Clefts. A cross-linguistic investigation

Inaugural-Dissertation
zur Erlangung des Doktorgrades der Philosophischen Fakultät der Universität zu Köln
im Fach Allgemeine Sprachwissenschaft

vorgelegt von
Kurt Malcher Moreno
geb. am 17.08.1971
in Mexiko-Stadt

Defensio 05.10.2021

Gutachter:
Prof. Dr. Nikolaus P. Himmelmann
Dr. Katharina Haude
Prof. Dr. Dejan Matić

Überarbeitete Fassung
09.09.2023

Contents

1 List of figures iv
2 List of tables iv
3 List of abbreviations iv
1 Introduction 1
1.1 Identifying clefts cross-linguistically 4
1.1.1 Major definitions and major types of clefts distinguished in the literature 4
1.1.2 The “cleft” vs. “pseudo-cleft” opposition cross-linguistically 5
1.1.3 Clefts as constructions distinct from equational/specificational sentences 8
1.2 On some basic notions 13
1.2.1 Cleft sentence types 13
1.2.2 Specificational sentences 21
1.2.3 Constructions referred to in the literature as “non-prototypical clefts” 23
1.3 Cross-linguistic literature on clefts 27
1.3.1 Clefting as a universal contrastive-emphasis marking device: Harries-Delisle (1973) 28
1.3.2 Grammaticalization: Heine & Reh (1984), Harris & Campbell (1995), Creissels (2021). 29
1.3.3 “Clefts” vs. “pseudo-clefts”: Lehmann (1984) 31
1.3.4 Cleftability: Luo (2009) 32
1.4 Chapter overview 33
2 Specificational sentences 38
2.1 Notions 38
2.1.1 Specification 38
2.1.2 Note on reference 39
2.1.3 Sentence types 41
2.1.4 Inverse predication approach to specification 43
2.1.5 Quasi specification 44
2.1.6 Specificational relations 46
2.2 Linguistic reflexes of the predicational-equational distinction 48
2.2.1 Formal distinctions between equational and nominal-predicate sentences 48
2.2.2 Equational vs. predicational copulas 51
2.2.2.1 Thai copulas 51
2.2.2.2 Akan/Twi copulas 54
2.3 Linguistic reflexes of specification 57
2.3.1 On the special status of the VARIABLE expression 57
2.3.1.1 Pronominalization patterns (tag questions and resumptive pronouns) 57
2.3.1.2 Hebrew pronominal copulas 59
2.3.1.3 Japanese/Korean coding of the cleft clause 63
2.3.2 Problems in distinguishing clefts from nominal-predicate sentences 66
2.3.2.1 Upper Nicola Okanagan 68

2.3.2.2 Rapa Nui ... 72
2.3.3 On tense-aspect congruence and case marking on the clefted constituent ... 76
2.4 Summary ... 77
3 Oriented clausal nominalizations ... 80
3.1 Basic notions ... 80
3.2 Nominalization strategies ... 81
3.2.1 Zero/minimal marking ... 81
3.2.2 Noun anchoring /support ... 82
3.2.3 Relative pronoun ... 82
3.2.4 Nominal classifier ... 84
3.2.5 Article ... 86
3.2.6 Article/ general subordinator ... 86
3.2.7 Relative marker ... 88
3.3 Nominal verbal forms ... 89
3.3.1 Turkish: subject-oriented and non-subject oriented participles ... 90
3.3.2 Yaqui: nominalization suffixes ... 92
3.3.3 Ingush: contextually-oriented participles ... 97
3.3.4 Tagalog: nominal predication as basic sentence pattern ... 101
3.4 A note on the noun anchoring/support strategy ... 105
3.5 Summary ... 110
4 Word order ... 112
4.1 VARIABLE-VALUE / VALUE-VARIABLE configurations ... 112
4.1.1 Mandarin ... 112
4.1.2 Akan/Twi ... 114
4.1.3 Hausa ... 118
4.1.4 Korean ... 122
4.1.5 Amharic ... 126
4.2 Dislocated, correlative, and correlative-like constructions ... 129
4.2.1 Hindi ... 130
4.2.2 Lithuanian ... 132
4.2.3 Hungarian ... 138
4.2.4 Nahuatl ... 144
4.3 Summary ... 150
5 (Non-pseudo-) clefts ... 153
5.1 Problems ... 153
5.1.1 The form of the cleft clause ... 154
5.1.2 Contrast with so-called presentative/presentational constructions ... 155
5.1.3 The status of the cleft pronoun ... 156
5.2 Two patterns: Germanic and Romance ... 158
5.2.1 The Germanic pattern ... 158

5.2.2 The Romance pattern 160
5.2.3 Cleft(-like) constructions involving complement clauses 163
5.3 English 164
5.3.1 Characteristics of the English *it*-cleft 165
5.3.2 Discontinuous constituent account: Bolinger (1972) 168
5.3.3 Two historical accounts of the development of the English *it*-cleft 170
5.3.3.1 Nominalization strategies in Old English 170
5.3.3.2 On (*h*)*it*+relative clause constructions in the history of English 172
5.3.3.3 On the origin of the English *it*-cleft: Old or Middle English? 173
5.4 French 175
5.4.1 General characteristics of the French *c'est…qu-…* cleft 176
5.4.2 The ce qu-…c'est… cleft 179
5.4.3 The extraposition account for the French *c'est…qu-…* cleft 180
5.4.4 Historical development of the French *c'est…qu-…* cleft 183
5.5 "Clefts" vs. "pseudo-clefts" outside Germanic and Romance 187
5.5.1 Persian 187
5.5.2 Bantu (Kirundi and Makhuwa-Enahara) 189
5.5.3 Santa María Peñoles Mixtec 195
5.6 Summary 199
6 Conclusions 202
7 References 206

1 List of figures

Figure 1: Pseudo-cleft schema according to Lehmann (n.d.) 5
Figure 2: Cleft schema according to Lehmann (n.d.) 5

2 List of tables

Table 1 Count and proportion of cleft clause types in *it*-clefts (Collins 1991:35) 167

3 List of abbreviations

ACC	accusative
ADJ	adjective
ADN	adnominal (modifying)
AL	alienable possession
ALL	allative
AND	andative
ANT	anterior
APPL	applicative
AOR	aorist
ART	article
AS	morphologically adjusted stem
ASP	aspect marker
ASRT	assertive
ASSOC	associative particle
AUG	augment
AV	actor voice
B	B-gender agreement (Ingush)
BEG	begun aspect
C	complementizer prefix
CAUS	causative
CL	(phrase-final) clitic
CLF	classifier
CLM	clause linkage marker
CM	cleft marker/ copula (Kirundi)
COMP	complementizer
COMPL	completive
COND	conditional
CONJ	conjunction
CONNEG	constituent negation
CONT	continuous
COP	copula
CTR	controlled complement clause
CUST	customary/habitual
CV	conveyance voice
CVSIM	simultaneous converb
D	D-gender agreement (Ingush)
DAT	dative
DECL	declarative
DEF	definite
DEI	deictic
DEM	demonstrative
DES	desiderative
DET	determiner
DIR	directional marker
DIR	directive transitivizer (Okanagan)
DIST	distal
DJ	disjoint conjugation
DOM	differential object marker
DP	determiner phrase
DU	dual
DUB	dubitative
DUR	durative
DX	deictic prefix
EMPH	emphatic
EP	epenthetic
ERG	ergative
EXCL	exclusive
EXP	experiential
EZ	ezafe/linker
F	feminine
FOC	focus
FUT	future
GEN	genitive
GENR	general/generic (agreement)
GER	gerund
H	high tone
H	H-series demonstrative (Hebrew)
HON	honorific
HUM	human
I	predicate marker/*particule inclusive*
ICMPL	incompletive
ID	identification particle

IMP	imperative
IMPF	imperfect
INAL	inalienable
INCL	inclusive
INDF	indefinite
INFR	inferential
INS	instrumental
IPFV	imperfective
IRR	irrealis
J	J-gender agreement (Ingush)
L	low tone
LAT	lative
LK	linker
LOC	locative
LV	locative voice
M	masculine
MED	medial
MID	middle (voice) marker
MOD	mood/modality marker
N	neuter
NEG	negation
NEGF	sentence-final negation marker
NEUT	neuter/neutron-passive (Kirundi)
NMLZ	nominalizer
NOM	nominative
NP	noun phrase
NP	noun prefix (Kirundi)
NPST	non-past
NSP	non-specific
NTR	neutral aspect
NUM	numeral marker
NW	non-witnessed tense
OBJ	object
OBL	oblique
OCC	occupation
OP	object prefix (Kirundi)
PASS	passive
PFV	perfective
PFX	prefix
PL	plural
PP	preposition phrase
POSS	possessor/possessive
POT	potential
PP	pronominal prefix (Kirundi)
PRED	predicate marker
PRED	predicative determiner (Rapa Nui)
PREP	preposition
PRF	perfect
PRO	pronoun
PROG	progressive
PROM	prominence marker
PROP	proper (name) article
PROX	proximal
PRS	present
PRST	presentative
PST	past
PTCL	particle
PTCP	participle/participial
PV	patient voice
Q	question marker
RED	reduplication
REL	relative
REM	remote
RFL	reflexive
RLS	realis
RN	relational noun
RPO	reflexive-possessive pronoun
RSTR	restrictive (relative) suffix
SBJ	subject
SG	singular
SP	subject prefix (Kirundi)
SS	secondary stem
STAT	stative
SUB	subordinator
SUBS	subsequent
TAM	tense-aspect-mood/modality
TOP	topic
UV	undergoer voice
V	V-gender agreement (Ingush)
VBZ	verbalizer
VN	verbal noun
VP	verb phrase
WP	witnessed past tense
YNQ	yes/no question
Z	Z-series demonstrative (Hebrew)

1 Introduction

This thesis is an attempt to systematically study the cross-linguistic variation found in the morpho-syntactic structure of cleft constructions. The term 'cleft' is found in many descriptions of languages from all around the world. However, it is often unclear whether the term actually denotes comparable structures. This is reflected by the frequent choice of terms such as 'cleft-like constructions' and 'constructions reminiscent of clefts' in describing the relevant constructions—especially in accounts of less-well studied languages. The present investigation makes a contribution to solving this problem by developing and testing a cross-linguistically applicable definition of cleft constructions.

A first very general or basic definition of clefts can be phrased as follows: Clefts are complex sentences that have the form of specificational sentences. They may be used as pragmatically marked alternatives to simple sentences but are in principle equivalent to these in terms of propositional content.

In his influential account of clefting, Lambrecht (2001) describes clefting as an information-structure-manipulating device that partitions the propositional content of a sentence into "presupposition" and "focus". Lambrecht (2001) further defines presupposition as "the set of propositions lexico-grammatically evoked in a sentence that the speaker assumes the hearer already knows or believes or is ready to take for granted at the time the sentence is uttered" and focus as "the component of a pragmatically structured proposition whereby the pragmatic assertion differs from the presupposition" (p. 474). Pragmatic assertion in turn is defined as "the proposition expressed by a sentence that the speaker expects the hearer to know or believe or take for granted as a result of hearing the utterance" (p. 474). Clefts, in Lambrecht's (2001) terms are "focus-marking devices used to prevent unintended predicate-focus construal of a proposition. Clefts serve to mark as focal an argument that might otherwise be construed as nonfocal, or as nonfocal a predicate that might otherwise be construed as focal, or both" (p. 489). This information structural manipulation is achieved by syntactically partitioning a sentence to (re-)align subject and predicate with presupposition and focus. Consider the following examples, contrasting a basic transitive sentence (1)(a) with sentences clefting the object (b) or the subject (c):

(1) English (constructed)

a. *The boy stole the hammer*.

b. [*What the boy stole*] *was* <u>*the hammer*</u>.

c. [*The one who stole the hammer*] *was* <u>*the boy*</u>.

The cleft sentences in (b) and (c) convey the same propositional content as the simple declarative sentence in (a). Assuming a neutral prosodic realization for the sentence in (1)(a), the information structure of the sentence is such that no single constituent is unambiguously "focal". In the corresponding clefts, the post-copular constituent is construed as focal and the relative clause construction is construed as non-focal. Clefts alternate (and interact) with other means such as accent shift or non-canonical word order to manipulate the information structure of a sentence. What distinguishes clefts from other information-structuring devices is their particular syntactic configuration, which is, as stated above, that of copular sentences.

