

<https://doi.org/10.22363/2687-0088-41329>


EDN: ITPFNA

Research article / Научная статья

The cleft construction: A formal definition

Igor MEL'ČUK  

Observatoire de linguistique Sens-Texte, Université de Montréal, Montreal, Canada

 igor.melcuk@umontreal.ca

Abstract

The paper addresses the task of systematizing linguistic notions and corresponding terminology: it presents a rigorous definition of the notion of cleft construction. The study is carried out in the theoretical framework of the Meaning-Text approach. The substantive and formal requirements on a rigorous linguistic definition are formulated. The cleft construction in English is described as a basis for the definition: the semantic [SemR], the deep-syntactic [DSyntR] and the surface-syntactic [SSyntR] representations of three English cleft sentences are given, as well as five formal DSyntR \Leftrightarrow SSyntR rules for the expression of a focalized Rheme by the cleft construction. The cleft construction is defined as a particular type of linguistic sign; it is a grammatical (surface-syntactic) idiom, headed by a lexeme of the copula verb 'be' with fairly complex syntactics. An overview of cleft constructions in several languages structurally different from English—French, Spanish, German, Irish, Kinyarwanda, and Mandarin Chinese—follows. Finally, pseudo-cleft sentences are considered; in contrast to cleft sentences, they are special only from a semantic, but not from a syntactic viewpoint (there is no term **pseudo-cleft construction*): they present a particular organization of the starting meaning. The results of the paper: 1) It proposes a formal notion of cleft construction, which allows the researchers to distinguish—in various languages—syntactic phenomena that serve the same informational purpose (namely, the expression of a focalized Rheme or Theme), but are structurally different; in this way the paper contributes to General syntax. 2) It provides a sketch of a formal description of the cleft construction in English, thus contributing to English studies.

Keywords: *formal linguistic notions, syntax, cleft construction, pseudo-cleft sentences, English, Kinyarwanda, Mandarin Chinese*


For citation:

Mel'čuk, Igor. 2025. The cleft construction: A formal definition. *Russian Journal of Linguistics* 29 (2). 218–249. <https://doi.org/10.22363/2687-0088-41329>

Расщепленная конструкция: формальное определение

Игорь МЕЛЬЧУК  

*Лингвистический центр Смысл-Текст, Монреальский университет,
Монреаль, Канада*

igor.melcuk@umontreal.ca

Аннотация

В статье решается задача систематизации лингвистических понятий и соответствующей терминологии: дается строгое определение понятия «расщепленная конструкция». Исследование проводится в рамках теоретического подхода «Смысл–Текст». Формулируются содержательные и формальные требования к строгому лингвистическому определению. В качестве базы определения описывается расщепленная конструкция в английском языке: даются семантические [SemR], глубинно-синтаксические [DSyntR] и поверхностно-синтаксические [SSyntR] представления трех английских расщепленных предложений, а также пять формальных $DSyntR \Leftrightarrow SSyntR$ правил для выражения фокализованной ремы конструкцией с расщеплением. Данная конструкция определяется как особый тип языкового знака; это грамматическая (поверхностно-синтаксическая) идиома, в которой синтаксической вершиной является глагол-связка ‘быть’ с достаточно сложной синтактикой. Представлен обзор расщепленных конструкций в нескольких языках, структурно отличных от английского, – во французском, испанском, немецком, ирландском, киньяруанда и современном китайском. Затем рассматриваются псевдо-расщепленные предложения, которые, в отличие от расщепленных, являются особенными только с семантической, но не с синтаксической точки зрения (термин **псевдо-расщепленная конструкция* не используется): в них представлена особая организация исходного значения. В статье предложены: 1) формальное определение расщепленной конструкции, которое позволяет исследователям различать в разных языках синтаксические явления, служащие одной и той же информационной цели (а именно выражению фокализованной ремы или темы), но являющиеся структурно различными; 2) краткое формальное описание расщепленной конструкции в английском языке. Таким образом, работа вносит вклад в исследование общего и английского синтаксиса.

Ключевые слова: *формальные лингвистические понятия, синтаксис, расщепленная конструкция, псевдо-расщепленные предложения, английский язык, язык киньяруанда, современный китайский язык*

Для цитирования:

Mel'čuk I. The cleft construction: A formal definition. *Russian Journal of Linguistics*. 2025. Vol. 29. № 2. P. 218–249. <https://doi.org/10.22363/2687-0088-41329>

1. Introduction: Toward a formal definition of cleft construction	220
2. What is a rigorous linguistic definition?	223
3. The cleft construction in English	224
3.1. The formal representations of three English cleft sentences.....	224
3.2. Expression of a focalized Rheme by means of the cleft construction.....	227
3.3. The cleft construction as a linguistic sign.....	232
4. The notion of cleft construction	235
5. The cleft construction in languages of the world	238
6. Pseudo-cleft sentences	242

7. Conclusions	244
Acknowledgments	244
Appendices	244
Appendix 1. Some linguistic notions relevant for this paper	244
Appendix 2. Logically possible morphological expressive means and morphological signs.....	246
References	246

1. Introduction: Toward a formal definition of cleft construction

As is well known, one of the serious obstacles to the development of, and progress in, linguistics is the absence of a unified **notional system** of the type that exists in hard sciences. Many futile discussions are provoked by mutual misunderstandings and/or by different interpretations of even the most current terms. That is why I have been tackling this problem for almost half a century: an attempt to create a system of formal notions for linguistic morphology was made in (Mel'čuk 1982, 1993–2000, 2006), for semantics and syntax—in (Mel'čuk 1988, 2012–2015, 2021), and for phraseology—in (Mel'čuk 2023). Speaking of syntax, such syntactic notions as syntactic actant, surface-syntactic subject, government and agreement, passive construction, ergative construction, relative clause and pseudo-relative clause have been formally introduced. Here, another syntactic notion is considered: the cleft construction.

■ The name of a linguistic notion (= a technical term) on its first mention is printed in Helvetica; if need be, the notion is explained either in the subsequent text or in Appendix 1, pp. 244–245.

What follows is couched in terms of the Meaning-Text approach (e.g., Mel'čuk 1974, 2012–2015, 2016, 2018, among others).¹ The three main pillars of this approach are as follows:

- The description of a linguistic entity is carried out in the direction from Meaning to Text; it shows how a given meaning is expressed by this entity (rather than how this entity is understood).

- All formal representations of utterances are stated in terms of **dependency**—semantic, syntactic and morphological.

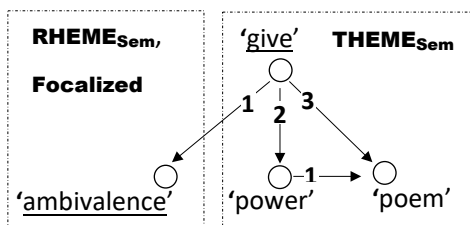
- Two levels of syntactic representation are distinguished: the deep-syntactic representation [DSyntR], which is closer to meaning, and the surface-syntactic representation [SSyntR], closer to text. The DSynt-structure [DSyntS] contains only semantically full lexical units (no grammatical lexical units), while the SSynt-structure [SSyntS] is made up of all actual lexemes of the utterance (including grammatical lexemes).

¹ A more complete list of publications on the Meaning-Text approach, as well as a brief overview of them, is presented in (Ivanova & Larina 2022).

The exposition presupposes sufficient familiarity of the reader with the conceptual apparatus and formalisms of the Meaning-Text approach. However, short explanations are provided wherever it seems necessary.

The prototypical cleft construction is illustrated in (1a), where its three lexical components are boldfaced; for a rigorous definition, see Section 4, pp. 235–236. In (1b) a (simplified) semantic representation [SemR] of sentence (1a) is given.

- (1) a. ***It is ambivalence that gives the poem its power.***
b.



- The underscoring of a semanteme within a communicative area in a semantic representation (in this case, of the semantemes 'ambivalence' and 'give') indicates its communicative dominance.

The cleft construction serves to express the focalization of the semantic Rheme [**RHEME_{Sem}**] of a sentence; in (1b) it is the semanteme 'ambivalence' that is focalized. In other words, the cleft construction plays a strictly communicative role: it marks on the surface the communicative value Focalized of the communicative category **Focalization** (see **Appendix 1**, p. 245).

The **RHEME_{Sem}** of a sentence is expressed on the deep-syntactic level by its **RHEME_{DSynt}**. Therefore, what is discussed in this paper is the expression of the focalized **RHEME_{DSynt}**.

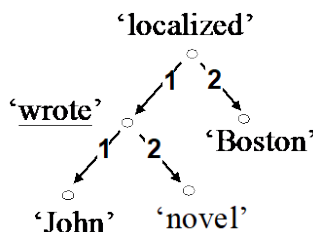
NB: For the semantic-communicative structure of sentences adopted here, see (Mel'čuk 2001).

A sentence including a cleft construction is called a **cleft sentence**. (The expression *cleft clause*, encountered in the literature, is to be avoided: it is a sentence that can be “cleft,” i.e. cleaved, into two clauses.)

The nature of the cleft construction in English was established by Otto Jespersen (1937: 83–89). Jespersen's description was developed in a huge number of studies: a Google search (2024.02.09) for “cleft sentences” produced 6,250,000 hits! Here I will mention just a few studies that have been most useful for myself. Thus, the classical English grammar (Quirk et al. 1985/1991: 1383–1387) offers a detailed characterization of English cleft constructions; see also (Sornicola 1988), (Collins 1991) and (Davidse 2000). The available information on cleft constructions in English, German and French is elegantly summarized in (Lambrecht 2001), an article that is basic for any study of clefts, including, of course, the present paper. I do not introduce new facts; all my data are borrowed from published research.

However, no rigorous general definition of the cleft construction as a particular type of linguistic unit has been supplied. It is this lacuna that I will be trying to fill.

It is convenient to start with a simple example. Let there be semantic structure [SemS] (2)—the depiction of a particular propositional meaning:
(2)



NB The SemS (2), as all the SemSs cited further, is incomplete: the meanings of grammemes (nominal number, verbal mood and tense) are not shown.

This SemS contains the communicatively dominant predicate semanteme—‘wrote’, its two Sem-actants—‘John’ and ‘novel’, and a semantic modifier—‘in [= localized.in] Boston’. Accordingly, from SemS (2) it is possible to produce several sentences with different semantic-communicative structures without focalization of the rhematic meaning; see, for instance, (3):

- (3) a. [*John*]**Sem-Theme** [*wrote this novel in Boston*]**Sem-Rheme**.
 b. [*This novel*,]**Sem-Theme, Focalized** [*John wrote it in Boston*]**Sem-Rheme**.
 c. [*In Boston*,]**Sem-Theme, Focalized** [*John wrote this novel*]**Sem-Rheme**.
 d. [*This novel was written by John*]**Sem-Theme** [*in Boston*]**Sem-Rheme**.
 e. [*This novel was written in Boston*]**Sem-Theme** [*by John*]**Sem-Rheme**.

(All these sentences are, of course, supposed to be uttered with the neutral prosody.)

At the same time, SemS (2) also underlies the cleft sentences in (4), where the focalized rhematic phrase is boldfaced:

- (4) a. *It is **John** that/who wrote this novel in Boston.*
 b. *It is **this novel** (that) John wrote in Boston.*
 c. *It is **in Boston** that John wrote this novel.*

The following fact is crucial:



All the sentences in (3) and (4)—the cleft and non-cleft ones—are produced from the same starting SemS: (2). In other words, all of these sentences have the same propositional meaning and differ only in their communicative organization.

|| The goal of this paper is a rigorous definition of the notion “cleft construction” and a formal characterization of this construction.

