenden, Alan. 2006. The de-accenting of given information: A cognitive universal? In Giuliano Bernini & Marcia L. Schwarz (eds.), *Pragmatic organization of discourse in the languages of Europe*, 311–55. Berlin and New York: Mouton de Gruyter.
- Dalrymple, Mary & Irina Nikolaeva, I. 2011. *Objects and information structure*. Cambridge: Cambridge University Press.
- De Haan, Pieter. 1989. *Postmodifying clauses in the English noun phrase: A corpus-based study*. Amsterdam: Rodopi.
- Denison, David & Marianne Hundt. 2013. Defining relatives. *Journal of English Linguistics* 41(2), 135–67.
- Depraetere, Ilse. 1996. Foregrounding in English relative clauses. *Linguistics* 34, 699–731.
- Dik, Simon C. 1997. *The theory of Functional Grammar*, part II: *Complex and derived constructions*. Berlin: Mouton de Gruyter.
- Dobrovie-Sorin, Carmen. 1997. Classes de prédicats, distribution des indéfinis et la distinction thétiq–catégorique. *Le Gré des Langues* 12, 58–97.
- Erteschik-Shir, Nomi. 2007. *Information structure: The syntax–discourse interface*. Oxford and New York: Oxford University Press.
- Francis, Elaine J. 2022. Relative clause extraposition and PP extraposition in English and German. In *Gradient acceptability and linguistic theory*, 126–56. Oxford: Oxford University Press.
- Francis, Elaine J. & Laura A. Michaelis. 2014. Why move? How weight and discourse factors combine to predict relative clause extraposition in English. In Brian MacWhinney, Andrej Malchukov & Edith Moravcsik (eds.), *Competing motivations in grammar & usage*, 70–87. Oxford: Oxford University Press.
- Francis, Elaine J. & Laura A. Michaelis. A. 2017. When relative clause extraposition is the right choice, it’s easier. *Language and Cognition* 9(2), 332–70.
- Goldberg, Adele E. 2013. Backgrounded constituents cannot be ‘extracted’. In Jon Sprouse & Norbert Hornstein (eds.), *Experimental syntax and island effects*, 221–38. Cambridge: Cambridge University Press.
- Hannay, Mike. 1991. Pragmatic function assignment and word order variation in a functional grammar of English. *Journal of Pragmatics* 16, 131–55.
- Huddleston, Rodney & Geoffrey K. Pullum et al. 2002. *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.
- Kirkwood, Henry W. 1977. Discontinuous noun phrases in existential sentences in English and German. *Journal of Linguistics* 13, 53–66.
- Komen, Erwin R. 2014. Chechen extraposition as an information ordering device. In Rik van Gijn, Jeremy Hammond, Dejan Matić, Saskia van Putten & Ana Vilacy Galucio (eds.), *Information structure and reference tracking in complex sentences*, 99–126. Amsterdam and Philadelphia: John Benjamins.
- Levy, Roger, Evelina Fedorenko, Mara Breen & Edward Gibson. 2012. The processing of extraposed structures in English. *Cognition* 122, 12–36.
- Loock, Rudy. 2010. *Appositive relative clauses in English*. Amsterdam and Philadelphia: John Benjamins.
- Matić, Dejan & Daniel Wedgwood. 2013. The meanings of focus: The significance of an interpretation-based category in cross-linguistic analysis. *Journal of Linguistics* 49, 127–63.
- Miller, P. 2001. Discourse constraints on (non)extraposition from subject in English. *Linguistics* 39(4), 683–701.
- Rydén, Mats. 1974. On notional relations in the relative clause complex. *English Studies* 55, 542–5.
- Sasse, Hans-Jürgen. 2006. Theticity. In Giuliano Bernini & Marcia L. Schwarz (eds.), *Pragmatic organization of discourse in the languages of Europe*, 255–308. Berlin and New York: Mouton de Gruyter.
- Schultze-Berndt, Eva. 2022. When subjects frame the clause: Discontinuous noun phrases as an iconic strategy for marking thetic constructions. *Linguistics* 60(3), 865–98.