In terms of its semantico-syntactic configuration, a cleft can be described as a complex sentence consisting of a specificational sentence which involves a subordinate clause (the 'cleft clause') and a nominal expression (the 'clefted constituent') in a subject-predicate-like relation. The cleft clause is a subordinate clause that describes an entity in terms of a state of affairs in which it is involved (i.e. an oriented

nominalization). The clefted constituent specifies the identity of this entity. The relationship between cleft clause and clefted constituent can be formulated in terms of a semantic VARIABLE contained in the cleft clause, the VALUE of which is specified by the clefted constituent (Akamajian 1970: 163). The correspondence of the different levels of analysis is illustrated in (2):

(2) English (constructed)

[*The one who stole the hammer*] *was*	*the boy*.
cleft clause	clefted constituent
presupposition	focus
VARIABLE (x stole the hammer)	VALUE (x = the boy)

In the following I will use interchangeably the terms 'cleft clause' and 'VARIABLE expression (as shorthand for 'the term containing the VARIABLE') on the one hand, and 'clefted constituent' and 'VALUE expression' (as shorthand for 'the term specifying the VALUE') on the other. Note that I will use the terms VARIABLE and VALUE also to refer to the constituents of simple specificational sentences (i.e. specificational sentences in which the VARIABLE expression does not involve a nominalized clause but a basic nominal expression).

It has been suggested that clefts are universally available syntactic constructions (Harris & Campbell 1995: 56). Creissels (2021) suggests that

> [t]o the extent that a language has an equative predication construction and a participant [oriented KM] nominalization construction […] (and it seems difficult to imagine a language that would lack these two types of constructions), it is always possible to combine them to make explicit the exclusive identification of a participant in a presupposed event. (Creissels 2021: 17)

This quote from Creissels (2021) points the way to delimiting the cross-linguistically applicable comparative concept that is necessary for carrying out this investigation. The comparative concept of the cleft that is used in the current work (corresponding to "plain" clefts in the sense of Creissels 2021), is based on two further comparative concepts: 'specificational sentence' and 'oriented (clausal) nominalization'. Both notions are obviously themselves in need of further scrutiny with regard to their cross-linguistic applicability and will therefore be discussed much more extensively in Chapter 2 and 3. The particular properties of the language-specific instantiations of clefts are of course expected to vary along with those of copular sentences and oriented nominalizations but there is also variation specific to clefts.

Thus, a refined definition of clefts can be phrased as follows: A cleft is a specificational sentence in which the VARIABLE expression involves an oriented clausal nominalization.

Many constructions described as clefts in the literature do not correspond to the definition I propose in this investigation because they cannot (to a greater or lesser extent) straightforwardly be described as specificational sentences. It is sometimes difficult to distinguish clefting from other alternative sentence patterns such as those involving non-canonical word order or focus marking by means of specialized particles or inflections. Furthermore, some languages have more than one construction which can be described as a cleft (or as exhibiting cleft-like structural properties). It is often the case that, among these, some correspond less clearly to specificational copular sentences in the language. Consider the following (constructed) English examples below, illustrating respectively (3)(a) a simple declarative sentence, (3)(b) a *wh*-cleft (involving a free relative introduced by the relative pronoun *what*), (3)(c) a *th*-cleft (involving a relative clause headed by the nominal *the thing*), and (d) an *it*-cleft (involving the pronoun *it* and a subordinate clause with the formal properties of a restrictive adnominal relative clause):

(3) English (constructed)

a.	*I saw the ship*.	canonical verbal sentence
b.	[*What I saw*] *was* <u>*the ship*</u>.	*wh*-cleft
c.	[*The thing I saw*] *was* <u>*the ship*</u>.	*th*-cleft
d.	*It was* <u>*the ship*</u> [*that I saw*].	*it*-cleft

The *wh*-cleft in (3)(b) and *th*-cleft in (c) correspond straightforwardly to the definition of clefting proposed in the present investigation. The fact that the *th*-cleft (3)(b) involves a headed relative clause construction may be problematic from the perspective of some accounts of clefting, but from a cross-linguistic perspective the difference between (3)(b) and (c) can be regarded simply as reflecting two different nominalization strategies which, in the case of English, happen to co-exist in the language. In contrast, the status of *it*-cleft (3)(d) as a specificational copular sentence is problematic. The construction consists of a matrix copular clause involving a pronoun (the status of which, as will be discussed further on, is itself controversial) and a subordinate clause (i.e. the expression in square brackets) which does not have (without an antecedent noun) the distribution of a nominal expression. Due to its formal properties, the cleft clause in the *it*-cleft cannot be straightforwardly analyzed as a term of the matrix copular clause (which already has the form of a saturated clause to begin with). Furthermore, cleft clauses in *it*-clefts often resemble (non-oriented) complement clauses in the language, rendering an analysis of *it*-clefts as specificational copular sentences even more problematic.

Note, however, that the term 'cleft' was in fact originally coined (in Jespersen 1969[1937]) to describe exclusively the *it*-cleft (and similar constructions in more or less closely related languages) and was not originally used to refer to constructions such as the *wh*- and *th*-cleft sentences in (3)(b-c) above. As will be detailed in the following section, there are different positions in the literature concerning the use of the term 'cleft' and the way I am using it here is clearly at odds with some of them. Thus, a few further remarks concerning terminology are in order. I use the term 'cleft' to refer to all constructions corresponding to the definition presented above. It is often the case in the literature that constructions more or less closely corresponding to my definition are referred to as "pseudo-clefts" and the term "cleft" is used (contrastively) to refer to constructions which (like the English *it*-cleft) do not. I will avoid using the term "pseudo-cleft" and the term "cleft" in a contrastive sense (meaning "non-pseudo-cleft", as it were) as far as possible. In some cases, however, I will refrain from substituting them and will use quotation marks instead. Substituting the terminology in these cases would often make the discussion of an account unnecessarily cumbersome. Whenever this is the case, it should be clear that "cleft" and "pseudo-cleft" (both used contrastively) are used in the sense of the particular author or philological tradition discussed. This should be clear from the context.

The remainder of this chapter is organized as follows: Section 1.1 is primarily concerned with different definitions of cleft constructions that have been proposed in the literature. Section 1.2 is dedicated to the discussion of some basic notions relevant to the understanding of clefts as defined in this dissertation. In Sub-section 1.2.1, I will discuss different systematizations that have been provided for clefts. In 1.2.2, I further elaborate upon the notion of 'specificational sentence' as understood in this dissertation (to be continued in Chapter 2). In 1.2.3, I turn to a discussion of a number of constructions sometimes described as "clefts" in the literature, but which do not correspond to the definition of clefts adopted in this investigation because they do not exhibit the structure of specificational sentences. Section 1.3 presents a

brief survey of the literature that has approached clefts from a cross-linguistic perspective. In Section 1.4, I present a chapter overview of the dissertation.

1.1 Identifying clefts cross-linguistically

1.1.1 Major definitions and major types of clefts distinguished in the literature

Most definitions of clefts in the literature agree in the main features characteristic of cleft constructions, namely the presence of a matrix copular clause and a subordinate clause describing a state of affairs in which a nominal expression in predicate position (of the matrix clause) refers to a participant (or a circumstant) involved in the state of affairs described in the subordinate clause. Correspondingly, there is some agreement that constructions lacking these features may be functionally close to clefts but should not be described as involving clefting. This is the case for instance with constructions involving non-canonical word order, marked prosodic patterns, or specialized particles or inflectional patterns marking a constituent as the focus of a sentence, but which do not exhibit the structure of a specificational sentence in which the VARIABLE expression involves an oriented clausal nominalization. In many cases, however, it may be difficult to identify clefts and distinguish them from what should rather be described as cleft-like constructions or even constructions which are merely functionally similar to clefts. This is often a problem in the absence of specialized marking of copular (more precisely: equational) relations and/or in the absence of overt marking of subordination (to be discussed in chapters 2 and 3). We will mainly be concerned with constructions whose status as clefts is accepted in most of the relevant literature.

In this section I will address major approaches to clefting found in the literature. These can be summarized as follows:

I Clefts are complex specificational sentences.

IIa There are two categorically distinct clefting construction types: “clefts” and “pseudo-clefts”. “Clefts” may (though need not) be historically related to specificational sentences (and hence to “pseudo-clefts”) but are fundamentally distinct from them. “Pseudo-clefts”, in contrast, can be analyzed as specificational sentences.

IIb Clefts (of any type, including “pseudo-clefts”) are constructions sui generis. They may (though need not) be related historically to specificational sentences but are fundamentally distinct from them.

The approach in I is the one advocated in this dissertation and shall be argued for explicitly or implicitly throughout. I will argue in the following sub-sections that the approaches in IIa and IIb are not adequate from a cross-linguistically perspective.

In Section 1.1.2, I will discuss aspects of approaches assuming a categorical distinction between “clefts” and “pseudo-clefts”. The discussion takes the account of clefting constructions in Lehmann (n.d.) as a point of departure because of its exemplarily systematic and explicit exposition of the subject. In accounts which operate with this terminological opposition, “pseudo-clefts” are constructions that correspond straightforwardly to the definition I propose for clefts in this investigation. “Clefts” proper in these approaches are either constructions which may be (analyzed as) structurally very similar to the English *it*-cleft or—in languages lacking such constructions—to constructions that may exhibit some features characteristic of clefts but that do not transparently correspond to specificational sentences. In my view this approach is problematic mainly because it does not offer a principled way to distinguish clefts from other (functionally) related constructions.

Section 1.1.3 focusses on the highly influential account of clefting in Lambrecht (2001). Lambrecht proposes an approach to clefting that assigns clefts a status of constructions fundamentally distinct from specificational sentences (assuming, however, a historical relationship between the former and the latter in most cases). In his framework for the analysis of cleft constructions, Lambrecht builds on the account of clefting in Jespersen (1969[1937]). Both accounts are primarily concerned with cleft constructions in English. While Jespersen's account deals exclusively with the English *it*-cleft (mentioning comparable constructions in other more or less closely related languages), Lambrecht's approach extends the analysis to "pseudo-clefts". From my point of view, the main problem with Lambrecht's (2001) account of clefting is his view of clefts as being in principle (i.e. by definition) something distinct from specificational sentences. The approach to clefting taken here is the opposite. I consider clefts to be in principle specificational copular constructions, which may develop idiosyncratic patterns.

1.1.2 The "cleft" vs. "pseudo-cleft" opposition cross-linguistically

Lehmann (n.d.) offers a very explicit cross-linguistically oriented characterization of clefting. He describes "clefts" and "pseudo-clefts" as two categorically distinct varieties of clefting constructions. For Lehmann, the crucial difference between "clefts" and "pseudo-clefts" concerns "the syntactic category and function" of the cleft clause. The schematic structure of both constructions is shown in the figures below:

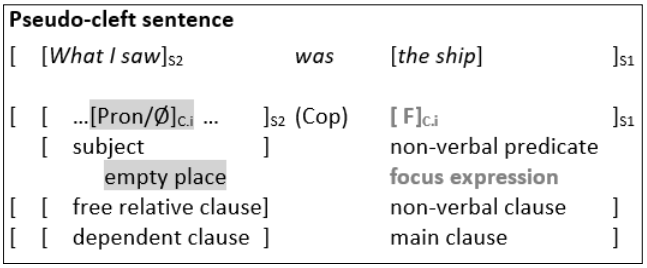

Figure 1: Pseudo-cleft schema according to Lehmann (n.d.)

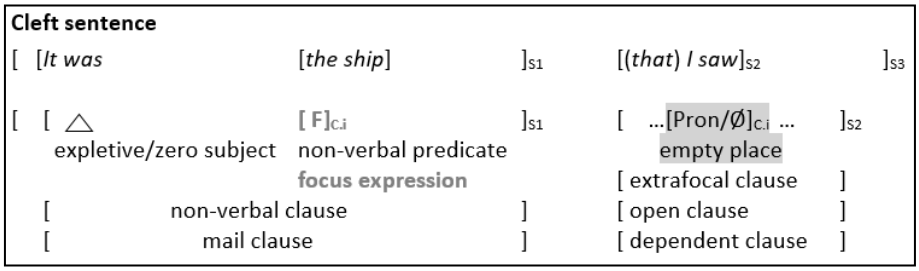

Figure 2: Cleft schema according to Lehmann (n.d.)

Several structural differences can be observed between analyses of the "pseudo-cleft" in Figure 1 and the "cleft" in Figure 2. In the "pseudo-cleft", S2 (the dependent clause) is part of S1 (the non-verbal clause), but in the "cleft", the dependent clause S2 is not a constituent of the main clause S1. Rather, both S1 and S2 are parts of S3, but the relationship between S1 and S2 is not clear. Whatever this relationship is, S2 is not a constituent of S1. The subject position of S1 is taken by an "expletive or zero" subject in the cleft sentence.