Thus, the paper constitutes another contribution to the long-term work on the notional system of linguistics mentioned above, at the beginning of this section.

2. What is a rigorous linguistic definition?

Since my aim is a rigorous definition of cleft construction, it is necessary to formulate first the requirements imposed on rigorous linguistic definitions. The most important class of linguistic definitions is that of deductive definitions; I think that the cleft construction must be defined deductively, and therefore I will characterize here deductive definitions only.²

A definition of a linguistic entity **E** should satisfy two sets of requirements: substantive ones and technical ones.

Three substantive requirements on a linguistic definition

1) A linguistic definition must be **strictly deductive**: when formulating a definition, the researcher proceeds from the more general to the more particular.

Usually, what is an **E** is intuitively more or less clear in central cases, but for many marginal specimens the answer is far from obvious; therefore, the notion of **E** must be sharpened. For this, it is necessary to define the most general notion of which **E** is a particular case. Attention: what is meant is the **most general notion**, not *genus proximum*, i.e. not ‘the nearest kind’: the notion the researcher is after must represent the most inclusive class of items to which **E** belongs, along, of course, with many other items, which are similar to, but essentially different from, **E**. This “superclass” must then be partitioned into the biggest subclasses available (ideally, into two subclasses), and this operation is repeated again and again, until we get a class that consists only of entities **E**.

Substantive requirement No. 1 ensures that the place of **E** among other similar notions (in terms of class inclusion) is well established.

2) A linguistic definition must strive for **maximal “narrowing” of E’s defining properties**: these properties must be distinguished and separated as much as possible, thus creating finer-grained notions. The researcher is supposed to place under the notion of each property as little as he can, so that **E** is defined by the maximal set of the finest possible relevant properties.

Substantive requirement No. 2 guarantees that the notion will be the least inclusive, i.e., the narrowest possible.

3) A linguistic definition must **be based on the prototypical cases of E**. This approach essentially follows what Charles Hockett (Hockett 1956) proposed about 70 years ago for the concept of grammatical case. Namely, “(grammatical) case” must be defined strictly on the basis of a prototypical case system, for instance, that of Latin or Ancient Greek, and then the notion has to be reasonably generalized—

² Two other types of linguistic definitions are inductive definitions (such as those of syntactic clause elements—subject, direct object, etc.) and enumerative definitions (such as that of lexical unit: a lexeme or an idiom).

such that new phenomena subsumed under the definition thus obtained are sufficiently similar to the Latin case.

No Eurocentrism is implied here. It is not that the Latin concept of case should be imposed on a completely different language. On the contrary, the name *case* should be used strictly for phenomena that are similar enough in criterial ways to the Latin case. If the phenomenon considered is not sufficiently similar to what we call *case* in Latin it simply should not be called *case*.

Substantive requirement No. 3 pushes the researcher towards intuitively more attractive notions.

Four technical requirements on a linguistic definition

A linguistic definition must be:

- 1) **Formal**—it should be applicable automatically, that is, literally.
- 2) **Rigorous**—it should contain only previously defined notions and/or else undefinable notions (*indefinibilia*), which must be listed as such. More precisely, it should be a definition of the axiomatic type: *per genus proximum et differentia specifica* ‘by the nearest kind and specific differences’, as established by Boethius (480–524), who was following the ideas of Aristotle.
- 3) **Sufficient and necessary**—it should cover all the phenomena that are perceived as being subsumable under the corresponding notion, and nothing but such phenomena.
- 4) **Logically universal**—it should be applicable to any relevant item of any language in order to check whether this item satisfies the given definition or not. (But this item itself as well as the entities in its definition are not meant, of course, to be necessarily language-universal.)

I use the cleft construction in English as a prototypical representative of cleft constructions in all languages. If a particular construction in a given language does not satisfy the proposed definition, it is not similar enough to the English cleft construction and therefore should not be called *cleft construction*.

3. The cleft construction in English

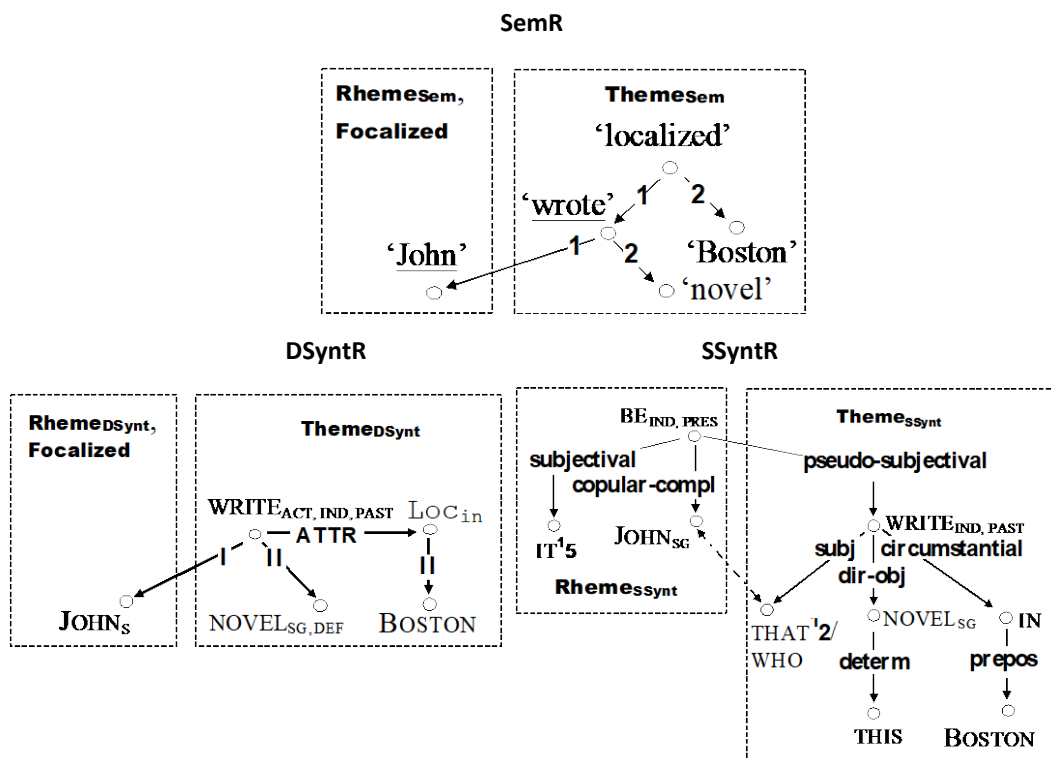
The English cleft construction will be characterized in two steps. First, the semantic, the deep-syntactic and the surface-syntactic representations of three illustrative cleft sentences—that is, sentences containing a cleft construction—are given (3.1). Second, the syntactic rules that ensure the expression of the focalization of a Rheme by a cleft construction are introduced (3.2).

3.1. The formal representations of three English cleft sentences

Here are the semantic [SemR], the deep-syntactic [DSyntR] and the surface-syntactic [SSyntR] representations of the cleft sentences in (5).

- A dashed two-headed arrow \longleftrightarrow in a Synt-structure shows a coreference link between two lexemes.

(5) [= (4a)] *It is John that/who wrote this novel in Boston.*



For the DSynt- and SSynt-relations, see (Mel’čuk 2021: Ch. 2).

Comments

1) As stated above (p. 221), a focalized Sem-Rheme corresponds to the focalized DSynt-Rheme, so that the cleft construction marks, strictly speaking, the focalization of the deep-syntactic Rheme.

2) The lexeme *IT*¹⁵ is a semantically empty pronoun, appearing as a dummy SSynt-subject; *THAT*¹² is a relative pronoun: \cong *WHICH*. (Lexicographic numbering here and below follows *Longman's Dictionary*.)

3) The subordinate clause in a cleft sentence depends on the copula verb BE by the **pseudo-subjectival** SSynt-relation (Mel’čuk 2021: 51–52). This is the SSyntRel used for all cases where a dummy SSynt-subject IT¹⁵ appears with the copula BE (or a similar verb such as SEEM or HAPPEN), the latter governing an expression that is, so to speak, “coreferential” with this IT¹⁵:

It←subjectival-is-[convenient]-pseudo-subjectival→to use the following technique.

or

It←subjectival—is—[known]—pseudo-subjectival→*that John is in town*.

4) The relative pronouns *THAT*¹² and *WHO* need the indication of coreference with their source (in this case, with the noun *JOHN*), since the grammatical number of this source noun controls the form of the Main Verb (*WRITE*) of the pseudo-subjectival subordinate clause of a cleft sentence:

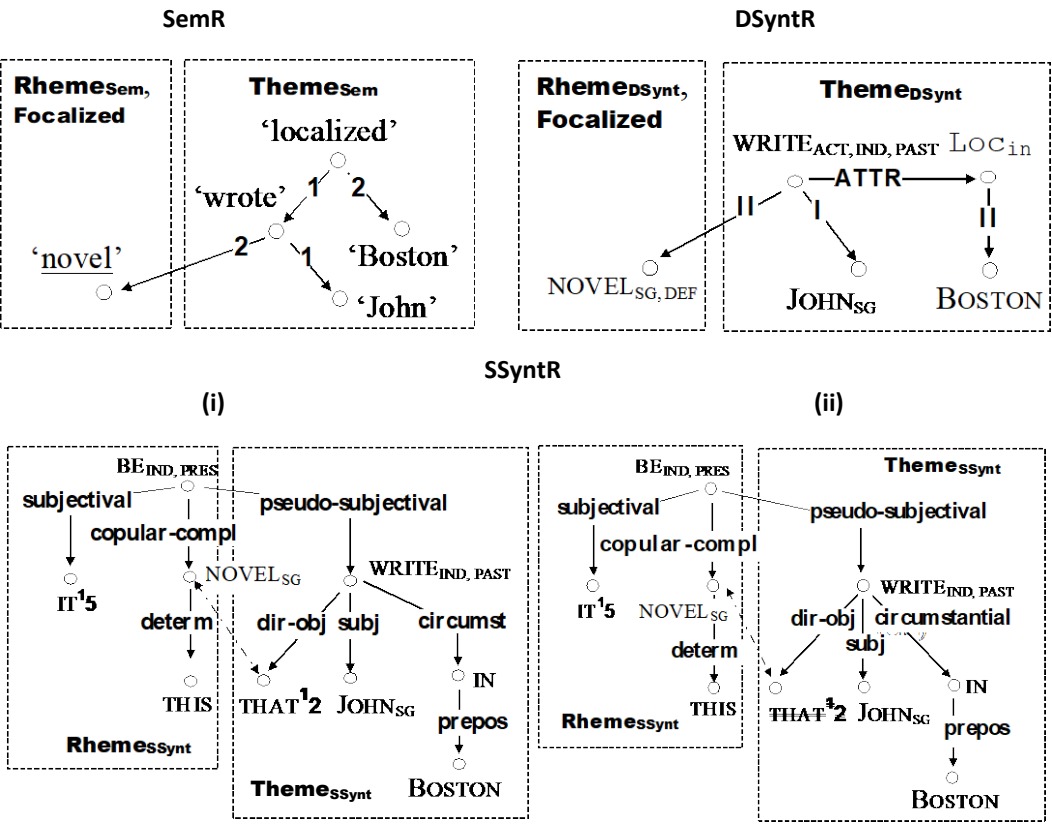
It is this person that/who←is writing the novel.

vs.