Lehmann notes that in the "pseudo-cleft" the dependent clause is an oriented nominalization ("free relative clause"). In Figure 1., Lehmann describes it as a dependent clause "whose empty position […] is coreferential with the predicate of the main clause." He points out that "[c]onsequently, the main predication

states identity of the subject with the predicate nominal." In other words, the "pseudo-cleft" construction is in principle a kind of specificational sentence consisting of two terms denoting the same entity. In the "cleft" construction, by contrast, the dependent clause is, in Lehmann's terms, open but not oriented. The precise wording is: "extrafocal clauses, although open, are not oriented in principle." This means that the "extra-focal clause" is not saturated (i.e. an argument slot determined by the valency of the subordinate predicate is not filled). Lehman remarks that "the extrafocal clause commonly takes the form of a complement clause subordinated by the universal subordinator (that')" and that "the extrafocal clause of a cleft-sentence generally has the same internal structure as an open complement clause" (Lehmann n.d.). Note however, that while the universal subordinator *that* may in principle always be used to introduce the cleft clause in English *it*-clefts, the use of relative pronouns is also possible, as illustrated in example (4).

(4) English (www)
It was <u>me</u> [*who let the dogs out*].

Now the status of the cleft clause [in square brackets] in (4) in relation to the rest of the sentence may be problematic, but I would argue that a relative clause introduced by the pronoun *who* is clearly oriented. Apart from involving a clearly oriented cleft clause, however, the sentence fits quite well in the schema proposed in Figure 2.

Note that in English there are constructions involving clausal arguments (i.e. complement clauses) which may resemble clefts. Consider the following sentences:

(5) English (www)
It was a blessing that she enjoyed good health for most of her life.

The sentence in (5) involves a clausal subject (i.e. a subject complement clause introduced by *that*) and a nominal predicate (*a blessing*). The complement clause is clearly not oriented and the nominal in predicate position does not refer to a participant involved in the state of affairs described in the complement clause. In these constructions, however, the complement clause is necessarily saturated. As will be discussed further in 1.2.1, in cases where the clefted constituent is not an argument but an adjunct, a cleft sentence (of the sort described in Figure 2) may be difficult to distinguish from a sentence involving an extraposed clausal subject (in a language that marks a complement and a relative clause in the same way). A (potential) ambiguity arises precisely because both cases involve (in principle) saturated predications. To the extent that these constructions can be distinguished, the distinction concerns the oriented vs. non-oriented status of the subordinate clause rather than the saturated vs. open status of the subordinate predicate. In this sense, the open/unsaturated status of the cleft sentence suggested in Figure 2 cannot be a defining characteristic of "clefts". Consider the examples below. The sentence in (6)(a) involves an extraposed (non-oriented) complement clause, the one in (7)(a) is an *it*-cleft. Consider the (constructed) paraphrases in (6)(b) and (7)(b).

(6) English (www)
a. *It was a shame that Paul crashed out*.

b. (*The fact*) *that Paul crashed was a shame*.

(7) English (www)
a. *It was last Tuesday that Paul crashed out*.

b. *The day that Paul crashed out was last Tuesday*.

Admittedly, the paraphrase in (7)(b) is not the only possible one. Another possible paraphrase would be something like *Paul's crashing out happened last Tuesday,* clearly involving a non-oriented clausal nominalization and not involving a specificational relation between the terms involved. Clearly, some *it*-clefts are more problematic than others. I will come back to this later. Now, according to my definition, clefting necessarily involves oriented nominalization. A useful notion to consider is that of inherent and contextual orientation in the sense of Shagal (2019). That is, whether the orientation of a nominalization is overtly encoded (as in the case of relative clauses introduced by a relative pronoun such as *who*) or not (as in the case of relatives introduced by the general subordinator *that*). The fact that a cleft clause in an *it*-cleft may be introduced by the universal subordinator *that* (instead of a form such as e.g. *who*) does not automatically make it non-oriented. I will return now to a further problem in Lehmann's (n.d.) account concerning the syntactic category and function of the cleft clause in "clefts" and "pseudo-clefts".

While the structure of the "pseudo-cleft" in Lehmann's (n.d.) account is rather straightforward, that of the (non-pseudo-) "cleft" is much less so. The "extrafocal clause" in a "cleft" is described by Lehmann on the one hand as not having any syntactic function in the main clause. On the other hand, he remarks with respect to its possible syntactic status that "it behaves in some respects as the subject [of the matrix clause, KM] but differs from normal subjects in being obligatorily extraposed." Lehmann does not explain in what way the "extrafocal clause" behaves as the subject of the matrix clause. I assume, as argued above, that the cleft clause involves an oriented nominalization and not a complement clause. This points to a similar relationship between the clefted constituent and the cleft clause in both "clefts" and in "pseudo-clefts". The problem remains, however, that the cleft clause in clefting constructions corresponding to the scheme in Figure 2 often does not have the form of a nominal expression but rather of an adnominal modifier (which, by itself, cannot not function as a constituent of a matrix clause). In fact, this is (in my view) the decisive trait that characterizes this type of constructions.

To summarize, there are two key distinctions between Lehmann's cleft types: First, in the "pseudo-cleft", the dependent (i.e. cleft) clause is a constituent of a non-verbal sentence and can be analyzed as being in a subject-predicate-like relationship with the clefted constituent. In the "cleft", the dependent clause ("extrafocal clause" in Lehmann's terms) is not a constituent of the matrix clause. The matrix clause in a "cleft" involves a nominal predicate and an empty (expletive or null) subject as its terms. Second, the cleft clause in the "pseudo-cleft" is a nominal expression and is co-referential with the nominal expression in predicate position. In a "cleft" sentence this is not so clearly the case.

A cross-linguistically applicable categorical distinction between "clefts" and "pseudo-clefts" in the sense just discussed is, however, problematic. In the first place, it is often difficult to decide whether a construction should belong to one category or the other (or perhaps to neither). Consider the Yucatec Maya sentence in (8), discussed in Lehmann (n.d) as an instance of the type "cleft" in this language.

(8) Yucatec Maya (Lehmann n.d.)

Ma'	*teech*	*in*	*t'an-ik=i'*
NEG	thou	[SBJ.1SG	call-ICMPL]=NEGF

'It's not you I'm calling.'

As Lehmann points out, the requirement of an expletive pronoun and an overt copula is language-specific. Yucatec lacks an obligatory overt copula. Yucatec also lacks a universal subordinator such as *that* in English. Lehmann analyzes the sentence in (8) as a "cleft" consisting of a matrix clause ('it is not you') and an extra-focal cause ('I'm calling'). As the language lacks expletive pronouns, Lehmann argues, "the main

clause of a cleft sentence reduces to the focused constituent". But the "extra-focal" clause is unmarked as well. So, as Lehmann notes, "if all segmental material that codes clefting is zero […] the cleft-sentence may differ from a simple sentence only by the positioning of [the clefted constituent]". In some accounts of similar constructions in Yucatec (see e.g. Verhoeven & Skopeteas 2015; Gutiérrez-Bravo 2017), this construction would not be considered an instance of clefting at all but a sentence involving focus fronting. Now, leaving aside the problem of distinguishing clefting from focus fronting in this language and assuming that (8) does involve clefting, the analysis of the sentence as a "cleft" (and not a "pseudo-cleft") is not without problems either. Yucatec has free relative clauses that do not involve an overt subordinator (cf. Gutiérrez-Bravo 2013: 28). Hence, the sentence in (8) might just as well be described as a "pseudo-cleft" consisting of a nominal expression in sentence-initial position (the canonical predicate position in the language), followed by a (bare) headless relative clause construction as the subject.

I acknowledge that the "cleft" vs. "pseudo-cleft" distinction can be used coherently in languages that have constructions similar to the English *it*-cleft (to be discussed in Chapter 5). In the case of English, the use of the term "pseudo-cleft" (though counter-intuitive from my point of view) has the advantage that it is a well-established cover term used to designate collectively a number of clefting constructions other than the *it*-cleft. I prefer, however, to follow those authors that use the term 'cleft' non-contrastively to designate all kinds of clefting constructions and avoid a terminology which would suggest a cross-linguistically coherent class of constructions exhibiting some but not all the properties of clefts as I propose to define them.

1.1.3 Clefts as constructions distinct from equational/specificational sentences

As was pointed out at the beginning of Section 1.1, there is a non-contrastive use of the term 'cleft' that differs from the one used here and which was characterized as following approach IIb. This approach does not posit a fundamental opposition between "clefts" and "pseudo-clefts" but differs from approach I in that it proposes a fundamental distinction between clefts and specificational copular sentences in a sense which I reject. In some accounts, the distinction in question is pointed out but not further discussed. In other cases, the adherence to this approach to clefts is implicit and is simply reflected in the types of clefting constructions which are discussed and the ones which are ignored. In his elaborate and highly influential account of clefting, Lambrecht (2001) explicitly proposes that clefts should be regarded as constructions sui generis, which may be historically related to specificational copular constructions, but should be regarded as distinct from them. I agree in principle that clefts may differ considerably from basic specificational copular sentences (as in the case of the *it*-cleft in English) and that clefts may undergo changes through historical processes resulting in cleft-like constructions which may have little in common with specificational copular sentences. My contention with respect to Lambrecht's account concerns his insistence on the necessarily constructionalized/grammaticalized status of clefts, which from my point of view makes his framework too restrictive to be applied cross-linguistically (and even, I argue, to the languages it is applied to in his account).

The account of clefting in Lambrecht (2001) concentrates on clefting in English and French and makes some reference to a few other Germanic and Romance languages but the definition of cleft constructions provided is arguably intended to be general enough to be applied cross-linguistically. The definition of a cleft in Lambrecht (2001: 467) is formulated as follows:

> A CLEFT CONSTRUCTION (CC) is a complex sentence structure consisting of a matrix clause headed by a copula and a relative or relative-like clause whose relativized argument is coindexed with the predicative argument of the copula. Taken together, the matrix and the relative express a logically

> simple proposition, which can also be expressed in the form of a single clause without a change in truth conditions. (Lambrecht 2001: 467)

In principle, the definition above is compatible with the view of clefting I propose (with the proviso that the presence of an overt copula should be taken as a language-specific trait). A point not obvious at first glance, which is, however, central in Lambrecht's account, is that a cleft, while involving "a matrix clause headed by a copula", is not a specificational copular clause. That is, Lambrecht (2001) argues that clefts are (or rather, may be) historically related to specificational copular clauses but are to be regarded as (categorically) distinct from these. In Lambrecht's terms, his analysis of clefts

> presupposes a view of the grammatical system in which information structure is a component of sentence grammar, on a par with syntax and semantics. In such a system, pragmatic, semantic, and grammatical relations are mapped onto each other in complex networks of linking relations, sometimes at the expense of semantic compositionality. In some sense […] CCs [cleft constructions] are "jerry-built of available parts used by ancestors for other purposes." In the case of clefts, the ancestor is the copular subject–predicate construction, whose available parts are now used by the grammar for a special purpose, that of focus-marking an argument of another proposition (Lambrecht 2001: 471-472).

I agree that clefts can generally be characterized as information-structure manipulating constructions and that clefts (i.e. particular kinds of clefts in particular languages) may exhibit idiosyncratic properties that set them apart from basic specificational sentences. In particular, I agree that the specificational sentence analysis of English *it*-clefts and similar constructions is problematic. However, I reject the decision to categorically distinguish clefts from specificational copular sentences. The reason is that, in my view, the status of clefts as specificational copular constructions is what makes clefting a cross-linguistically coherent and comparable phenomenon (and not the fact that clefts may develop distinctive properties diachronically).

The previous section discussed a major distinction often made in the literature on clefts, namely that between "pseudo-clefts" and "clefts". An interesting feature of Lambrecht's account is the way he treats this distinction. Going back to Lambrecht's (2001: 467) definition of cleft constructions presented above, note that the (syntactic and categorial) status of the "relative(-like)" clause in the sentence is not specified. This is compatible with the way the cleft clause is characterized in "clefts" (i.e. "non-pseudo-clefts") in Lehmann's (n.d.) account. That is, the cleft clause is not described as a term of the matrix. This is in accordance with Lambrecht's view of a cleft construction as *not* being a "copular subject-predicate construction". Unlike other accounts, however, Lambrecht (2001) does not make a distinction in this respect between "pseudo-clefts" and "clefts", where the former can in principle be described as "copular subject-predicate constructions" and the latter as sui generis focus-marking constructions. Instead, Lambrecht applies a sui generis construction analysis to all clefting constructions. The argumentation runs as follows:

In Lambrecht's (2001) account, all clefting constructions involve semantically empty elements (copulas, expletive pronouns, and complementizers) that serve the purpose of marking the information structural organization of a sentence. This is illustrated in the following scheme, where X stands for a focused constituent and Y for an open sentence (i.e. a clause expressing, in Lambrecht's terms, a proposition containing a variable the value of which is specified by the focused constituent). Lambrecht identifies three kinds of ("argument focus") cleft sentences in English:

(9) English (Lambrecht 2001: 469)

a. Y X
I like champagne. canonical verbal sentence

b. [it] [is] X [that] Y — IT-cleft
It is champagne that I like.

c. [it + that] Y [is] X — WH-cleft
What I like is champagne.

d. X [is] [it + that] Y — reversed WH-cleft
Champagne is what I like.

The elements in square brackets in the examples above are described in Lambrecht's (2001) framework as semantically empty dummy elements whose contribution is to manipulate the information structure of the sentence. In his account, Lambrecht does not use the terms "cleft" and "pseudo-cleft" and points out that these terms stem from "early transformational analyses" (p. 467).[1] He replaces these terms with the terms "IT-clefts" and "WH-clefts" respectively. These designations obviously allude to the forms of elements found in the respective constructions in English but, in Lambrecht's terms, "are to be understood as abstract labels for cross-linguistic formal types" (p. 467). In Lambrecht's account, the difference between "WH-clefts" and "IT-clefts" is that "WH-clefts" involve a composite version of the "empty pronoun" and the "empty complementizer", which are expressed as separate elements in "IT-clefts". Thus, in Lambrecht's account "IT-clefts" and "WH-clefts" are not regarded as two fundamentally different constructions but rather two formally distinct types of clefts. He notes that "the choice of one rather than another of these […] types is determined by various formal and pragmatic factors" (p. 497), but the analysis he makes of them is in principle the same. The fact that the free relative clause *what I like* in sentences such as (9)(b) and (c) has potentially the distribution of a nominal expression and that the subordinate clause *that I like* in (9)(a) does not, is, from the perspective of this analysis, not important.

As noted earlier, Lambrecht's (2001) analysis draws on an account by Jespersen (1969[1937]). I will now briefly discuss Jespersen's (1969[1937]) approach to clefts, as well as that of an earlier account by Jespersen (1949[1927]) to put Lambrecht's analysis into perspective. Unlike Lambrecht's account, both of Jespersen's (1949[1927], 1969[1937]) accounts are concerned with the English *it*-cleft (although references are made to comparable constructions in French, Danish, and a few other languages with similar constructions). In both of Jespersen's accounts, the construction in question is discussed in the context of a general description of relative clauses in English and focuses on the anomalous status of a sentence type involving an element formally identical to an adnominal, restrictive relative clause but differing from a regular relative clause in not restrictively modifying its apparent antecedent. The difference between Jespersen's alternative accounts is very instructive and points to two quite different approaches to *it*-clefts. In his original analysis, Jespersen (1949[1927]: 89) notes that the relative clause in a cleft construction does not "belong" to the constituent preceding it but rather to the pronoun *it* which functions as the subject of the matrix clause. Evidence for the relative clause not "belonging" to its apparent antecedent is found in sentences where the relative clause follows items such as personal names and pronouns, which cannot be modified by restrictive relative clauses. In his second account, Jespersen (1969[1937]) provides the following formalization of his original (1949[1927]) (but subsequently rejected) analysis:

(10) English (Jespersen 1969[1937]: 73)
It is the wife who decides. S* V P 2* (S_2^c V)

[1] The term 'pseudo-cleft' is attributed to Peters & Bach (1968).

In the scheme presented in (10), the pronoun *it* and the relative clause are represented as a discontinuous constituent. The asterisk in Jespersen's notational scheme marks "words standing apart, but belonging together" (p. 8). The sentence as a whole is analyzed as a copular clause consisting of a subject (notated with a capital S), a copular verb (V) and a predicative noun (P). The subject is a discontinuous expression involving the pronoun *it* and a relative clause (*It + who decides*). The subordinate clause thus is described as a restrictive relative clause modifying the pronoun *it*. In this way, Jespersen's original (1949[1927]) account of the English *it*-clefts seeks to explain the anomalous character of the construction by describing it in terms of a more regular one.

In his second account, Jespersen (1969[1937]) notes that "[t]he necessity of finding the best way of symbolizing such sentences has given […] occasion to reconsider the whole question" (p. 74). He proposes to "leave out of the account the question of the origin of such constructions" and focus on "how they are felt now, and how they are accordingly to be analyzed and symbolized" (p. 74). In this alternative analysis, Jespersen (1969[1937]) proposes to "take […] *it is* (*is it*) together with the connective word [subordination marker] (if any such is found) as a kind of extraposition […] and to treat the rest of the sentence as if there had been no intercalation" (p. 76). In Jespersen's (1969[1937]) terminology "extraposition" is the placement of "a word, or a group of words, […] as it were, outside the sentence as if it had nothing to do there" (p. 35).[2] Accordingly, he formalizes an *it*-cleft as follows:

(11) English (Jespersen 1969[1937]: 76; cited in Lambrecht 2001: 465)

a. [*It is*] *the wife* [*that*] *decides.* [SV] S [3^{e}] V

b. [*It is*] *the wife* [*who*] *decides* . [SV] S [S^{e}] V

The scheme in (11) shows an analysis of the *it*-cleft which presents it as a basic sentence to which "extraposed" elements (in square brackets) are intercalated. These elements are described as "lesser subjects" (small cap S), "lesser verbs" (small cap V) and "connectors" (3^{e} and s^{e}). Jespersen (1969[1937]) characterizes the insertion of the "lesser subject" and "lesser verb" *it is* as "a demonstrative gesture to point at one particular part of the sentence to which the attention of the hearer is to be drawn especially" (p. 76). The presence of the "connector" (i.e. *that* and *who* in (11)(a) and (b) respectively) is—as far as I can see—not further elaborated upon.

Jespersen's original (1949[1927]) account approaches clefting (i.e. the English *it*-cleft and similar constructions in other languages) as a kind of copular clause (albeit one with highly idiosyncratic properties). In his second analysis, the *it*-cleft is viewed as a completely sui generis construction. Lambrecht (2001) takes up Jespersen's second proposal and proposes some refinements. The refinements introduced by Lambrecht allow him to account for the fact that the matrix copula in a cleft sentence, though semantically empty in his view, behaves in principle like a regular copular verb in the language. A cleft sentence may be introduced in many—perhaps even the majority—of cases by an invariant *it's* but the copula has in principle the same inflectional possibilities of a regular finite verb in the language. In Lambrecht's (2001) account, the predicate of the cleft clause assigns ("indirectly") a theta role to the clefted element. Simultaneously, the clefted element is an argument of the copula. The contribution of the copula

[2]Note that Jespersen (1969[1937]: 73) refers to the discontinuous-constituent analysis (i.e. his original (1949[1927]) analysis) to as 'transposition'. In several later accounts in the literature, discontinuous-constituent analyses use the term 'extraposition' to refer to the displacement of a relative clause away from its antecedent. Jespersen uses 'extraposition' exclusively in the sense just clarified.

is information structural (and not semantic) and marks its predicate argument as the focus of a sentence. In Lambrecht's terms:

> There is a kind of functional division of labor between the two predicators of a CC [cleft construction]. Both introduce their own syntactic predicate-argument structure, but the type of (nonsyntactic) role they assign to their arguments is different in each case. While the role assigned by the RC [relative clause] predicator is semantic, that assigned by the copula is pragmatic. (Lambrecht 2001: 471)

Lambrecht (2001) repeatedly argues for the status of the copula (together with the "empty" pronoun *it*) as a kind of focus marker but insists on treating the verb *be* as "a regular bivalent predicator" (p. 472) and its object (i.e. the focus constituent) as the predicate of the matrix clause. The difference with respect to a "copular subject-predicate construction" is that in a cleft there is only a dummy subject (an expletive *it*) and hence no subject-predicate structure. The insistence on the semantically empty status of the information-structure modifying elements allows Lambrecht, as will be discussed further on, to include a host of constructions, none of which are based on specificational copular sentences, in his typology of clefts. But perhaps the most interesting move in Lambrecht's account is the extension of his analysis of the *it*-clefts to constructions which (in my view) are clearly specificational copular clauses. Namely, "WH-clefts".

Lambrecht (2001) argues that both "IT-clefts" and "WH-clefts" can be described as structures involving semantically empty elements whose function is to manipulate the information structure of the sentence. In his analysis, both "IT-" and "WH-clefts" involve, beside a semantically empty copula, "dummy" or "empty" pronouns and complementizers. The latter in English are (mainly) *that* and *what* (and *what* is, for Lambrecht, a composite of the "dummy" pronoun *it* and the "dummy complementizer" *that*). Lambrecht admits some extensions such as the relative pronoun *who* (optionally used when a human referent is involved) and *the one(s)* heading relative clauses (e.g. *The one that/who decides is the wife*). He explicitly rejects, however, any elements which may have a lexical content beyond marking a human vs. non-human distinction. In fact, he explicitly argues for the dummy status of *the one* because it cannot be freely used as an argument expression (p. 469). I disagree with these restrictions and argue that any item which can be used as a pro-noun (in the sense of Halliday & Hasan 1976), should qualify as a head of a cleft clause. Consider for instance the following sentence pairs, used by Jespersen (1949[1927]) to illustrate the correspondence between *it*-clefts and propositionally equivalent constructions:

(12) English (Jespersen 1949[1927]: 89)
a. *It is <u>champagne</u>* [*I like best*].

b. *<u>Champagne</u> is* [*what I like best*].

(13) English (Jespersen 1949[1927]: 89)
a. *It is <u>only the milk</u>* [*that is bad*].

b. [*The only thing that is bad*] *is <u>the milk</u>*.

(14) English (Jespersen 1949[1927]: 89)
a. *It was <u>the Colonel</u>* [*I was looking for*].

b. *<u>The Colonel</u> was* [*the man I was looking* for].

In Jespersen's account, the paraphrases in the (b) sentences in the examples (12) to (14) are not intended to be anything but ordinary copular clauses (the term 'cleft' being reserved for *it*-clefts). Only in (12)(b) does the paraphrase of the *it-cleft* involve a free relative clause with the relative pronoun *what* and would qualify

as a WH-cleft in Lambrecht's (2001) terms. In (13) and (14), the relatives are headed by expressions involving more or less general nouns (*the thing, the man*), standing in a hyperonymic relation with respect to the clefted constituent (*the milk,* and *the Colonel,* respectively). Now, following Lambrecht (2001: 469), out of the (b) sentences above, only the sentence in (12)(b) should be considered a cleft. In his account it is very important that the elements involved in marking the cleft clause be "semantically empty dummy elements" and not common nouns. In my view, however, the constructions in (13)(b) and (14)(b) should also be considered clefts. I consider clefts to be in principle specificational copular sentences and, in my view, there is no need for a restriction such as Lambrecht's that would permit only "semantically empty" elements—as long as a specificational relation between the two terms of the copular clause holds. Given a hyperonymy relation (or co-reference) between the item heading the relative clause and the clefted constituent, this is arguably the case and I see no reason to exclude (as a matter of principle) constructions involving lexical nouns heading relative clause constructions. When items such as *the one* are used instead of forms such as *the man,* the denotational restrictions which the latter type would impose are generally recoverable from the discourse context (or happen not to be relevant). Furthermore, some languages lack headless relative clause constructions and cleft clauses (as defined here) can only be formed with lexical nouns or are subject to these restrictions if the relativized argument is not a core argument (e.g. local or temporal adjuncts). Excluding headed relative clause constructions as possible terms in cleft sentences would imply that in these languages clefting is not possible. This is not a problem in Lambrecht's (2001) account because he does not discuss languages exhibiting such restrictions but poses a considerable problem from a cross-linguistic perspective.

Now, Lambrecht's (2001) approach to clefting can be considered on the one hand too restrictive with respect to the approach I follow. On the other hand, however, it extends the notion of clefting to include constructions which under my definition should not be considered clefts. That is, his framework includes constructions which are arguably not directly related (historically or otherwise) to specificational sentences but which resemble in some respects *it*-clefts (and similar constructions in Romance and Germanic). These types of constructions have been discussed in recent literature (e.g. Karssenberg et al. (eds) (2018)) under the rubric "non-prototypical clefts". I will briefly discuss such constructions in Section 1.2.3 later on.