It is these people that/who←are writing the novel.

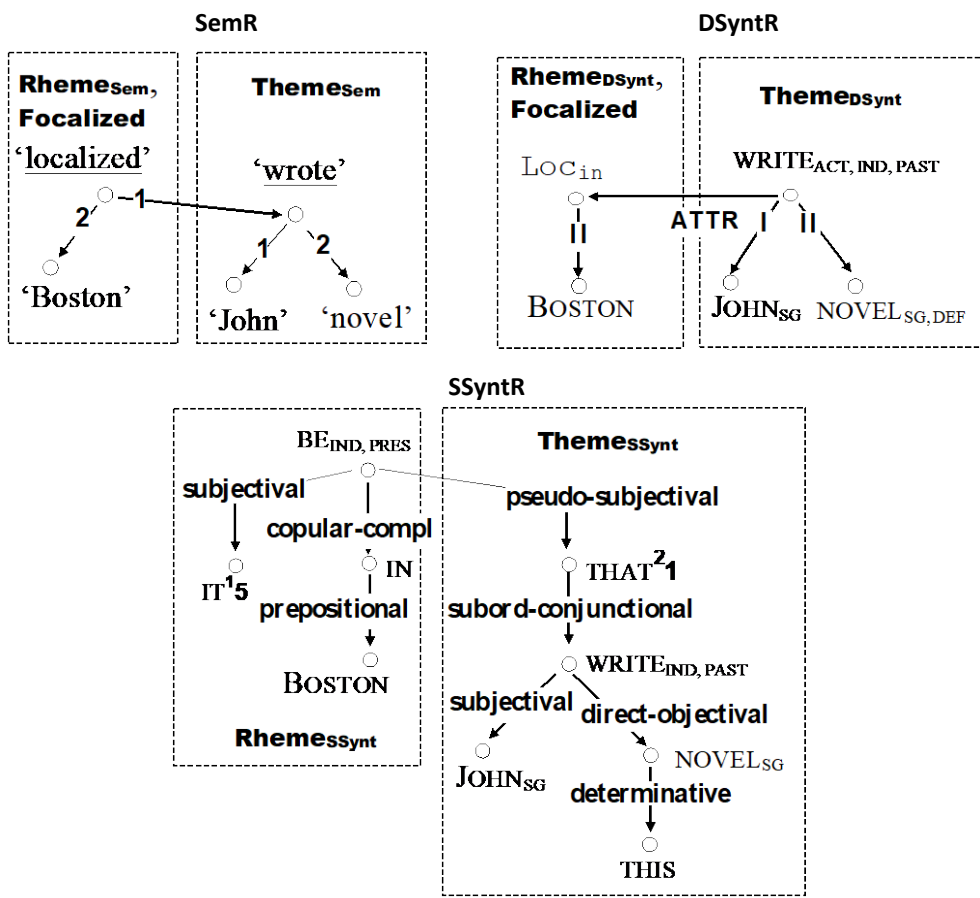
In some other languages that have the cleft construction, the Main Verb of the subordinate clause in a cleft sentence also reflects the person and gender of the relative pronoun's source (see Section 4, the examples (18) and (19)), so that the indication of coreference is justified.

(6) [= (4b)] *It is this novel (that) John wrote in Boston*.



In the SSyntR of (6-ii) the pronoun *THAT*¹² is elided, which happens typically when it is the direct object of the Main Verb in the subordinate clause of the cleft sentence. (The elision is shown by double strikethrough, which indicates that this lexeme does not appear in the morphological string, that is, in the deep-morphological representation.)

(7) [= (4c)] *It is in Boston that John wrote this novel.*



Now I can spell out the deep-syntactic rules (rules of the DSyntS \Leftrightarrow SSyntS form), which establish the correspondence between the DSynt- and the SSynt-structures of a cleft sentence. These rules, from now on, are referred to as Foc-Cleft rules.

3.2. Expression of a focalized Rheme by means of the cleft construction

The English Foc-Cleft rules given below claim neither exhaustiveness nor high accuracy: the goal is not to provide an ideal and exhaustive description of the English cleft construction, but only to offer a representative illustration of the proposed formalisms. (For factual details, see the titles mentioned in Section 1.) The English cleft construction subsumes a series of more or less marginal, slightly deviant cases, which cause doubts among speakers. I am unable to sort out the relevant facts and will rely on the intuition of my few informants.



The Foc-Cleft rules are presented in order to show how my formalisms can be used for the description of various expressions of this kind; these rules are more about the formal notion of cleft construction rather than about the real cleft construction in English.

Five Foc-Cleft rules are needed according to the SSynt-role that the focalized rhematic element L_1 plays in the corresponding sentence without such focalization.

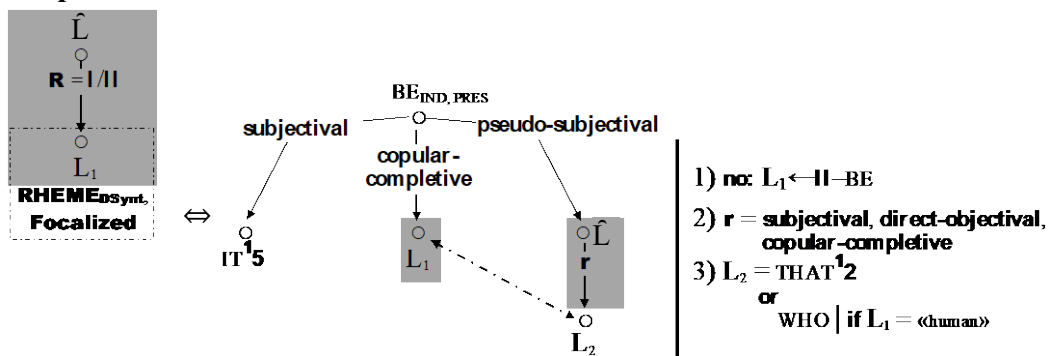
- Rule 1: L_1 is an SSynt-actant expressed by a prepositionless noun or an adjective; that is, it is a SSynt-subject, a direct object or a copular complement.
- Rule 2: L_1 is an SSynt-actant expressed by a preposition-introduced noun; that is, it is a surface-syntactic indirect or oblique object.
- Rule 3: L_1 is a possessor complement of a subject or a direct object.
- Rule 4: L_1 is a possessor complement of an indirect or oblique object.
- Rule 5: L_1 is a circumstantial.

These rules represent a rather rough picture of the cleft construction in English; the simplifications and omissions accepted are indicated on p. 231.

Foc-Cleft rules for English (DSyntR \Leftrightarrow SSyntR)

- The symbol \hat{L} stands for the lexeme L that is the syntactic head of the sentence. The symbols **R** and **r** stand, respectively, for a particular DSynt-relation and a particular SSynt-relation.
- Shading** indicates the context of the rule—that is, the components of the manipulated structures that are not affected by the rule, but whose presence is necessary for the rule to apply correctly.

Foc-Cleft rule 1: Focalizing subjects, direct objects and copular complements

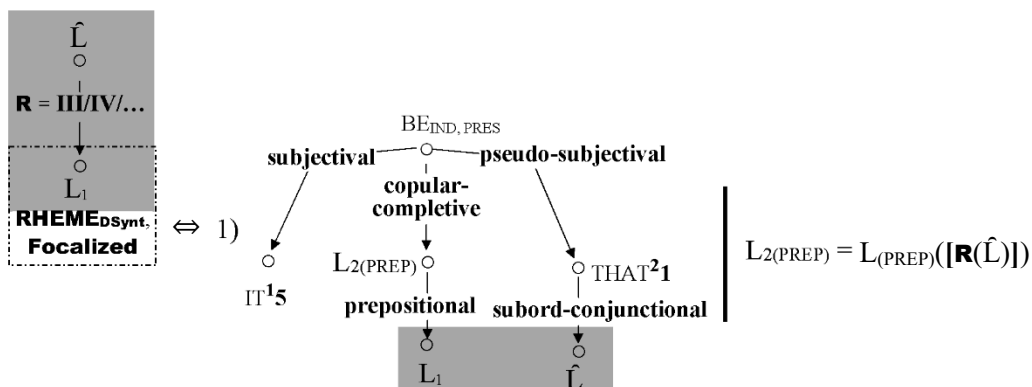


- (i) *It is food_{L1} that_{L2} ← subj-determines_L the shape of the mind.*
- (ii) *It is McGregor_{L1} and Househoffer that/who_{L2} ← subj-determine_L our plans.*
- (iii) *It is food_{L1} that_{L2} ← dir-obj-[we]-consider_L to be the determining factor.*
- (iv) *It is McGregor_{L1} that/who_{L2} ← dir-obj-[we]-consider_L to be the determining factor.*
- (v) **It is American_{L1}, not British, that_{L2} ← cop-compl-[McGregor]-is_L. [Condition 1)]*

vs.

It is American_{L1}, not British, that_{L2} ← cop-compl-[McGregor_{L1}]-became_L.

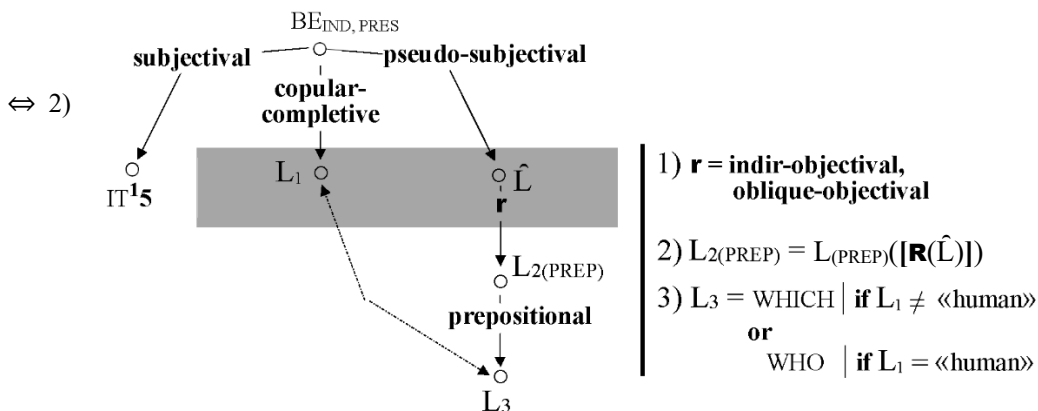
Foc-Cleft rule 2: Focalizing indirect and oblique objects



■ The expression $L_{(PREP)}([R(\hat{L})])$ means “the preposition that is indicated in the column **R** of \hat{L} ’s government pattern as a possible surface means marking the corresponding SSynt-actant.”

(vi) IndirO: *It is* to_{L_2} $McGregor_{L_1}$ *that*–[we]–subord-conj→*gave* $_{\hat{L}}$ *the book*.