1.2 On some basic notions

1.2.1 Cleft sentence types

As mentioned earlier, there are different types of clefts and terms used to refer to language-specific cleft constructions (e.g. the English *it*-cleft and the French *c'est…qu-* cleft). In distinguishing major types of clefts, the discussion above focused on accounts using the terms "clefts" and "pseudo-clefts" contrastively. It was noted that the use of the term "pseudo-cleft" in much of the literature corresponds quite closely to what I consider canonical clefts. It was further noted that the term is used more or less consistently in this sense. In contrast, the term "cleft" (used in contrast to "pseudo-cleft") is frequently used to refer to constructions similar to the English *it*-cleft. In languages that lack constructions similar to the English *it*-cleft, however, the term "cleft" is used in some accounts to refer to constructions which do not correspond to canonical clefts but that do not resemble English *it*-clefts and similar constructions either. I will briefly discuss some other classifications regularly used in the literature and not mentioned (or only mentioned briefly) up to this point. Note, however, that I will not discuss pragmatic-discursive categories such as the

"informative presupposition" [3] cleft and the like but shall concentrate on two criteria that may be used to classify clefts. The first criterion concerns the formal make-up of the term involving the cleft clause and the second the notional role/function of the clefted constituent. I will begin with formal classifications.

Beside *it-clefts* and "pseudo-clefts", the most widely used formally motivated classification in English distinguishes *wh*-clefts and *th*-clefts. The terms are derived from the forms of the elements introducing the cleft clause. In the case of *th*-clefts, the definite article *the* and in the case of *wh*-clefts, a relative pronoun. The distinction between *wh*- and *th*-clefts basically corresponds to that between free and headed relative clauses. In English, the distribution of *th*- and *wh*- clefts is determined to some extent by a general dispreference for headless relative clauses (see e.g. Collins 1991). Only the wh-pronoun *what* can be freely used in clefts but this form is not normally appropriate for free relatives referring to humans (but see example in f.n. 32 in Chapter 2). In these cases, a relative clause headed by a pronoun or a common noun such as *the one, the person*, or *the girl* is preferred. *It*-clefts, in contrast, are not subject to this particular constraint. Note, however, that the restriction against *wh*- forms other than *what* is not absolute and, in some collocations, they may be freely used. According to Collins (1991), this is the case where the clefted constituent is a demonstrative and precedes the cleft clause as in the following examples.

(15) English (Collins 1991: 29)

a.	?*[*Who took your purse*] *was* <u>*Mary*</u>.	*wh*-cleft
b.	[*The one who took your purse*] *was* <u>*Mary*</u>.	*th*-cleft
c.	<u>*That*</u>*'s* [*who took your purse*].	*wh*-cleft
d.	<u>*That*</u>*'s* [*the one who took your purse*].	*th*-cleft

(16) English (Collins 1991: 29)

a.	?[*How I did it*] *was* <u>*by rotating the knob*</u>.	*wh*-cleft
b.	[*The way I did it*] *was* <u>*by rotating the knob*</u>.	*th*-cleft
c.	<u>*That*</u>*'s* [*how I did it*].	*wh*-cleft
d.	<u>*That*</u>*'s* [*the way I did it*].	*th*-cleft

The examples in (15) and (16) illustrate in a concise manner the relationship between *wh*- and *th*-clefts in English. In principle they are interchangeable but *wh*- forms other than *what* are restricted in their use. On the other hand, *what* is incompatible with human referents but also with clefted adjuncts. In these latter cases, forms such as *when* (time), *where* (location), *how* (manner), or *why* (reason) are semantically appropriate but restricted in their use and a headed relative clause with an appropriate noun must be used in many contexts. Note again that the *it*-cleft is not subject to these restrictions (cf. *It was by rotating the knob*

[3] The term "informative presupposition" cleft is used in in many accounts to refer to sentences in which the cleft clause conveys information which, though presented as presupposed, is new to the hearer. An example follows:

(i) English (Prince 1979: 898)
The leaders of the militant homophile movement in America generally have been young people. It was <u>*they*</u> [*who fought back during a violent police raid on a Greenwich Village bar in 1969*], *an incident from which many gays date the birth of the modern crusade for homosexual rights.*

The fact that somebody fought back during a particular occasion is not assumed to be known by the reader but the writer presents this information as a generally-known fact. Formally the (underlined) sentence in the example above is an *it*-cleft and, with respect to the function/role of the clefted constituent, it is an argument (subject) cleft.

that I did it). As pointed out in the literature (e.g. Prince 1979 and a vast body of subsequent work), however, the discourse contexts under which *it*-clefts and *wh*- or *th*-clefts can be felicitously used may differ considerably.

The distinction between clefts involving "headed" and "headless" relative clause constructions can be made more or less straightforwardly in many languages. As far as I am aware, there is no cross-linguistically oriented shorthand term (comparable to the English terms) to describe clefts involving headless or headed relative clauses. As noted earlier, in some accounts both are indistinctly referred to as "pseudo-clefts" (even if there exist no contrasting "clefts" in the language in question). It was also noted in the discussion at the end of the previous section that in accounts on clefting, clefts involving "headed" relative clause constructions (i.e. corresponding to English *th*-clefts) are often ignored (i.e. simply not discussed) and sometimes even explicitly excluded. As will be discussed in Chapter 3, the formal make-up of the VARIABLE expression in clefts may involve nominalization and subordination strategies different from those typically used in English. But even in English there are some minor types that do not, strictly speaking, correspond to either *th*- or *wh*-clefts. A case in point are *all*-clefts, illustrated in the following example contrasting an *all*-cleft with *wh*- and *th*-clefts (and the corresponding simple sentences):

(17) English (Collins 1991: 32)

a. [*What/the thing the car needs*] *is* <u>*a new battery*</u>. *wh-/th-* cleft

b. [*All the car needs*] *is* <u>*a new battery*</u>. *all*-cleft

c. *The car needs a new battery*.

d. *The car only needs a new battery*.

Collins (1991) describes the difference between cleft sentences such as (17)(a) and (b) in terms of their equivalence to sentences such as (c) and (d) with respect to their propositional content. The *wh*- and *th*-variants in (17)(a) are truth-conditionally equivalent to (c). They can, however, trigger an exhaustivity implicature rendering them pragmatically equivalent to the sentence in (17)(d). That is, cleft variants in (17)(a) may be interpreted as expressing the message that is overtly encoded and entailed (not only implicated) in (17)(d). In the case of the *all*-cleft in (17)(b), the propositional content fully corresponds to that of the sentence in (d). Formally, however, the sentences in (17)(a) and (b) have in common that the relative clause construction has the same potential distribution as a basic nominal expression (unlike the cleft clause in *it*-clefts). Beside *all*-clefts, there are other cleft variants in English which do not (strictly speaking) correspond to *wh*- and *th-clefts.* Consider the following examples.

(18) English (www)
[*Something you really do need*] *is a* <u>set of *digital scales that will weigh oils, lye etc. accurately*</u>.

(19) English (www)
[*A thing you need*] *is* <u>*the attribute action for html forms*</u>.

In principle, the sentences in (18) and (19) correspond (like *all*-clefts) to the same basic configuration as *wh*- and *th*-clefts, although neither involves a *wh*-pronoun nor a definite determiner introducing the specified (VARIABLE) term. As far as I am aware, there is no specialized terminology to describe these types of clefts. But note that such sentences are hardly ever discussed in the literature. As will be discussed in Chapter 3, a classification of clefts based on the formal properties of the term involving the cleft clause depends on the strategies available in a given language to form oriented clausal nominalizations. I may point out to the

reader that, according to my definition of clefts, the presence of a finite verb is not a necessary condition. Consider the following examples:

(20) English (www)
[T*he one who held the gun*] *was* *a fellow by the name of Tele Denton*.

(21) English (www)
[T*he one who was holding the gun*] *was* *Maximo Hernandez alias "Putol"*, *sir*.

(22) English (www)
[*The one holding the gun*] *was* *thirteen-year-old Ian Manuel*.

The sentences in (20) to (22) may be classified as *th*-clefts. However, taking into account the form of the subordinate predicate, the sentence in (22) can be considered to belong to a different type than those in (20) and (21) because the VARIABLE term does not involve a finite clause. Sentences such as (22) are typically not discussed in accounts of clefting in English but they are considered as cleft variants in accounts of clefting in other languages. The exclusion of clefts involving non-finite subordinate forms would be highly problematic from a cross-linguistic perspective, considering the fact that in some languages this is the main (or only) strategy available to form oriented clausal nominalizations.

The second classification criterion found in the literature concerns the notional role/function of the clefted element. Classifications based on this criterion usually distinguish argument clefts, adjunct clefts, and predicate or verb clefts. The bulk of the literature concentrates on argument clefts and for many languages there is little or no information available for other types. Note that the degree to (and the way in) which clefts may differ formally (or not) with respect to the notional status of the clefted constituent in the cleft clause is language specific and is to a considerable extent determined by constraints on forming oriented nominalizations. In Mandarin Chinese, for example, clefts involving direct arguments may be formed with "headless" relative clauses, but adjuncts are reported to require a "head" noun and (optionally) the marking of the function corresponding to the clefted constituent. Consider the sentences in (23) and (24).

(23) Mandarin (adapted from Tsai 2008: 969)

[*Zuotian*	*zai*	*jiaoshi*	*da*	*Xiaodi*	*de*	(*ren*)]	*shi*	*Akiu.*
yesterday	at	classroom	hit	Xiaodi	NMLZ	person	COP	Akiu

'The one/person who beat Xiaodi in the classroom yesterday was Akiu.'

(24) Mandarin (adapted from Tsai 2008: 969)

[*Zuotian*	*Akiu*	*da*	*Xiaodi*	*de*	*(*didian*)]	*shi*	(*zai*)	*jiaoshi.*
yesterday	Akiu	hit	Xiaodi	NMLZ	location	COP	at	classroom

'The location where Akiu beat Xiaodi yesterday was in the classroom.'

This restriction concerns the construal of free relative clauses in the language generally and is not particular to clefts. It should be pointed out that in some languages distinct patterns are required not only for argument and adjunct clefts but for different types of arguments—distinguishing subjects from non-subject arguments or direct from oblique arguments such as goals, beneficiaries, and instruments (which may or may not be treated in the same way as adjuncts). The available information on non-argument clefts for most languages is limited to local and (to a lesser extent) temporal adjuncts. According to the literature, it seems that in English virtually all kinds of adjunct expressions can be clefted, including means and manner (see examples in (16) above), reason/cause (25), and purpose (26). These may involve finite or non-finite subordinate

clauses as clefted constituents. Consider the following examples, described as cleft sentences in the respective sources.

(25) English (Collins 2001: 61)
a. [*the reason it was quiet before* […]] *was* <u>*because the British didn't stir up the* […] *Ulster Protestants*</u>.

b. *It was because the British didn't stir up the Ulster Protestants* [*that it was quiet before*].

(26) English (Huddleston & Pullum 2002: 1418)
It's certainly not <u>*to make life easier for us*</u> [*that they are changing the rules*].

Unfortunately, there is not much information available on constructions comparable to the sentences in the examples in (16), (25), and (26) above for most languages (and indeed the extent to which they are possible at all remains unclear). For English, I note in Chapter 5 that, as argued in Ball (1991, 1994a), the development of adjunct *it*-clefts such as (25)(b) and (26) may be related to the development of the formally similar clausal subject extraposition constructions such as the following:

(27) English (www)
It was soon forgotten that they were ever pronounced.

(28) English (www)
It was not without reason that I went into the temple.

(29) English (www)
It was a tragedy that such a young woman should lose her mobility.

Sentences such as those in (27)-(29), however, differ fundamentally from clefts in that the post-copular term does not specify the reference of the denotatum of an oriented subordinate clause. Rather, the post-copular expression is used to predicate something of a state of affairs described in an extraposed, non-oriented clausal subject. Note that, as pointed out by Pérez Guerra (1998) and Calude (2008), these constructions may be difficult to distinguish from clefts.[4] (I must point out that I find the status of the examples in (25)(b) and (26) above problematic.)

As suggested earlier, clefting is not limited to argument and adjunct expressions but can involve predicates and even whole propositions. Consider the examples in (30) below, illustrating respectively a simple declarative sentence (30)(a) and two corresponding clefts (b) and (c). The sentence in (30)(b) illustrates a construction widely known in the literature as '*do*-cleft'. The construction in (30)(c) is occasionally described as '*happen*-cleft'.

(30) English (Halliday & Matthiessen 2014: 94-95)
a. *The Duke gave the teapot to my aunt*.

b. [*What the duke did with that teapot*] *was* <u>*give it to my aunt*</u>.

c. [*What the Duke did*] *was* <u>*give the teapot to my aunt*</u>.

[4] Calude (2008) suggests a test to distinguish *it*-clefts from (non-oriented) clausal subject extraposition with *it*-insertion. The test consists in paraphrasing the sentence by placing the nominalized clause introduced by *that* in canonical (pre-copular) subject position, replacing the pronoun *it*. If the result is in principle an acceptable sentence (however stylistically awkward), it is not an *it*-cleft.

d. [*What happened*] *was* <u>*that the Duke gave the teapot to my aunt*</u>.

Formally, the *do*- and *happen*-clefts in (30)(b-c) and (d) respectively can be described as *wh*- clefts. Similar sentences can be construed as *th*-clefts (*wh*-cleft sentences with *what* are in principle interchangeable with *th*-cleft sentences with *the thing that*). Note, however, that they are not normally construed as *it*-clefts (Collins 1991). Constructions very similar to English *do*-clefts are attested in a number of genetically and areally unrelated languages. There is not much information available for constructions similar to *happen*-clefts. In English, *do*- and *happen*-clefts have been reported on the basis of corpus studies (see e.g. Weinert & Miller 1996) to be among the most frequently used *wh*-clefts together with *wh*-clefts involving the verb *say* such as the following:

(31) English (www)
[*What he said*] *was* <u>*that he might propose extending Israeli law to Jewish communities in Judea and Samaria*</u>.

Say-clefts (31) can be (more or less) straightforwardly described as argument clefts involving verbs taking clausal complements. This analysis may be perhaps extended to *happen*-clefts (where *happen* would take a clausal subject). In the case of *do*-clefts, the verb *do* can be perhaps best analyzed as a pro-form co-indexed with verbal predicates.

The term term 'predicate cleft' is used in the literature to describe constructions (mainly in African languages) reminiscent *do*-clefts, the cleft status of which is unclear. The pattern in question involves a nominalized (often reduplicated) verbal form preceding a clause involving a finite form of the same verb. Consider the Yoruba (Atlantic-Congo, Ede) and Buli (Atlantic-Congo, Gur) sentences in (32) and (33).

(32) Yoruba (Kandybowicz 2004; cited in Kandybowicz 2009: 140)

a. *Ri-ra ni Olú ra ife*
RED-buy FOC Olu buy cup
'It's BUYING that Olú did to the cup.'

b. *Ri-ra ife ni Olú ra ife*
RED-buy cup FOC Olu buy ife
'It's BUYING A CUP that Olú did.'

(33) Buli (Hiraiwa 2005; cited in Kandybowicz 2009: 140)

a. *Dɛ̄-kā àlī /àtì Àtìm dɛ̀ mángò-kŭ dīēm.*
eat-NMLZ COMP Atim ate mango-DEF yesterday
'It is eating that Atim ate the mango yesterday.'

b. *Mángò-kŭ dɛ̄-kā àlī /àtì Àtìm dɛ̀ dīēm*
mango-DEF eat-NMLZ COMP Atim ate yesterday
'It is eating the mango that Atim ate yesterday.'

The cleft status of the constructions in question may vary from language to language. It could be the case that the Buli sentences in (33) may be described as cleft sentences (corresponding to my definition) as the elements glossed as "complementizers" are used to mark oriented subordinate clauses (denoting human and non-human entities respectively) in the language (see Hiraiwa 2003). The cleft status of the Yoruba sentences in (32) is less clear as there is no overt marking of nominalization in the clause following the nominalized verb. Jones (2006: 144) refers to an analysis in Déchaine (2002: 5) arguing for a zero-marked nominalization and thus justifying a cleft analysis of the construction. But note that not only the verb but

also the object (and not a resumptive pronoun as is more commonly the case cross-linguistically) is repeated in the clause following the focused expressions, making a cleft interpretation more problematic. I do not have any information regarding the form of free relative clauses or comparable nominal expressions in either of these languages. Accounts in the literature of similar constructions in other languages (mostly describing the constructions in question in terms of movement operations) do not provide the kind of information we would need to discuss their status as clefts in my terms.

In accounts of English, a distinction roughly corresponding to that between direct arguments on the one hand and obliques and adjuncts on the other, is often described in terms of the syntactic realization (noun phrases vs. prepositional phrases). In the same spirit, clefts are described as also involving adjectives, adverbs, (nominal forms of) verbs, and clausal constituents. It is crucial to note that whatever the formal category of the clefted constituent (and its notional role/function in the cleft clause), the constituent in question is treated as a nominal expression. Consider the following examples:

(34) English (Bolinger 1972: 88)
a. [*The way he behaved toward her*] *was* <u>*offensively*</u>.

b. [*The way he behaved toward her*] *was offensive*.

In the cleft sentence in (34)(a), the expression in predicate position is formally an adverb but it is not used as an adverb in this context. Rather, it is used as a nominal expression with the same denotation (a manner) as the relative clause construction. Consider now the sentence in (34)(b). In this case, the adjective *offensive* functions as a regular predicative adjective in the language. This is not a cleft, however, but a basic declarative sentence with a predicative adjective attributing a quality to the nominal expression in subject position. The expression *offensively* in (34)(a) does not primarily qualify a manner of behavior but specifies it. Note that in the qualifying sentence in (34)(b), the manner of behavior is qualified (perhaps from the point of view of the person who may have produced this utterance) but the behavior in question might as well be a friendly one (and judged for some reason offensive to the utterer of the sentence). In the cleft sentence in (34)(a), where the adverb *offensively* is used as a specifying term, this possibility is ruled out. I will discuss my understanding of specificational relations further in the following section. For now, it is sufficient to note that the relationship between the VARIABLE and VALUE terms in a cleft (and in specificational copular sentences generally) is equational in the sense of Bolinger (1972).[5] That is, both terms denote the same kind of entity (in the sentence in (34)(a) a manner of behavior) and that both are treated as nominal expressions. If a nominalizing strategy is available for the kind of expression clefted (e.g. in the case of verbal predicates), it is applied. If not, the nominal status of the expression in question is attained simply by coercion. Note that I do not assume nominalization by coercion in clefts to be possible in every language. Consider now the following sentences, illustrating the clefting of an adjectival predicate in (35)(a) and a propositionally equivalent simple sentence in (b).

(35) English (www)
a. [*What he was*] *was* <u>*angry*</u>.

b. *He was angry*.

The cleft sentence in (35)(a) in involves an adjective as a specifying term. But note that the adjective does not function in this sentence as a predicative adjective. In the corresponding simple sentence in (35)(b), the post-copular adjective can be described as predicatively ascribing a property to the subject. In the cleft

[5] Chapter 2 offers further elaboration of the understanding of this notion in this dissertation.

sentence (35)(a), the VARIABLE expression does not denote a person but this person's mood or condition. The adjective *angry* does not qualify this person's mood but specifies the nature of this mood. The contrast can be more clearly appreciated in Spanish, where a distinction is made between the general copula *ser* and the stative/locative copula *estar*. Consider the examples in (36) corresponding to the English sentences in (35).

(36) Spanish (www)

a. [*Lo* *que* *estaba*] *era* *enojado*.
ART.SG.N REL be.STAT.3SG.IPFV be.3SG.IPFV angry.SG.M
'What he was was angry.'

b. *Estaba* *enojado*.
be.STAT.3SG.IPFV angry.SG.M
'He was angry.'

In the simple declarative sentence in (36)(b), only the stative/locative copula may be used as a support element for the predicative adjective *enojado* 'angry'. In the cleft sentence in (36)(a), the stative/locative copula occurs in the cleft clause but the general copula *ser* is required as a support element for the VALUE expression. I interpret this as evidence that the adjective in predicate position in the cleft is not being used as a predicative adjective but as a nominal expression denoting a property (like the free relative in subject position). The adjective is not in any way marked as a nominal (as opposed to adjectival) predicate and I assume that its nominal status is attained (as in the case of the adverb *offensively* in (34)(a) above) simply by coercion. A parallel can be drawn to locative expressions. In clefts, these denote locations but pattern like nominal expressions in predicative position in equational sentences. Consider the following examples. The Spanish sentence in (37) is a cleft and the one in (38) a sentence featuring a locative predicate.

(37) Spanish (www)
[*El* *lugar* *donde* *vivía*] *era* *en* *la* *casa* *de* *ancianos*.
ART.DEF.SG.M place where live.3SG.IPFV be.3SG.IPFV in ART.DEF.SG.F house of elderly.person.PL.M
'The place where he lived was in the home for the elderly.'

(38) Spanish (www)
[*El* *lugar* *donde* *vivía*] *estaba* *en* *un* *sitio* *bastante* *sucio*.
ART.DEF.SG.M place where live.3SG.IPFV be.3SG.IPFV in ART.DEF.SG.M site enough dirty.SG.M
'The place where he lived was in a rather dirty location.'

As noted by Bolinger (1972), sentences such as the English translations of (37) and (38) are potentially ambiguous in English. The Spanish sentences in (37) and (38) are, however, not ambiguous. In English, they permit a cleft reading but also a reading under which the sentence is interpreted as involving a nominal expression with a locative denotatum and a locative predicate. Bolinger refers to these alternative readings as "equational" and "place in a place" respectively. In the first case, the expression in post-copular position is understood as specifying the reference of the expression in subject position which denotes a location, the reference of which must be specified. Under the "place in a place" interpretation a locative predicate is understood as providing information concerning the location of a spatial entity with respect to another. This (very subtle and for the purposes of effective communication in most cases hardly relevant) ambiguity is,

however, present only in languages (like English) in which nominal and locative predicates are marked identically (in the case at hand, by means of the copula *be* as a support element). Another condition for the ambiguity to arise is the possibility (not necessarily given in every language) for clefted constituents to be marked as locative expressions in the first place. Note that in the Spanish cleft sentence in (37) as well as in its English translation, the locative preposition may be omitted because the notional role/function in the cleft sentence (i.e. a location) of the nominal expression *la casa the ancianos* 'the home for the elderly' is clear. Without the locative preposition in the English translation, the sentence would preclude a "place in place" interpretation—which is already difficult (if possible at all) in Spanish, given the use of the general copula *ser*. Spanish and English permit (but not always require) the locative (and more generally, the oblique role/function) marking of the clefted constituent by means of prepositions. As will be discussed in Section 2.3.3 in Chapter 2, and in Chapter 5 (see especially 5.4.4), the conditions with respect to the marking of the notional role/function of the clefted constituent varies from language to language. In some languages (e.g. in Classical Latin, with the exception of locative adverbs) case and/or adpositional marking of the notional role/function in the cleft clause on the clefted constituent itself is virtually always absent. In principle, this would constitute the canonical case if we expect the VALUE expression in a cleft to behave exactly like the VALUE expression in a simple specificational sentence. Given the scarce information on clefting of constituents other than direct arguments in most languages, it is difficult to examine cross-linguistic variation in this respect. I will assume, however, that clefts, as specificational sentences, necessarily involve two terms denoting the same kind of entity and which behave (in principle) as nominal expressions.

1.2.2 Specificational sentences

In the previous sections, clefts were described as specificational copular sentences and a very brief characterization of my understanding of this notion was given at the beginning of this chapter. In this section, I will try to make my understanding of the notion as explicit as possible. (The discussion will continue in more detail in Chapter 2). It is obviously not my intention to discuss the vast literature on the subject but to briefly elaborate on a few points which are necessary for a further discussion of clefts within the context of this investigation. That is, I will concentrate on the features which I consider to be the basic distinctive properties of specificational copular sentences. I agree with accounts in the literature that consider clefts to be necessarily specificational (going back at least to Akamajian 1970). My understanding of why this should be so is the following: I consider clefts to be alternative constructions used to manipulate the information structure by construing a sentence as an equational one. (As will become clearer in Chapter 2, I adopt the view that specificational sentences may be best regarded as a subtype of equational sentences.) The motivation behind the use of oriented nominalizations as subject terms in predicational sentences (which are not equational in the scheme I am adopting) is simply to describe the referent of the subject expression (for whatever reason) in terms of a state of affairs in which the referent is involved. That is, the choice of such a construction can be described simply as a choice of an expression to refer to an entity over another.

The distinction between predicational (also called ascriptive or attributive in some accounts) and specificational sentences can be characterized as follows. In predicational sentences a property is predicated of the subject term (the reference of which is not in question). A specificational sentence is a sentence involving a term that opens up a variable (the VARIABLE expression), the value of which is specified by a VALUE expression. That is, the VALUE expression is used to establish the reference of the VARIABLE expression. Typically, the VARIABLE and VALUE expressions correspond to the subject and predicate of a nominal sentence (or the pre-copular and post-copular terms in a language like English). But notions such

as subject and predicate (or pre- and post-copular) should better kept separate from the notions of VARIABLE and VALUE.