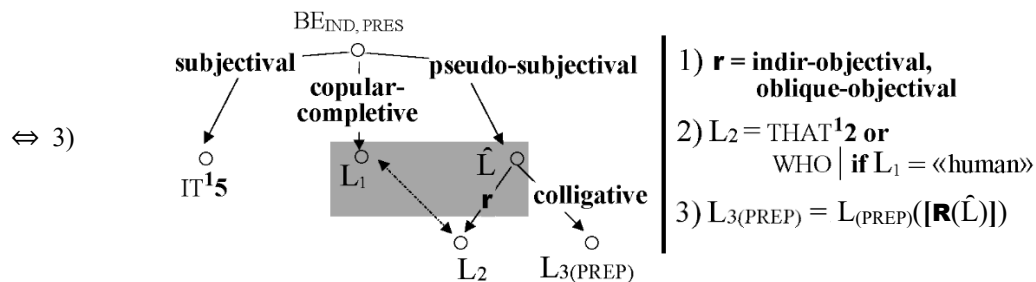
(vii) Oblo: *It is* on_{L_2} *this aid* $_{L_1}$ / on_{L_2} $McGregor_{L_1}$ *that*–[we]–subord-conj→*depend* $_{\hat{L}}$.



(viii) IndirO: *It is* $McGregor_{L_1}$ to_{L_2} –prepositional→*whom* $_{L_3}$ *we gave* $_{\hat{L}}$ *the book*.

(ix) Oblo: *It is* $McGregor_{L_1}$ on_{L_2} –prepositional→*whom* $_{L_3}$ *we depend* $_{\hat{L}}$.

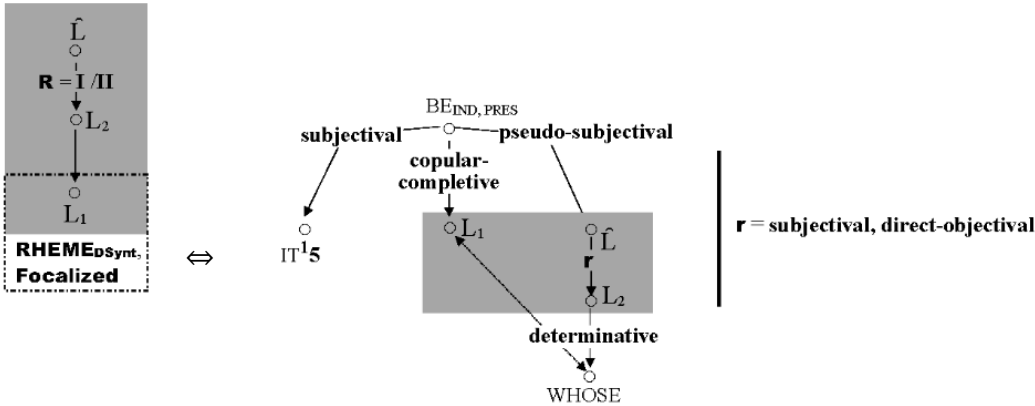
(x) Oblo: *It is these promises* $_{L_1}$ on_{L_2} –prepositional→*which* $_{L_3}$ *we depend* $_{\hat{L}}$.



(xi) IndirO: *It is* $McGregor_{L_1}$ *that/who(m)* $_{L_2}$ *we gave* $_{\hat{L}}$ –[*the book*]–colligative→ to_{L_3} .

(xii) Oblo: *It is these promises* $_{L_1}$ *that* $_{L_2}$ *we depend* $_{\hat{L}}$ –colligative→ on_{L_3} .

Foc-Cleft rule 3: Focalizing the possessor complement of the subject or the direct object

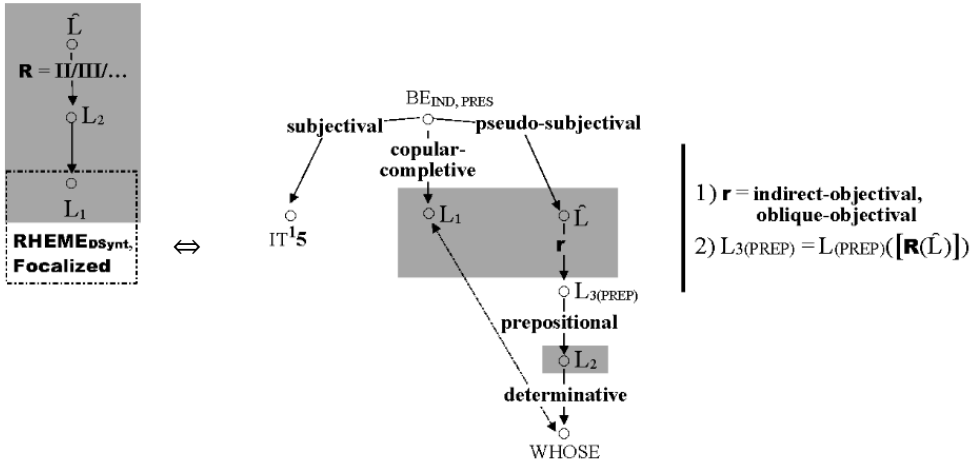


(xiii) *It is McGregor_{L1} whose←determinative-intervention_{L2} saved_L the day.*

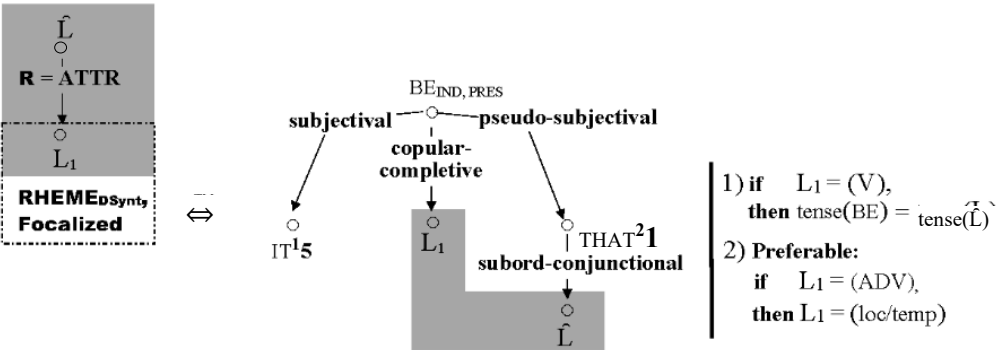
(xiv) *It is McGregor_{L1} whose←determinative-paintings_{L2} we adore_L.*

Foc-Cleft rule 4: Focalizing the possessor complement of the indirect or oblique object

(xv) *It is McGregor_{L1} on_{L3} whose←determinative-support_{L2} we depend_L.*



Foc-Cleft rule 5: Focalizing circumstantials



- (xvi) *It is just there_{L1} that-[we]-subord-conj→met_L McGregor.*
 (xvii) *It is yesterday_{L1} that-[we]-subord-conj→met_L McGregor.*
 (xviii) *It is with_{L1} <together_{L1} with> McGregor that-[we]-subord-conj→visited_L Boston.*
 (xix) *It is on_{L1} "The National" that-[we]-subord-conj→saw_L this news.*
 (xx) *It is because_{L1} of McGregor that-[we]-subord-conj→are_L late.*
 (xxi) **It is to see_{L1} McGregor that-[we]-subord-conj→came_L to Boston. [Condition 1)]*

vs.

It is to see_{L1} McGregor that-[we]-subord-conj→come_L to Boston.

and

It was to see_{L1} McGregor that-[we]-subord-conj→came_L to Boston.

- (xxii) *?It is carefully_{L1} that-[McGregor]-subord-conj→opened_L the box. [Condition 2); cf. examples (xvi) and (xvii)]*

vs.

It is with_{L1} care that-[McGregor]-subord-conj→opened_L the box.

Comment

Condition 1): In case of the focalization of a circumstantial, if this circumstantial is a verb (an infinitive of purpose), the tense of the subordinate clause must be the same as the tense of the superordinate clause (= the tense of BE).

Condition 2): If L_1 is an adverb, it preferably is locative or temporal; for instance, *It is far away that John lives* or *It is very early that John begins his day*, but *?It is very warmly that John was received there*. However, qualitative adverbs are also possible: *It was sadly that he smiled* or *It was very slowly that he answered*.

Simplifications adopted in the Foc-Cleft rules

The Foc-Cleft rules presented above gloss over the following four important properties of the English cleft construction.

1) The cleft construction readily appears also in negative and interrogative sentences:

- (8) a. *It is not John who insists on our leaving.*
 b. *Is it John who insists on our leaving?*
 c. *Is it not John who insists on our leaving?*

Our Foc-Cleft rules ignore this fact.

2) The verb BE in the superordinate clause of a cleft sentence is not necessarily in the present tense (contrary to what is indicated in our Foc-Cleft rules):

- (9) a. *It was John who brought the booze.*
 b. *From day one, it has been the Americans who politicized the issue.*
 c. *If we fail, it will be overconfidence that is our undoing.*

3) If L_1 —the rhematic element to be focalized—does not depend immediately on the syntactic head \hat{L} of the sentence, the copular complement in the subordinate clause must be the highest SSynt-governor of L_1 that depends directly on \hat{L} . For instance, in the sentence *John reads only←restrictive veryRHEME, Foc short stories* the focalized rhematic element is the adverb VERY; then the corresponding cleft

sentence is *It is only VÉRY short stories that John reads*, where the copular complement is the head of the phrase *very←short←stories*, and the focalized element is marked by prosody (phrasal stress). This particularity is also not reflected in the Foc-Cleft rules.

4) \hat{L} in a Foc-Cleft rule can be not a single lexeme $L_{(V)}$, but is instead a standard syntactic subtree (Mel'čuk & Pertsov 1987: 193–195, 485–491), a.k.a. a verbal nucleus (Kahane 2001), that is, a syntactic chain of subsequently subordinated lexemes that is syntactically equivalent to a simple verb; for instance, instead of *John wrote_L this text* one may encounter *John might_L→have→written this text*. Such a standard subtree can be transferred from the subordinate to the superordinate clause of a cleft sentence *salva significatione* (replacing, of course, the label of the terminal node by the corresponding form of the verb BE):

- (10) *John_{RHEME, Foc} might_L have written this text.* ~
It is John who might have written this text. ≡
It might have been John who wrote this text.

Moreover, some adverbials (mostly, parentheticals) can be transferred from the subordinate to the superordinate clause of a cleft sentence:

- (11) *John_{RHEME, Foc} probably ⟨, as you know,⟩ wrote this text.* ~
It is John who probably ⟨, as you know,⟩ wrote this text. ≡
It is probably ⟨, as you know,⟩ John who wrote this text.

However, our Foc-Cleft rules do not account for this operation.

Similarly, the copula verb in the cleft construction can also be the terminal component of a syntactic standard subtree, as, for instance, in (12):

- (12) a. *It can be these moments that keep us from speaking up.*
 b. *It seems to be these people who were incarcerated.*

This possibility is not accounted for, either.

Our rules ignore as well three syntactic phenomena that are not specific to the cleft construction:

- The possible omission (= ellipsis) of the relative pronoun THAT¹²/WHO that plays the SSynt-role of the direct object or of a dangling preposition's object, as seen in *It is the last statement ⟨McGregor⟩ ~~that~~ ⟨who⟩ we believe* or *It is the last statement ⟨McGregor⟩ ~~that~~ ~~who~~ we believe in*, and of the conjunction THAT²¹, as in *It is there ~~that~~ we met McGregor*.

- The choice between lexemes WHO and THAT¹²/WHICH in case L_1 is a human noun.

- The choice between morphological forms *who* and *whom*, which is fluctuating in English.

3.3. The cleft construction as a linguistic sign

The cleft construction is a complex linguistic sign—a member of the huge family of linguistic signs that includes lexical units, morphological markers of inflectional and derivational significations (see the table of morphological signs in

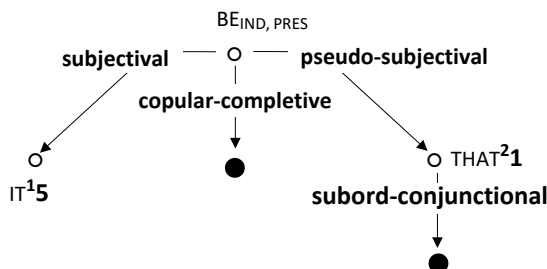
Appendix 2, p. 246), meaning-carrying SSynt-relations, etc. It must be described as all signs are—by specifying its three components: signified, signifier, syntactics.

The **signified** of the cleft construction is unproblematic: it is the DSynt-communicative value **Focalized**, characterizing the **RHEME**_{DSynt} of the sentence and expressing the Sem-Comm-value **Focalized** (see above, p. 221); there is a general consensus concerning this point. The expression of communicative values by segmental means—in this case, by the phrase of the type *It is* [X] *that/who*...—is rather infrequent in the world's languages, but nonetheless it is well known. Some stock examples include the English idiom 'AS FOR' [X] and the lexeme SPEAKING [*of* X], which express the DSynt-communicative value **Focalized** characterizing the **THEME**_{DSynt} of the sentence, the equivalent Russian idiom 'ČTO KASAETSJA' [X-*a*], or Russian lexemes—particles ÈTO³ [X], expressing the focalization of the **RHEME**_{DSynt} (*Èto Ivan prišël* 'This is Ivan who has arrived'), and [X]-TO⁴, which expresses the focalization of the **THEME**_{DSynt} (*Ivan-to pridët* 'As for Ivan, he'll come').

- The top corners ' ' enclose an idiom (= a non-compositional phraseme).

The **signifier** of the Foc-Cleft construction is, as said above, the phrase *It is* [X] *that/who...*; formally speaking, it is a SSynt-subtree of one of the two following forms:

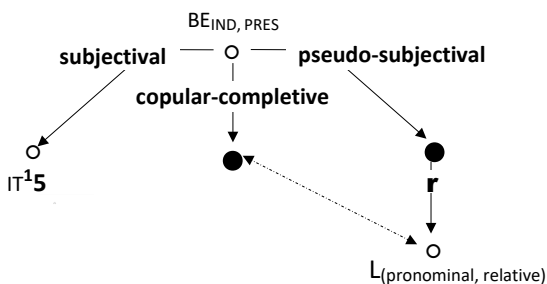
(13) a.



(the subordinate pseudo-subjectival clause of a cleft sentence is a regular clause introduced by the conjunction THAT²¹)

or
b.

b.



(the subordinate clause of a cleft sentence is a pseudo-relative clause introduced by the corresponding relative pronoun: **THAT**¹² or **WHO**).

- The blackened nodes of the above subtrees are “the contact points”: by means of these nodes the Foc-Cleft substructure is joined to the SSynt-structure of the cleft sentence under production.

These two cleft construction subtrees are approximate and schematic, in the sense that they illustrate the cleft construction along very general lines. Strictly speaking, for English, it is necessary to consider all the variants of the Foc-Cleft subtrees represented by the seven subrules in Subsection 3.2: these are, so to say, seven “allo-clefts” of one “clefteme,” distributed as function of the DSynt-context.

The **syntactics** of the cleft construction is quite involved: the way it is hooked up to the rest of the SSynt-structure of the sentence under construction cannot be spelled out by a series of sufficiently simple statements. (In any case, for the time being I am unable to do so.) It has to be specified by a set of complex conditions distributed between the five above rules.

Now that the sign under consideration is formally characterized, I have to answer an obvious question:

||What type of linguistic sign is the cleft construction?

It is not a morphological sign—that is, its signified is not expressed within a wordform.³ It is a syntactic sign: its signified is expressed within an SSynt-structure, in other words, within a sentence. It is, thus, similar to DSynt-relations, which are also expressed by SSynt-subtrees, except that the cleft construction carries a communicative rather than a syntactic signification. The type of a sign is established by the type of its signifier, and syntactic signs have just three possible types of signifiers: 1) segmental signifiers, i.e. lexical units; 2) prosodies; and 3) linear order, specified by SSynt-relations.

In a language with flexible word order, such as Russian, a communicative value of a phrase can control the linear implementation of an SSyntRel: thus, the SSynt-configuration

IVAN←**subjectival**–VSTAT'_{PAST} ‘Ivan stood up’

surfaces as *Ivan vstal* if the lexeme IVAN is the Synt-Theme, and as *Vstal IVÁN* (with prosodic emphasis on *Ivan*) if IVAN is the Synt-Rheme. Hungarian presents a more complex case: if a given clause has no focalization, a prefixed Main Verb appears in its SSyntS as such; but if focalization of a sentence element is present, the prefix of the verb splits from it and is linearly positioned according to fairly complex rules. For instance, consider the verb BE+JÖT(-ni) ‘enter, come into’ in sentences (14):

³ For a better orientation of the reader, a table of possible morphological expressive means and possible types of morphological signs, for which there exists a rigorous calculus, is presented in Appendix 2.

- (14) a. *Maria be+jötött a szobába* ‘Maria came into the room’.
vs.

b. *Maria jötött—prefixal-auxiliary→be a szobába*
‘**MARIA** [and not somebody else] came into the room’.

In (14b), MARIA constitutes a focalized Rheme, and this communicative value is expressed by the introduction (into the SSyntS) of a new node—a split prefix **be-**, which syntactically depends on the lexical part of the verb and is positioned by the **prefixal-auxiliary** SSyntRel.

The signifier of the cleft construction is, we have seen, segmental: it is a phrase, and this phrase is semantically non-compositional. Therefore, this sign is a lexical unit. More precisely, it seems to be an idiom: ‘IT IS [X] THAT/WHO ...’.

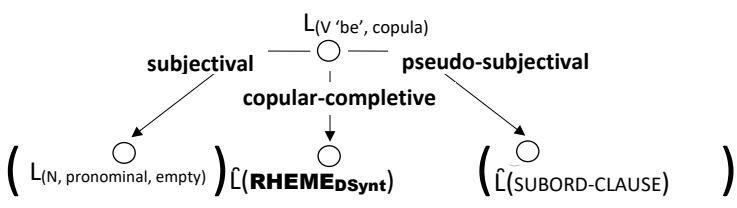
True, ‘IT IS [X] THAT/WHO ...’ is quite an unusual idiom: normally, an idiom appears as a label on one of the nodes in a DSyntS, while ‘IT IS [X] THAT/WHO ...’ does not. This idiom is a surface grammatical lexical unit—like, for instance, another English idiom, ‘AS FOR’ [X], and the similar Russian units mentioned above. As an idiom, it exists only in the lexicon, because it never appears in the DSyntS; and in the SSyntS its separate components appear individually. **NB:** This conclusion is made more precise in the next section.

4. The notion of cleft construction

Given its communicative role—marking the focalization of the DSynt-Rheme—and its quite complex syntactics, the expression ‘IT IS [X] THAT/WHO ...’ is specific enough to merit a special name—the *cleft construction*. Let us proceed to the main goal of the paper—a formal definition of the notion “cleft construction,” applicable universally. In other words, this definition based on the English cleft construction as the prototypical case must be good for Language in general.

Definition: cleft construction

A cleft construction is a grammatical lexical unit, namely, a surface-syntactic idiom, that expresses the focalization of the sentence’s **RHEME_{DSynt}** and whose signifier is an SSynt-subtree of the following general form:



where:

- L(V ‘be’) is a particular lexeme of the copula verb (with the meaning ‘be’).
- L(N, pronominal, empty) is a dummy pronoun such as Eng. IT¹⁵ (Fr. IL, Ger. ES or a zero lexeme— \emptyset_{SG}^{EMPTY} , e.g., in Spanish and Italian); the parentheses indicate that it can be absent (as, e.g., in Chinese).

$\hat{L}(\mathbf{RHEME}_{DSynt})$ is the lexeme that constitutes the \mathbf{RHEME}_{DSynt} or is the SSynt-head of the phrase that constitutes the \mathbf{RHEME}_{DSynt} .

$\hat{L}(\text{SUBORD-CLAUSE})$ stands for the SSynt-head of the subordinate pseudo-subjectival clause, this SSynt-head being:

- either an empty complementizer, such as Eng. THAT²¹ (Fr. QUE, Ger. DASS);
- or a finite verb that is the syntactic head of this subordinate clause and that has a particular relative pronoun as a dependent, such as Eng. THAT¹² or WHO; the parentheses mean that the complementizer and/or the relative pronoun can be absent (as, e.g., in Chinese).

As one sees, in a language where both the pronoun $L_{(N, \text{pronominal, empty})}$ and the relative pronouns are absent, the cleft construction is reduced to just a grammatical lexeme: one of the lexemes of the copula verb. Roughly (in plain English):

The cleft construction is a grammatical (more precisely, surface-syntactic) idiom headed by the copula verb, similar to Eng. BE; the lexical entry of this verb contains a complete characterization of the cleft construction.⁴ In a particular language this idiom can be reduced to a single lexeme—to a lexeme of the copula verb.

NB According to (Andrade 2019), in colloquial Brazilian Portuguese even the copula verb SER ‘be’ of a cleft sentence can be elided on the morphological surface (but the complementizer is preserved):

- (15) *O João que eu vi* ‘(It was) John that I saw’.
 the John that I saw

Thus, the goal of the paper, formulated on p. 222, is reached.

The proposed definition of cleft construction corresponds to the requirements formulated in Section 2. On the one hand, it is strictly deductive (the cleft construction is defined as a particular subclass of linguistic signs, more precisely, a subclass of grammatical lexical units), its defining features are separated enough (its signified, its signifier and its syntactics are characterized separately), and it is based on a well-known prototypical case—the English cleft construction. On the other hand, it is formal, rigorous, necessary and sufficient, and logically universal. (To avoid a possible confusion, let it be reminded that entities mentioned in the definition are not language-universal, just like the cleft construction itself; however, they are defined without mentioning specific features of a particular language.)

A cleft sentence is a complex sentence: it consists of two full-fledged clauses—the superordinate and the subordinate ones.

- The superordinate clause has the following SSynt-structure:

$L_{(N, \text{pronominal, empty})} \leftarrow \text{subjectival} - L_{(V, \text{copula})} \text{ ‘be’ } - \text{copular-completeive} \rightarrow \hat{L}(\mathbf{RHEME}_{DSyntS})$

⁴ This formulation corrects a previous attempt to describe the cleft construction: (Mel'čuk 2001: 180–190), where it was proposed to represent the cleft construction in the DSyntS by a fictitious lexeme «IT_BE» (instead of a particular lexeme of the vocable BE).

• The subordinate clause is introduced, as indicated above, either by an empty subordinating conjunction (such as THAT²¹ in English), or by a relative pronoun (such as the THAT¹² and WHO in English), or else it has no explicit marker, as in Chinese, see examples (27) and (28). The subordinate clause introduced by THAT¹² or WHO is formally similar to a relative clause; however, these two clause types are essentially different.