There are many approaches to the analysis of specificational sentences. According to one view, specification can be best described as inverse predication. Simplifying somewhat, in predicational sentences there is a referential subject and a non-referential nominal predicate while in a specificational sentence the referentiality of the terms is reversed. In other accounts, the relation between the VARIABLE and the VALUE is described in terms of hyperonymy relations, generality of denotation, or relative familiarity (for overviews of different positions in the literature see e.g. Keizer 1992; Mikkelsen 2005; Van Praet & Davidse 2015; Davidse & Van Praet 2019; Patten 2016). For the moment, it is sufficient to state that the VALUE expression establishes (specifies) the reference of an entity denoted by the VARIABLE expression.

The specificational vs. predicational distinction is sometimes conflated with another distinction made in the literature, namely the classifying vs. identifying (or categorizing vs. identifying) distinction (for accounts comparing these distinctions and suggestions for their combination see Keizer 1992, Van Praet & Davidse 2015). The main opposition in the classifying vs. identifying distinction concerns the referential status of the expression in predicate position in terms of specificity. Discussions of clefts more often than not involve argument clefts where the VALUE expression (i.e. the clefted constituent) refers to specific individuals. There is a strong affinity between identifying and specificational sentences and a specificational reading may obtain more readily in a copular clause if the VALUE expression refers to a specific individual. But a nominal expression without specific reference does not necessarily imply a predicational interpretation. Consider the following examples of (specificational) clefts, involving respectively specific and non-specific nominal expressions as VALUE expressions:[6]

(39) English (constructed)
a. [*What I'm looking for*] *is* *a* (*particular*) *pencil*.

b. [*What I'm looking for*] *is* *a pencil*. (*Any pencil will do*.)

Whereas *a pencil* in (39)(a) refers to a particular individual object, *a pencil* in (39)(b) refers to any arbitrary object belonging to the class of objects to which the description 'pencil' applies. The sentence in (39)(b), however is not a predicational sentence (in the most salient reading at least). Thus, rather than depending on the referential (specific vs. non-specific) status of the expression in post copular position, what makes a sentence specificational is the relationship between the expressions involved. The point of (39)(b) (again, assuming a specificational reading) is not to inform the addressee that a thing being sought after (the reference of which would not be at stake) has the property of belonging to the class of objects which may be described as pencils (or that have the property of being pencils, perhaps among other properties) but to establish the reference of the VARIABLE expression. The examples in (39) illustrate that the specificational status of a sentence is independent from the specificity status of the predicate term but they are not well-suited to illustrate the contrast between specificational and predicational sentences because a predicational reading is difficult to trigger given the properties of the predicate in the cleft clause (*look for*). Consider, however, the example in (40), which allows more readily both a specificational and a predicational reading.

[6] See Rivero (1975: 40) for a discussion of clefts involving unambiguously (non-)specific nominal expressions in Spanish. In Spanish, the specificity of the referent can be unambiguously determined in nominal expressions modified by restrictive relative clauses by the choice of the indicative or subjunctive mood in the subordinate verb. The author challenges the assumption in Kuno (1970), according to whom clefting does not apply to non-specific nominal expressions.

(40) English (Huddleston & Pullum 2002: 267)
[*What he gave her*] *was* <u>*a worthless piece of jewelry*</u>.

As Huddleston & Pullum (2002: 267) point out, there are two readings of this sentence. The specificational reading arises in a context where the addressee does not know what was given and the point of the sentence is to provide this information. The predicational reading arises where the identity of what was given is not in question and a property is ascribed to the object in question (an evaluative term like *worthless* makes a predicational reading more salient but the ambiguity does not depend on it).

Ambiguity between specificational and predicational readings is also discussed in Lambrecht's (2001) account of clefting. Consider the sentence in (41):

(41) English (Lambrecht 2001: 494)
[*What he bought*] *is* <u>*champagne*</u>.

In Lambrecht's (2001) account, the free relative clause *what he bought* in the predicational reading is described as a "full-fledged referential expression". In the specificational reading, he argues, "it merely refers to some past event performed by the individual in question" (p. 249). This is fully in line with the approach to clefting discussed in 1.1.3, where the status of clefts as specificational sentences is not acknowledged and both the markers of subordination and the copula are viewed as information structural markers. In the approach followed here, however, the free relative clause does not refer to some event but denotes an entity—in the case at hand, the object of a past event of buying. I agree that the VARIABLE expression in a cleft is not a "full-fledged referential expression" but it shares this trait with clearly nominal VARIABLE expressions in simple (non-cleft) specificational sentences of the sort *The winner is John.* The denotation of *a winner* is arguably a person and not an event. Note that, as in the case of the clefts such as *The one who won is John* or *It was John who won*, the sentence *The winner is John* implies an event of winning having taken place.

To recapitulate: Clefts are (by definition) specificational sentences. I take denotational congruence to be a necessary property of specificational sentences. Thus, a sentence with a subject denoting a thing and a predicate denoting a property such as *What he gave her was worthless* is necessarily predicational. Specificational sentences can be described as equational. They must be distinguished, however, from identity statements (see discussion in Chapter 2) but also from predicational sentences involving denotationally congruent terms in subject and predicate position (a nominal expression and a non-referential nominal predicate) in a predicational (attributive) relation. The defining characteristic of specificational sentences is the relationship between the VARIABLE and VALUE terms, which can be described in terms of the communicative intention of the sentence. The point of a specificational sentence is to establish (define) the reference of the VARIABLE expression. In a predicational sentence, by contrast, the reference of the subject term is not in question. The descriptive content provided by the nominal predicate is attributed to an entity whose reference is already established or is taken for granted. I also argued that the relation holding between two terms in a specificational sentence is independent from the specificity status of the VALUE expression. The VALUE expression may (and perhaps typically does) refer to a particular (specific) entity but does not need to.

1.2.3 Constructions referred to in the literature as "non-prototypical clefts"

As mentioned earlier at the end of Section 1.1.3, there is some discussion in the literature involving a number of constructions sometimes described as "non-prototypical clefts". These constructions exhibit formal and

functional similarities to clefts but are not specificational sentences. Such constructions will not be considered in the present investigation beyond a brief discussion in this section. The constructions in question resemble clefts of a particular type, namely, constructions structurally similar to the English *it*-cleft. (Note that the literature on the constructions I will be discussing focuses mainly if not exclusively on constructions in Romance and Germanic languages.) Karssenberg et al. (2018: 6) point out that these constructions have in common with clefts that they can be "unclefted" (i.e. the propositional content they express can be conveyed in a simple clause) and that they involve (like *it*-clefts) relative clauses that are "neither restrictive nor appositive". Karssenberg et al. (2018) distinguish "existential clefts" (42), "possessive clefts" (43), and "perception clefts" (44), which they classify according to the predicate of the matrix clause.

(42) Existential cleft (Karssenberg et al. 2018: 7)

a. French

Il	*y*	*a*	*le*	*telephone*	*qui*	*sonne*
it	there	have:3SG.PRES	the	phone	that	ring:3SG.PRES

'The phone is ringing.'

b. Italian

C'è	*un*	*signore*	*che*	*vuole*	*parlare*	*con*	*te.*
there_be.3SG.PRES	a	man	who	want:3SG.PRES	talk.INF	with	you

'There is a man who wants to talk to you.'

c. English

You are quite right David, it was engineered, seems there's only me and you who can see it.

(43) Possessive cleft (Karssenberg et al. 2018: 7)

I have a friend of mine in the history department who teaches two courses per semester.

(44) Perception cleft (Karssenberg et al. 2018: 7)

a. French

Voilà	*le*	*facteur*	*qui*	*arrive*.
see.there.IMP	the	mailman	who	arrive:3SG.PRES

'There's the mailman coming./Here comes the mailman.'

b. Italian

Ecco	*Maria*	*che*	*arriva*.
see.there.IMP	Maria	who	arrive.3SG.PRES

'There is Maria coming./There comes Maria.'

None of these involve a specificational copular matrix clause but all bear a family resemblance with *it*-clefts. Karssenberg et al. (2018) point out that not all accounts of these constructions agree upon their cleft status. For instance, Cruschina (2018) argues against the cleft status of the so-called presentational *ci* 'there' construction in Italian, which he views as an "extension of the existential construction" rather than a cleft. Beside the formal similarity to *it*-clefts, all of the constructions illustrated in the examples in (42) to (44) involve some information structural manipulation departing from the default structure of a canonical verbal sentence. I will concentrate below on the main information structural configurations identified in Lambrecht (2001: 504f), where "non-prototypical clefts" are contrasted with "standard" clefts.

Lambrecht (2001: 504f) distinguishes two major (information structural) types that contrast with "standard specificational clefts": "non-exhaustive specificational clefts" and "sentence-focus clefts". Among the

former, he discusses constructions which involve the "dummy subject" *there* instead of *it*, corresponding to "existential clefts" in Karssenberg et al.'s (2018) classification (illustrated in the examples in (42) above). The contrast between "standard specificational clefts" (45)(b) and "non-exhaustive specificational clefts (45)(c) is illustrated below. Both are propositionally equivalent to the simple (i.e. non-cleft) sentence in (45)(a).

(45) English (Lambrecht 2001: 505)

a. *He wants to explain the use of clefts*.

b. *It's* <u>*the use of clefts*</u> [*he wants to explain*].

c. *There's the use of clefts he wants to explain*.

Both (45)(b) and (c) are propositionally equivalent to the simple sentence in (45)(a) and both involve an information structural manipulation such that one constituent is placed in focus and the rest of the sentence is backgrounded. Both (45)(b) and (c) can be described (in the intended readings) as conveying the presupposition that there is something somebody wants to explain and that this thing is 'the use of clefts'. The difference consists in that in (45)(b) an exhaustive interpretation obtains while this is not (necessarily) the case in (45)(c). That is, the sentence (45)(b) is interpreted as expressing that the use of clefts is the only thing—to the exclusion of other alternatives—that somebody wants to explain while in (45)(c) there is no implication that other alternatives are excluded. In Lambrecht's terms:

> The use of the ''existential'' subject *there* instead of *it* conveys the notion that among the things capable of specifying the value of the variable there ''exists'' the one denoted by the FP [focus phrase]. While in IT [*it*-] clefts the FP [focus phrase] denotatum is equated with the value of the variable, in *there* clefts it is merely ''located'' within a set of possible values." (Lambrecht 2001: 505).

Similar (non-exhaustive) interpretations may hold for constructions that feature the "empty copula" HAVE instead of BE (i.e. involving the verb *have* or its translational equivalent as in example (43) above, characterized by Karssenberg et al. (2018) as a "possessive cleft"). Beside non-exhaustive focus, constructions featuring HAVE as a "dummy copula" can express sentence focus, as illustrated in (46). Consider the contrast between (46) and (47), which Lambrecht characterizes as non-exhaustive and exhaustive "clefts" respectively.

(46) French (Lambrecht 2001: 506)
Y avait André qui voulait encore de la viande, y avait Bertrand, mais Claude il en vopulait pas.
'André wanted some more meat, Bertrand did, but Claude (he) didn't want any.'

(47) French (Lambrecht 2001: 506)
C'etait pas André qui avait pris le jounal, c'etait Bernard.
'André didn't take the newspaper, Bertrand did.'

The point here is that a sentence such as (46), with the verb *avoir* 'have' as a "dummy copula" does not trigger the implicature that André was the only one from a relevant set of referents involved in a state of affairs described in the subordinate clause introduced by the relative pronoun *qui* 'who'. According to Lambrecht (2001), this is the case in (47), where the "empty copula" *être* 'be' is used.

A "sentence-focus/presentational-eventive" (non-specificational) "cleft" (Lambrecht's sense) is illustrated in (48).

(48) English (Lambrecht 2001: 507)
There's a LINGUIST who wants two explain CLEFTS.

In this case, Lambrecht (2001) argues, a presupposition is absent. Thus, the focus (which Lambrecht characterized as the non-presupposed part of an assertion) and assertion coincide (p. 508).