It is useful to indicate some of the surface-syntactic differences observed in English between the subordinate clause of a cleft sentence with focalized rhematic noun L (16a) and a genuine restrictive relative clause modifying the noun L that depends on the verb BE (16b):

- (16) a. *It is*—[*this gentleman*_L]**—pseudo-subj**→*who told me about the trip.* ~
b. *This is the gentleman*_L**—relative**→*who told me about the trip.*

These differences have been discussed by many authors, so that there is nothing original in the table below: I simply have collected and systematized well-known facts.

Table 1

Cleft sentence	Sentence with a restrictive relative clause
1. L—the copular complement of ‘be’—is the focalized rhematic element.	1. L—the copular complement of ‘be’—is a noun modified by a relative clause.
2. The subordinate clause depends on the verb ‘be’, that is, on the Main Verb of the superordinate clause, by the pseudo-subjectival SSyntRel.	2. The subordinate clause depends on L by the relative SSyntRel.
3. <i>It</i> is the empty pronoun <i>IT</i> ¹⁵ ; it does not alternate with anything.	3. <i>It</i> is the demonstrative pronoun <i>IT</i> ¹¹ ‘this/that thing/situation’; it is referential and alternates with <i>THIS/THAT</i> : <i>It/This/That is a novel that John is reading.</i>
4. <i>That</i> is the conjunction THAT ²¹ : <i>It is tomorrow that I am leaving.</i>	4. <i>That</i> is a relative pronoun THAT ¹² , which is coreferential with its antecedent; it alternates with <i>WHO</i> and <i>WHICH</i> .
5. L can be a proper name or personal pronoun: <i>It is John/him that [= THAT²¹] I want to talk to.</i>	5. L cannot be a proper name or a personal pronoun: <i>*This (man) is John/him that [= THAT¹²] I want to talk to.</i>
6. The verb BE agrees with <i>IT</i> (BE is in the singular, even if L is a noun in the plural): <i>It is novels that John enjoys reading.</i>	6. The verb BE agrees with L; if L is in the plural, a plural demonstrative pronoun <i>these</i> has to be used: <i>These are novels that John enjoys reading.</i>
7. The tense of the Main Verb of the subordinate clause of a cleft sentence constrains the tense of the Main Verb BE of the superordinate clause; thus, if the former is in the present, the latter cannot be in the past: <i>*It was the novel that John is still reading.</i>	7. The tense of the Main Verb of the relative clause does not constrain the tense of the Main Verb BE of the superordinate clause: <i>This was the novel that John is still reading.</i>
8. L carries a rising intonation and the subordinate clause, a falling one; a pause is possible between L and the subordinate clause: ↗ ↘ <i>It is the dish that John enjoys.</i>	8. L and its relative clause carry a single falling intonation; a pause is possible between <i>THIS</i> and the rest of the sentence: ↘ <i>This is the dish that John enjoys.</i>

Differences between cleft sentences with a focalized noun L and sentences with a relative clause that depends on the copular complement L of BE

The proposed definition of cleft construction excludes from this notion other constructions that express focalization of the Rheme as well, but have a completely different syntactic structure. For instance, the Russian construction with focalizing particle ÈTO³ (see subsection 3.3, **Signified**, pp. 233) should by no means be called “cleft construction”: it features no clefting. In the same vein, a pseudo-cleft shows no syntactic clefting, so its name is misleading.

5. The cleft construction in languages of the world

The following remarks are far from being a serious typological overview of cleft constructions in languages other than English; they are meant to be but a preliminary hint to wet the appetite of prospective researchers.

Besides English, the cleft construction is richly represented in French (see Doetjes et al. 2004):

(17) French

- | | | | | | | | | | |
|----|--------------|-------------|--------------|------------|--------------|-----------|--------------|-----------|---------------|
| a. | <i>C'est</i> | Jean | <i>qui</i> | <i>a</i> | <i>écrit</i> | <i>ce</i> | <i>roman</i> | <i>à</i> | <i>Paris.</i> |
| | this is | Jean | who | has | written | this | novel | in | Paris |
| b. | <i>C'est</i> | <i>ce</i> | roman | <i>que</i> | <i>Jean</i> | <i>a</i> | <i>écrit</i> | <i>à</i> | <i>Paris.</i> |
| | this is | this | novel | that | Jean | has | written | in | Paris |
| c. | <i>C'est</i> | <i>à</i> | Paris | <i>que</i> | <i>Jean</i> | <i>a</i> | <i>écrit</i> | <i>ce</i> | <i>roman.</i> |
| | this is | in | Paris | that | Jean | has | written | this | novel |

The French cleft construction has several surface differences with respect to the English one:

- Instead of a dummy pronoun *IT*, a regular equivalent of the Eng. *IT*, the demonstrative pronoun *CE* ‘this’ is used.

- Instead of the contrast between the relative pronouns *THAT*¹² [with any source] ~ *WHO* [with a human source only], French features another contrast: *qui* ‘which’-nominative case ~ *que* ‘which’-accusative case.

- In French the Main Verb of the subordinate clause of a cleft sentence whose SSynt-subject is a relative pronoun agrees in person and number with the source of this pronoun, while in English the Main Verb of such a subordinate clause agrees with this pronoun only in number (i.e., it is invariably in the 3rd person):

- (18) a. Fr. *C'est moi qui suis* *responsable pour cet effet.*
be-IND.PRES.1.SG

vs.

- b. Eng. *It is me who is* *responsible for this effect.*
be-IND.PRES.3.SG

- In French, the Main Verb of the subordinate clause of a cleft sentence—more specifically, the past participle in the compound tenses (19a)—and its copular complement (19b) agree with the source of the relative pronoun also in gender:

- (19) a. *C'est ce **roman** que nous avons compris+Ø.*

vs.

*C'est cette **nouvelle** que nous avons compris+e.*

- b. *C'est ce **roman** qui est intéressant+Ø.*

vs.

*C'est cette **nouvelle** qui est intéressant+e.*

A similar type of cleft construction is found in Italian, Spanish and Portuguese (but not in Romanian—see Gheorghe 2017: 2). For instance:

- (20) Spanish

- a. *Fue **Juan** quien escribió esta novela en Paris.*
was Juan who wrote this novel in Paris
- b. *Es esta **novela** la que Juan escribió en Paris.*
is this novel the that Juan wrote in Paris
- c. *Es **en Paris** donde Juan escribió esta novela.*
is in Paris where Juan wrote this novel

This construction existed already in pre-classical Latin (Dufter 2009: 1):

- (21) *Non ego sum qui te dudum conduxī* [Plautus 254 – 184 BC]
not I am who you just.now I.hired
'It is not me who hired you just now'.

Besides English, the cleft construction is found in other Germanic languages—German, Dutch and all Scandinavian languages (see Fischer 2009); for instance:

- (22) German

- a. *Es war **Johann**, der diesen Roman in Paris geschrieben hat.*
it was Johann which this novel in Paris written has
- b. *Es war dieser **Roman**, den Johann in Paris geschrieben hat.*
it was this novel which Johann in Paris written has
- c. *Es war **Paris**, wo Johann diesen Roman geschrieben hat.*
it was Paris where Johann this novel written has

NB In languages with a relatively flexible word order—such as Spanish, Italian and German—clefts are considered by purists as superfluous, since word order and prosody can do the job, that is, express focalization quite well.

The cleft construction is typical of Celtic languages; thus, we find it in Irish (Stenson 2020: 67–70):

- (23) a. *Is **mise** a d'oscail an doras* 'It is me that opened the door'.
is me REL I.opened the door

NB The lexeme A, glossed as REL(ator) \approx 'that', is a subordinator—a particle?—introducing a relative or a pseudo-relative clause.

- b. *Ba go Gaillimh a chuaigh sé* 'It was to Galway that he went'.
was to Galway REL he.went he
- c. *Is **róchliste** atá Seán* lit. 'It is too clever that John is'.
is too.clever REL.is John [atá \Leftarrow a + is]

- d. *Is é gur cheannaigh Seán carr a chreideann Máire*
 is it that bought John car REL believes Mary
 lit. 'It is that John bought a car that Mary believes'.

Examples (23c–d) are from (Maki & Ó Baoill 2014); the word-for-word equivalents of these Irish sentences are ungrammatical in English. The authors indicate that in Irish focalization clefting is much freer than in English, which allows one to speculate that the prevalence of the cleft construction in English might be due to the Celtic substratum.

It is worth emphasizing that the cleft construction is absent from Slavic languages—with the exception of Czech and Ukrainian:

- (24) a. Czech (Reeve 2012: 167)

Je to manželka, kdo rozhoduje 'It is the wife who decides'.
 is that wife who decides

- b. Ukrainian (Duma 2022: 1)

To je spadok, ščo xvyľjuje joho 'It is the inheritance that worries him'.
 that is inheritance that worries he-ACC

NB: Example (24b) is dubious. My colleague, linguist-Ukrainist Dr. Volodymyr Trub, in his personal communication (2025.06.21) denies the existence of the cleft construction in Ukrainian.

Turkic and Finno-Ugric (Finnish, Estonian, Hungarian) languages lack the cleft construction as well. However, it is found in Africa, in particular, in Bantu languages, for instance, in Kinyarwanda.

- (25) Kinyarwanda (Kimenyi 1980: 70–72; the relativization of a clause is expressed by the high tone ' on the stem of its Main Verb)

a. *Na ábá+ana b+a +gíí +ye kw' iishuûri*
 be II child II PAST go.REL COMPL(itive) to school
 '{It} is the children who went to school'.

b. *N'i í+kárámu umu+koóbwa y+a +háá +ye úmw+áana*
 be IX pen I girl I PAST give.REL COMPL I child
 '{It} is the pen that the girl gave to the child'.

c. *Ni ku maguru umw+áana y+a +gíí +ye*
 be on feet I child I PAST go.REL COMPL
 '{It} is on foot that the child went'.

And now, to the cleft construction on the other side of the globe: in Mandarin Chinese.

The situation with the cleft construction in Chinese is controversial—in the sense that the dozens of researchers who have discussed it have not been able to come to a (more or less) unanimous conclusion as to what is and what is not a cleft construction in this language. Nevertheless, one of the candidates for the title of cleft construction corresponds to the definition of Section 4, representing its extreme, or “limiting,” case; therefore, it seems necessary to consider this construction here.

Because of my ignorance of Chinese facts, I am not in a position to participate in the discussion; I will simply adopt the most neutral position (following, although not fully and not literally, Teng 1979 and Paul & Whitman 2008) and use only the least controversial examples.⁵ (26) gives the underlying—non-cleft—sentence; the sentences in (27) provide typical examples of the Chinese cleft construction in the strict sense of our definition—with the copula verb SHÌ ‘be’ taking the focalized Rheme (boxed in the examples) as its complement:

- (26) *Wǒ zuótiān gěi Lì dǎ-le diànhuà* ‘I called Li yesterday’.
 I yesterday to Li hit COMPL(ative) phone
- (27) a. *Shì→wǒ zuótiān gěi Lì dǎ-le diànhuà* ‘It is me who called Li yesterday’.
 be I yesterday to Li hit COMPL phone
 b. *Wǒ zuótiān shì→gěi Lì dǎ-le diànhuà* ‘It is Li that I called yesterday’.
 I yesterday be to Li hit COMPL phone
 c. *Wǒ shì→zuótiān gěi Lì dǎ-le diànhuà* ‘It is yesterday that I called Li’.
 I be yesterday to Li hit COMPL phone

NB 1. The underscoring identifies the subordinate pseudo-relative clause. The absence of a subordinating conjunction and of a relative pronoun in this clause corresponds to the nature of Chinese syntax: the absence of explicit subordinators in several types of subordinated clauses.

2. *Wǒ zuótiān* in (27b) and *wǒ* in (27c) are SSynt-prolepses, which express the Theme (‘speaking of ...’); a pause is possible after them. Chinese being a strongly Pro-Drop language, the repetition of *Wǒ* as the SSynt-subject of the verb *Dǎ* is elided.

The sentences in (27) consist each of two full-fledged finite clauses: the superordinate clause with the copula verb SHÌ ‘be’ and the subordinate pseudo-relative clause; this represents the clefting. The verbs in both clauses readily accept the negation—BÙ or MÉI—and a modal verb such as KĚNÉNG ‘may’; this demonstrates the genuine verbal finite character of both clauses.

NB BÙ negates stative facts, and MÉI, dynamic ones that did not take place in the past; BÙ changes its falling tone to the rising one and becomes BÚ, when followed by a syllable with the falling tone (tonal dissimilation); MÉI does not combine with the marker of the completive, *-le*.

- (28) a-i. *Bú shì wǒ zuótiān gěi Lì dǎ-le diànhuà*
 It is not me who called Li yesterday’.
 ii. *Kěnéng shì wǒ zuótiān gěi Lì dǎ-le diànhuà*
 ‘It may be me who called Li yesterday’.
 b-i. *Shì wǒ zuótiān méi gěi Lì dǎ diànhuà*
 ‘It is me who did not call Li yesterday’.
 ii. *Shì wǒ zuótiān kěnéng gěi Lì dǎ-le diànhuà*
 ‘It is me who may have called Li yesterday’.

⁵ The Chinese examples (26) – (29) have been verified, corrected, reverified, recorrected and reverified again by Li Liu, who I kindly ask to receive my most heartfelt gratitude.

Thus, it can be concluded that a cleft construction as defined in Section 4 does exist in Chinese.

However, what about a different type of sentence used in Chinese for the focalization of the Theme, but currently subsumed also under the name of “cleft construction”? I mean the sentences like those in (29):

- (29) a. *Zuótiān gěi Lǐ dǎ-le diànhuà-de shì wǒ*
yesterday to Li hit COMPL phone NOMINALIZER be I
‘[The one who called Li yesterday]_{THEME, Foc} is me’.
- b. *Wǒ gěi Lǐ dǎ-le diànhuà-de shì zuótiān*
I to Li hit COMPL phone NOMINALIZER be yesterday
lit. ‘[That I called Li]_{THEME, Foc} is yesterday’.

NB The marker *-de* is a nominalizer used to form a particular type of relative and pseudo-relative clauses. It means ‘(the one) who...’ or ‘(the fact) that...’.

The sentences in (29) are nothing else but so-called pseudo-cleft sentences, which are treated immediately below.

6. Pseudo-cleft sentences

Cleft sentences are commonly opposed to, but as a rule considered together with, so-called pseudo-cleft sentences (a detailed overview of pseudo-cleft sentences of various types is found in Collins 1991 and De Cesare 2017):

- (30) a. **The person** who wrote this novel in Boston is John.
b. **What** John wrote in Boston is this novel.
c. **The place** where John wrote this novel is Boston.

NB The examples in (30) illustrate only one of several types of pseudo-cleft sentences, but what is stated about this type holds about other types, too.

It is said that a pseudo-cleft sentence can also be used for focalization, but rather than the focalization of the DSynt-Rheme, pseudo-clefting expresses that of the DSynt-Theme (boldfaced in (30)). This is absolutely correct. But in sharp contrast with a cleft sentence, a pseudo-cleft sentence does not feature a particular syntactic construction to express focalization: it is absurd to speak about **pseudo-cleft constructions*—and nobody does.

To see better how the focalization of the Theme in pseudo-clefts is done, let us return for a moment to the sentences in (3). The focalization of the Rheme in them is shown by means of cleft constructions; but what about the focalization of the Theme? It is expressed by the syntactic operations of (linear) Fronting and the introduction of the corresponding pronoun:

- (31) a. *John_{THEME, Foc}, he wrote this novel in Boston.*
b. *This novel_{THEME, Foc}, John wrote it in Boston.*
c. *Boston_{THEME, Foc}, John wrote this novel there.*

These operations do not, of course, impact the starting meaning. But with pseudo-clefts, the picture is completely different—the focalization of the Theme is expressed, so to speak, in a much “deeper” way.

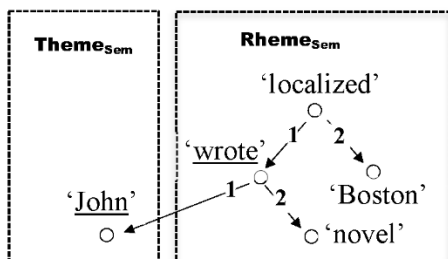


In a pseudo-cleft sentence the focalization of the Theme is expressed through a particular organization of the starting meaning, not by a syntactic construction.

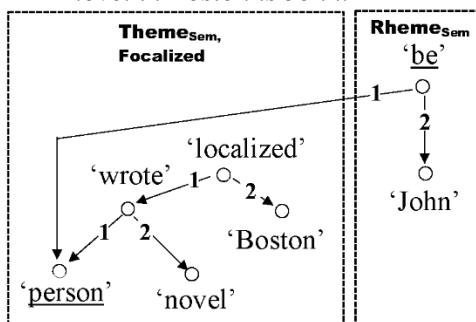
A cleft sentence **has the same propositional meaning** as the corresponding communicatively neutral sentence; yet a pseudo-cleft sentence **does not have the same propositional meaning** as the corresponding communicatively neutral sentence. In other words, the sentences (32a) and (32b) are produced from different semantic structures; this can be seen from the corresponding SemRs:

(32)

a. *John wrote this novel in Boston.*



b. *The person/The one who wrote this novel in Boston is John.*



The semantic difference between (32a) and (32b) is buttressed by the fact that (32b) implies the uniqueness of John ('John, and nobody else'), while (32a) does not carry this implication.

The sentence (32b) has a lot of “semantic” variants: *The person* (<*The guy, The author, The intelligent traveler, ...*> *who wrote this novel...*). Focalization of the semantic Rheme ‘ σ_1 ’ is done by imposing on the expression of ‘ σ_1 ’ a special SSynt-construction: the cleft construction. But focalization of the semantic Theme ‘ σ_2 ’ is done by reformatting this ‘ σ_2 ’ itself into ‘ σ_2 ’ and then expressing ‘ σ_2 ’ by means of standard syntactic rules. As a result, pseudo-cleft sentences should not be considered in syntax as a special case: from a syntactic viewpoint, they are the most usual sentences. In other words, cleft sentences must be described in syntax, and pseudo-cleft sentences, in semantics. It is not for nothing that pseudo-clefts exist in all languages, while clefts exist only in a small number of them.

It is true that sentences (4a) and (30a) are approximately equivalent:

- (33) [(4a)] *It is John who wrote this novel in Boston.* \cong
 [(30a)] *(The one) who wrote this novel in Boston is John.*

But in what sense are they equivalent? Not semantically, since they have different propositional meanings and different Sem-communicative organizations.

They are equivalent in the same sense that the following pairs of sentences are equivalent:

- (34) a. *Two and three make five.* ~ *Adding three to two you obtain five* [$2 + 3 = 5$].
 b. Eng. *Ten to three (o'clock).* ~ Rus. *Bez desjati tri* lit. 'without ten three' [14:50].

The sentences in (33), (34a) and (34b) are **conceptually** equivalent: they carry the same information about extralinguistic reality. However, **linguistically** they are not equivalent. To sum up:

One can, of course, consider pseudo-clefts in parallel with genuine clefts, but it is necessary to make absolutely clear their essential difference.

7. Conclusions

The results of this study can be summed up as follows.

- With respect to general linguistics: a new notion—that of cleft construction—is added to the inventory of formal syntactic notions. This is a contribution to the construction of general syntax. The notion introduced is illustrated with the data of several languages. A brief comparison with pseudo-cleft sentences is offered; it is shown that pseudo-cleft sentences are essentially different from cleft sentences, being particular on the semantic rather than the syntactic level.

- With respect to English studies: the large bulk of factual knowledge about the cleft construction in English collected over the years by numerous researchers is represented as a formal model—by five Meaning-Text type DSyntR \Leftrightarrow SSyntR rules. This is a modest contribution to the task of elaborating a formal description of English syntax in terms of syntactic dependency (see the first attempt in this direction in Mel'čuk & Pertsov 1987).

Abbreviations and Notation

DSyntR	: deep-syntactic representation	R	: a given deep-syntactic relation
DSyntS	: deep-syntactic structure	SemR	: semantic representation
Foc-Cleft	: focalizing cleft	Sem-	: semantic
L	: a given lexical unit	SSyntR	: surface-syntactic representation
Ĺ(P)	: lexeme that is the syntactic head of the phrase P	SSyntRel	: surface-syntactic relation
PREP	: preposition	SSyntS	: surface-syntactic structure
'σ'	: communicatively dominant semanteme	$L_1 \leftarrow \cdots \rightarrow L_2$: lexical units L_1 and L_2 are coreferential
r	: a given surface-syntactic relation	' $L_1 L_2 \dots$ '	: idiom $L_1 L_2 \dots$
		■	: explanation of a notation

Acknowledgments

The previous drafts of this paper were read and criticized by D. Beck, I. Boguslavsky, L. Iordanskaja, I. Mackenzie, S. Marengo, J. Milićević and E. Savvina; the help of D. Beck was absolutely essential. Some finishing touches were put by the anonymous reviewers of the *Russian Journal of Linguistics* and T. Larina. Dear all, it is your constructive proposals that have helped me significantly improve my text; thank you so much!

Appendices

Appendix 1. Some linguistic notions relevant for this paper

communicatively dominant semanteme

The semanteme 'σ₁' in a configuration of semantemes 'σ₁–σ₂' is communicatively dominant if the whole configuration can be reduced to 'σ₁' such that the meaning conveyed is impoverished, but not distorted; the communicative dominance of 'σ₁' is shown by underscoring. Cf.:

'people←1–sing' ⇔ *People sing*. vs. 'people←1–sing' ⇔ *singing people*.

focalization

The Sem-configuration 'σ' is focalized if the Speaker presents it as being logically most prominent for him, that is, as being the focus of his statement. The denotation of 'σ' is singled out as a specific element of a class: 'exactly σ, and not something else'. For instance:

- in *It is John who brought the booze*, the Rheme 'John' is **Focalized** (in *The booze was brought by John* the Rheme 'John' is **Non-focalized**);
- in *John, he brought the booze*, the Theme 'John' is **Focalized** (in *John brought the booze* the Theme 'John' is **Non-focalized**).

grammatical lexical unit

A lexical unit (a lexeme or an idiom) is grammatical if it expresses either an inflectional meaning (e.g., 'after now [future tense]' ⇔ WILL), or a communicative value (e.g., **Focalized** [Theme] ⇔ 'AS FOR'), or else marks a syntactic dependency (e.g., governed prepositions such as TO in *secretary to the Minister*). Grammatical lexical units do not appear in the DSynt-structure and are introduced into the SSynt-structure by grammatical DSyntR ⇔ SSyntR rules.

pseudo-relative clause

A subordinate clause is pseudo-relative if it has the form of a relative clause, but is an actant rather than a modifier—that is, if it is a syntactic equivalent of a noun phrase; e.g., *Listen to **who he likes!*** Pseudo-relative clauses are often called *free* or *headless*. See (Mel'čuk 2021: 235ff.).

relative clause

A subordinate clause is relative if it is a modifier of a lexical element in the super-ordinate clause; e.g., *the things **you gave your life to*** [R. Kipling].

semantic representation

A formal object representing an utterance at the semantic level—a set of four structures: 1) a semantic structure [SemS], which specifies the propositional meaning of the utterance by means of a semantic network; 2) a semantic-communicative structure, which specifies the communicative characteristics of the utterance by means of such communicative values as **Rheme ~ Theme, Given ~ New, Non-focalized ~ Focalized**, etc., associated with particular areas of the SemS; 3) a rhetorical structure, which describes the stylistic/artistic properties of the utterance; and 4) a referential structure, identifying the referents of semantic components in the SemS. See Mel'čuk (2012–2015: vol. 1, 125–128).

source of a pronoun

A lexical unit in the DSyntS that is replaced in the SSyntS and in the sentence by the corresponding pronoun; e.g., in the sentence *John saw his son*, the source of the pronoun *his* is JOHN, seen in the DSynt-configuration JOHN←II–SON.

Appendix 2. Logically possible morphological expressive means and morphological signs

Table 2

Morphological expressive means	Morphological signs	Examples
I. Segmental means segment	1. radical 2. affix	compounding: It. <i>capo+divisione</i> lit. ‘head [of] division’ affixing: <i>book+s, re+read</i>
II. Suprasegmental means prosody	3. suprafix	suprafixing: Ngbaka IMPERF high tone ´ : <i>mī gómǎ</i> ‘I am.cutting’ POSITIVE PERF middle toneˉ : <i>mí gǎmǎ</i> ‘I have.cut’ NEGATIVE PERF low tone ` : <i>mí gǎmǎ gǎ</i> ‘I have.cut not’
III. Operational means • Applicable to signifieds: metasemy1 • Applicable to signifiers: replication1 alternation –in a segment –in prosody • Applicable to syntactics: conversion1	4. metasemy2* 5. replication2 6. segmental apophony 7. suprasegmental apophony 8. conversion2	metasemizing: two1 ‘number 2’ (two plus two) ~ two2 ‘in quantity of 2’ (two books) reduplicating: Ancient Greek PRES <i>tla+ō</i> ‘I.suffer’ ~ PERF <i>tetlāk+a</i> ‘I.have.suffered’ PRES <i>graph+ō</i> ‘I.write’ ~ PERF <i>gegraph+a</i> ‘I.have.written’ vowel substitution: Yiddish SG <i>štat</i> ‘city’ ~ PL <i>štet</i> ‘cities’ SG <i>zun</i> ‘son’ ~ PL <i>zin</i> ‘sons’ stress substitution: <i>adréss</i> _(V) ~ <i>áddress</i> _(N) , <i>expórt</i> _(V) ~ <i>éxport</i> _(N) part of speech substitution: <i>bomb</i> _(N) ~ <i>bomb</i> _(V)

* See (Mel’čuk 2024). A metasemy is a linguistic sign whose signifier is a substitution operation on the signified of the target lexeme. Thus, in the sentence *John saw two excellent Rembrandts* the noun [a] REMBRANDT ‘[a] painting by Rembrandt’ is derived from the proper noun REMBRANDT by one of several productive English metasemies.

References

- Andrade, Aroldo Leal. 2019. Assessing the emergence of reduced clefts in Brazilian Portuguese: Rhetoric structure, information structure and syntax. *Revista de Letras* 99. 101–126. (available from https://www.researchgate.net/publication/336568567_Assessing_the_emergence_of_reduced_clefts_in_Brazilian_Portuguese_Rhetoric_structure_information_structure_and_syntax)
- Collins, Peter. 1991. *Cleft and Pseudocleft Constructions in English*. London–New York: Routledge.
- Davidse, Kristin. 2000. Constructional approach to clefts. *Linguistics* 38 (6). 1101–1131.
- De Cesare, Anna-Maria. 2017. Cleft constructions. In Andreas Dufter & Elisabeth Stark (eds.), *Manual of Romance morphosyntax and syntax*, 536–568. Berlin–Boston: Walter de Gruyter.
- Doetjes, Jenny, Georges Rebuschi & Annie Rialland. 2004. Cleft sentences. In Francis Corblin & Henriëtte de Swart (eds.), *Handbook of French semantics*, 529–552. Stanford, CA: CSLI Publications. (available from <http://fcorblin.free.fr/textes/FrenchSemantics.pdf>)
- Dufter, Andreas. 2009. *Beyond focus marking: Fine-tuning the evolution of cleft types from Latin to Modern French [conference paper]*. (available from https://www.academia.edu/3038036/Beyond_focus_marking_Fine_tuning_the_evolution_of_cleft_types_from_Latin_to_Modern_French)
- Duma, Yegor. 2021. *Structure of clefts in Ukrainian* [preprint]. (available from https://www.researchgate.net/publication/364904045_Structure_of_Clefts_in_Ukrainian)
- Fischer, Klaus. 2009. Cleft sentences: Form, function, and translation. *Journal of Germanic Linguistics* 21 (2). 167–191.
- Gheorghe, Mihaela. 2017. (Pseudo)cleft constructions in old Romanian. *Diacronia* 6. 1–10.
- Hockett, Charles. 1956. Case systems [typescript].
- Ivanova, Svetlana V. & Tatiana V. Larina. 2022. “Meaning ⇔ Text” Theory and the linguistic universe of Igor Mel'čuk. *Russian Journal of Linguistics* 26 (4). 857–880. <https://doi.org/10.22363/2687-0088-32635>
- Jespersen, Otto. 1937. *Analytic Syntax*. London: G. Allen & Unwin.
- Kahane, Sylvain. 2001. A fully lexicalized grammar for French based on Meaning-Text theory. In Alexander Gelbukh (ed.), *Computational linguistics and intelligent text processing, Second International conference, CICLing 2001, Mexico-City, Mexico, February 18–24, 2001, Proceedings [Lecture notes in computer science 2004]*, 18–31. Berlin/Heidelberg: Springer.
- Kimenyi, Alexandre. 1980. *A Relational Grammar of Kinyarwanda*. Berkeley–Los Angeles–London: University of California Press.
- Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39 (3). 463–516.
- Maki, Hideki & Dónall P. Ó Baoill. 2014. *The cleft construction in Irish*. (available from <https://elsj.jp/wp-content/uploads/Irish-Maki-and-O-Baoill.pdf>)
- Malcher Moreno, Kurt. 2023. *Clefts: A Cross-Linguistic Investigation*. Universität zu Köln, PhD dissertation (available from <https://kups.ub.uni-koeln.de/71050/>).
- Mel'čuk, Igor. 1974. *Opyt teorii lingvističeskix modelei “Smysl ⇔ Tekst”*. *Semantika, sintaksis [Outline of a Theory of Meaning-Text Type Linguistic Models. Semantics, Syntax]*. Moskva: Nauka [Reprinted: 1999. Moskva : Škola «Jazyki ruskoj kul'tury».] (In Russ.).
- Mel'čuk, Igor. 1982. *Towards a Language of Linguistics. A System of Formal Notions for Theoretical Morphology*. München: Wilhelm Fink.
- Mel'čuk, Igor. 1988. *Dependency Syntax: Theory and Practice*. Albany, N.Y.: The SUNY Press.

- Mel'čuk, Igor. 1993–2000. *Cours de morphologie générale*. Vols 1–5. Montréal: Les Presses de l'Université de Montréal – Paris: CNRS Éditions.
- Mel'čuk, Igor. 2001. *Communicative Organization in Natural Language. The Semantic-Communicative Structure of Sentences*. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, Igor. 2006. *Aspects of the Theory of Morphology*. Berlin/New York: Mouton de Gruyter.
- Mel'čuk, Igor. 2012–2015. *Semantics: From Meaning to Text*. Vols 1–3. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, Igor. 2016. *Language: From Meaning to Text*. Moskva/Boston: Academic Studies Press.
- Mel'čuk, Igor. 2018. Anna Wierzbicka, semantic decomposition, and the Meaning-Text approach. *Russian Journal of Linguistics* 22 (3). 521–538.
- Mel'čuk, Igor. 2021. *Ten Studies in Dependency Syntax*. Berlin: De Gruyter Mouton.
- Mel'čuk, Igor. 2023. *General Phraseology: Theory and Practice*. Amsterdam/Philadelphia: John Benjamins.
- Mel'čuk, Igor. 2024. A new type of linguistic sign: Metasemy². *Linguisticae Investigationes* 47 (1). 1–29.
- Mel'čuk, Igor & Nikolaj Pertsov. 1987. *Surface Syntax of English. A Formal Model Within the Meaning-Text Framework*. Amsterdam: John Benjamins.
- Paul, Waltraud & John Whitman. 2008. *Shi ... de* focus clefts in Mandarin Chinese. *Linguistic Review* 25 (3–4). 413–451. (available from <https://shs.hal.science/halshs-01574252/file/Cleft%20TLR%202008.pdf>)
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech & Jan Svartvik. 1985/1991. *Comprehensive Grammar of the English Language*. London: Longman.
- Reeve, Matthew. 2012. *Clefts and Their Relatives*. Amsterdam/Philadelphia: John Benjamins.
- Sornicola, Rosanna. 1988. *It*-clefts and *Wh*-clefts: Two awkward sentence types. *Journal of Linguistics* 24 (2). 343–379.
- Stenson, Nancy. 2020. *Modern Irish: A Comprehensive Grammar*. Routledge: London–New York.
- Teng, Shou-hsin. 1979. Remarks on cleft sentences in Chinese. *Journal of Chinese Linguistics* 7 (1). 101–114. (available from <https://www.jstor.org/stable/23753035>)

Article history:

Received: 29 October 2024

Accepted: 25 January 2025

Bionote:

Igor MEL'ČUK is Professor Emeritus at the University of Montreal, Canada. He is the proponent of the Meaning-Text approach to natural language and one of the pioneers of Machine Translation (since 1954). His research interests include general linguistics (with special attention to its conceptual apparatus and terminology), semantics (with special attention to lexicology and lexicography), syntax (with special attention to dependency structures), and morphology. His work is based mostly on Russian and French, but concerns also English and several other languages (Spanish, Hungarian, Lezghian, Alutor, Dyirbal, Bafia, Kirundi, Korean).

e-mail: igor.melcuk@umontreal.ca

<https://orcid.org/0000-0002-4520-0554>

Сведения об авторе:

Игорь Александрович МЕЛЬЧУК – заслуженный профессор Монреальского университета (Канада) и член Королевского общества Канады, является автором лингвистического подхода «Смысл–Текст» и одним из пионеров машинного перевода (с 1954 года). Его научные интересы включают общую лингвистику (в особенности ее понятийный аппарат и терминологию), семантику (в особенности лексикологию и лексикографию), синтаксис (в особенности структуры зависимостей) и морфологию. Работы И.А. Мельчука основываются, в первую очередь, на данных русского и французского языков, но затрагивают также английский и ряд других языков (испанский, венгерский, лезгинский, алуторский, дырбал, бафия, рунди, корейский).

e-mail: igor.melcuk@umontreal.ca

<https://orcid.org/0000-0002-4520-0554>