The use of the term 'cleft' to describe the "non-prototypical cleft" constructions discussed above is fully in line with Lambrecht's general approach to clefts, discussed in Section 1.1.3. Recall that in Lambrecht's (2001) account, cleft constructions may (but need not) be historically related to specificational sentences and that clefts are explicitly described as being fundamentally different from them. Recall also that the cleft pronoun, the copula, and the "complementizer" in clefts are viewed as "semantically empty" "dummy" elements. A cleft can be ultimately described, in Lambrecht's (2001) terms, as a construction in which an argument "has its pragmatic role assigned by one predicator and its semantic role by another" (p. 470). In these terms, the nature of the matrix clause involving the predicator "assigning a pragmatic role" seems to be of little importance as long as the predicator in question is "semantically empty". That is, as long as the predicator does not contribute information which would be absent in a corresponding "unclefted" sentence. In some sense, this can be said of the predicators involved in the constructions identified in Karssenberg et al. (2018) as "existential", "possessive", and "perception" clefts. They coincide to large extent with the predicators found cross-linguistically in "existential" or "presentational" constructions (see Creissels 2014), which express the semantically rather abstract notions of presence or availability. Arguably, the presence or availability of a referent is generally (though perhaps not necessarily in all languages) implied in simple declarative sentences by default. That is, the presence or availability of a participant in a state of affairs is in principle taken for granted (unless questioned or negated) even if this is not explicitly asserted. It is thus to be expected for a cleft-like sentence involving an existential or presentational predication to be propositionally equivalent to its "unclefted" counterpart. Like (specificational) clefts, these constructions manipulate information structure by placing a constituent in predicate position (i.e. as the complement of a predicator). There is, however, a crucial difference between these constructions and clefts. In clefts, the cleft clause can be analyzed as one of the terms of the matrix clause just like any VARIABLE expression in a specificational sentence generally. This is not the case in the cleft-like constructions with existential or presentational matrices. The difference is obvious with respect to constructions such as English *wh-* and *th-*clefts (e.g. *What I like is champagne; The one who gave the teapot to my aunt was the Duke*) but clearly less so with respect to *it*-clefts and related constructions where the status of the cleft clause is problematic (and often does not resemble an oriented clausal nominalization). I acknowledge the similarities between "non-prototypical clefts" and *it*-cefts (and similar constructions). In a way, they can be said to partake (formally) of a general, overarching "cleft schema", in the sense of Patten (2012a: 76). Viewed primarily as information-structure manipulating devices involving means beyond prosody and non-canonical word order, it makes perfect sense to consider them as cleft-like.

As an aside, note that there are constructions with similar information-structural properties as those discussed above but which are construed as specificational sentences and clearly fit my definition of clefts. Firstly, there are clefts that do not need to be associated with exhaustivity implicatures. Consider the examples in (49) and (50).

(49) English (Craig 1971: 306)
[*Something everyone should learn to accept*] *is* <u>*maturity*</u>.

(50) English (Craig 1971: 306)
[*Someone/ a person I admire*] *is* <u>*Joe Cash*</u>.

Unlike the "non-prototypical clefts" discussed earlier, these constructions would indeed differ from more familiar argument-focus clefts only in lacking exhaustivity presupposition. Sentences such as those in (49) and (50), however, are rarely (if ever) discussed in the standard literature on clefts. In fact, these examples stem from an "investigation of syntactic-semantic relationships in the selected writing of students in grades 4-2" (Craig 1971). The reason they are generally ignored in the literature is arguably to do with the restrictive approach to clefts discussed in Section 1.1.3.

As for sentence-focus clefts, *happen*-clefts can arguably be described (in some sense) as involving "sentence focus" (the "presupposition" amounts to the fact that something happened). Consider the example in (51), repeated for convenience from (30).

(51) English (Halliday & Matthiessen 2014: 94)
[*What happened*] *was* <u>*that the duke gave that teapot to my aunt*</u>.

Constructions of this kind are not discussed in Lambrecht (2001). They may be considered "non-prototypical", assuming the prototypical function of clefts "to mark as focal an argument that might otherwise be construed as nonfocal" (Lambrecht 2001: 508). They are reported, however, to occur particularly frequently—perhaps more frequently than information-structurally "prototypical" clefts in English (Weinert & Miller 1996).

1.3 Cross-linguistic literature on clefts

This section provides a brief overview on the comparatively little cross-linguistic research that has been done on clefts. Cross-linguistic here is understood as encompassing data from genetically and areally diverse languages. Most comparative work on clefts deals with more or less closely related languages and, particularly, with (mostly better-described) Germanic and Romance languages. Of course, most if not all accounts of clefting dealing with the phenomenon in single languages (other than major, especially Germanic and Romance languages) involve cross-linguistic comparison in the sense that descriptions of cleft constructions in any language almost invariably involve comparison with English clefts (and to some extent, clefts in other languages such as French).

The studies reviewed below address particular aspects of clefting. They are very disparate in scope, methodology and agenda.[7] The first one discusses the status of clefting as a cross-linguistic focus-marking device. Harries-Delisle (1973) argues that clefts are universal means to mark focus ("contrastive emphasis" in her terms). Three further studies to be discussed in this section deal with clefts as a diachronic source of focus-marking constructions (Heine & Reh 1984: 147-183, Harris & Campbell 1995: 151-168, Creissels 2021). The main point made in these studies is that clefts tend to develop from bi-clausal into monoclausal constructions, becoming progressively less cleft-like. A further account, Lehmann (1984: 358-363), deals with the characterization of clefts cross-linguistically. Lehmann (1984: 358-363) is a section in a typological study of relative clause constructions. It considers clefting from the perspective of its role as a strategy to

[7] To the list of works to be discussed below, one might add Harmann & Veenstra (eds) (2013). This edited volume contains information on clefting in several languages. The individual contributions, however, focus mostly on individual languages. A notable exception is the contribution by Hole & Zimmermann (2013). The authors compare clefting patterns in Burmese, Japanese, and Mandarin. In the case of the latter language, however, only the *shì…de* construction is discussed. This construction does not qualify as a cleft under the definition adopted in this dissertation.

manipulate information structure and argues for a fundamental distinction between "clefts" and "pseudo-clefts" cross-linguistically. Luo (2009) is an investigation about "cleftability", i.e. a study of what constituents can occur as clefted constituents in cleft sentences cross-linguistically. This is the only systematic monograph-length cross-linguistic study on clefts I am aware of. It considers data from more than 50 languages (though much of the discussion concentrates on English and Chinese) but follows a definition of 'clefts' very different from that proposed in the present investigation.

1.3.1 Clefting as a universal contrastive-emphasis marking device: Harries-Delisle (1973)

The first account I shall discuss can be best understood in the context of a then current discussion concerning the derivation of "clefts" from "pseudo-clefts". The paper by Harries-Delisle (1973) takes position in this debate. The author characterizes clefts as constructions marking "contrastive emphasis" involving equational sentences. Harries-Delisle (1973) is one of the very few papers on clefts that considers data from a wider variety of genetically and areally diverse languages.[8] The paper proposes that clefting is a universal phenomenon but also argues that all constructions used to convey "contrastive emphasis"—including cases involving (only) marked word order or prosodic means—are ultimately derivable from cleft sentences by the application of transformational rules such as deletion of a neutral noun heading the relative clause, deletion of the relative clause introducer, deletion of the copula element, and extraposition of the cleft clause. The application of these rules may be optional or obligatory depending on the language and specific configurations.

The notion of clefting in Harries-Delisle (1973) corresponds to that adopted in the present study. Harries-Delisle resolves the problem concerning the "cleft" vs. "pseudo-cleft" opposition (see discussion in 1.1.2) using a set of transformational rules to derive the former from the latter. She also notes that the occurrence of the constructions cross-linguistically suggests the basic status of "pseudo-clefts". Harries-Delisle points out that, in her sample "all languages […] have cleft constructions of the form a) *the one who...is X*, only relatively few languages have a corresponding form like b) *it is X who....* [suggesting that] whenever a language has b), it also has a)" (p. 99).

As already noted, the account in Harries-Delisle (1973) can be situated in the context of the discussion concerning the derivation of cleft sentences and the relationship between different kinds of cleft constructions in English (i.e. *it*-clefts and so-called "pseudo-clefts"). Most of this discussion seems to have focused mainly on English constructions but at the time Harries-Delisle's study was written, it had already been extended to other languages. A prominent account involving a non-European language (and cited in Harries-Delisle 1973) was the one in Takizala (1972), which described cleft (and, more generally "focus") constructions in Kihung'an (aka Hungana-Saamba, Atlatic-Congo, Bantu).[9] Like Harries-Delisle (1973), Takizala (1972) argued for the basic status of clefts as a source for other focus constructions. Takizala (1972) argued that the evidence for the cleft source of other focus constructions is not as clear in English as it is in Kihung'an (where the relation between a number of focus-marking constructions co-existing in the language is more evident), but that the processes through which clefts are derived into other focus constructions might

[8] English, German, Mandarin, Hungarian, Kihung'an (Bantu), Swahili (Bantu), Japanese, Tagalog, Arabic, Javanese, Indonesian, Amharic, Dera (Chadic), Diola-Fogny, Kikuyu (Bantu), Malagasy, Kanuri (Saharan), Hausa (Chadic), Somali, Akan, Karekare (Chadic), Ngizim (Chadic), Bade (Cushitic), Marathi, Lugbara (Nilo-Saharan), Hindi, Tswana (Bantu), Telugu, Vietnamese, Bengali, French, Hebrew, Russian, Navajo (Athabascan), Turkish, Murle (Eastern Sudanic, Nilo-Saharan), Zapotec (Otomanguean), Galla (Cushitic), Temne (Mel, Niger-Congo), in order of appearance. Note that for several of the languages mentioned no examples are discussed.

[9] Takizala's (1972) Kihung'an data are also discussed i.a. in Givón (1979: 246-186; 2001: ch 15)

be analogous. Harries-Delisle (1973) attempts to illustrate that this could be the case in all languages, drawing from a selection of areally and genetically diverse languages for which information was available to illustrate the point. Note that the investigation in Harries-Delisle (1973), like many accounts in the literature, is limited to argument clefting but she explicitly states that this limitation was primarily due to the lack of information on the clefting of other constituents in the relevant sources (p. 86).

1.3.2 Grammaticalization: Heine & Reh (1984), Harris & Campbell (1995), Creissels (2021)

The next three accounts on clefting to be discussed (Heine & Reh 1984: 147-183, Harris & Campbell 1995: 151-168, Creissels (2021)) deal with clefts cross-linguistically focusing on the historical development of cleft constructions. The authors argue that clefts tend to evolve from clearly bi-clausal structures into monoclausal focusing constructions (and in some cases further to pragmatically unmarked sentences).

Heine & Reh's (1984: 147-183) account is a chapter of a monograph on grammaticalization in African languages. Heine & Reh (1984: 147-183) discuss clefting as the historical source of "completive" focus marking in a number of (genetically and areally unrelated) African languages. The authors characterize "completive" focus as a type of focus that "does not involve contrast" but is rather used "to fill the gap in the pragmatic knowledge of the addressee [and] most clearly […] manifests itself in answers to WH-questions [referring] to either (term focus), or predicates (predicate focus), possibly also to entire clauses" (p. 148). In principle, this type of focus is not incompatible with the kind of focus expected in cleft constructions (minus the preoccupation with exhaustivity implicatures common in much of the literature on clefts). The reason Heine & Reh (1984) exclude the expression of "contrastive" focus from the account we are interested in now is that some constructions more clearly expressing contrast in some African languages are realized, they argue, by means not related to clefting (p. 147-148). What they have in mind in this case are constructions where "contrast is marked morphologically by partial reduplication of the verb stem" such as in the case of the Yoruba "predicate cleft" constructions illustrated in examples (32) and (33) in Section 1.2.

Heine & Reh (1984: 147-183) contrast weakly and strongly grammaticalized focus-marking constructions and argue that clefting is pervasively (though not without exceptions) a source for ("completive") focus-marking constructions. Clefting is understood in this account as involving subordination (of some sort) and copular matrix clauses (Heine & Reh 1984: 147). The focus of the discussion is the relationship between focus-marking constructions more or less clearly involving clefting and those involving segmental morphological marking (as opposed phonological alternations, distinctive sentence prosody, or word order permutation). It is argued that constructions involving segmental morphological marking can be generally explained as being historically derived from clefts. Note, however, that the cleft status of the source constructions discussed in this account is not always clear. The question concerning how (and whether) these "cleft" constructions are related to specificational copular sentences is discussed in passing with short references to Givón (1979: 246-48; based on data from Takizala 1972) and Schachter (1973). In the end, the authors merely note the possibility of a diachronic relationship between specificational sentences and the cleft-like constructions they discuss but this question seems to be of no particular importance to their understanding of the notion of clefting (p. 178-179).

The account in Harris & Campbell (1995: 151-168) refers to and builds on the account in Heine & Reh (1984: 147-182) discussed above. In fact, Harris & Campbell (1995) mention Heine & Reh (1984: 149-150) and Givón (1979: 246-8)—the latter, as already mentioned, is a brief discussion based on the Kihung'an

data in Takizala (1972)—as the "only two discussions of the universals of this phenomenon" they are aware of (p. 151). Harris & Campbell's account is (like the account in Heine & Reh's 1984: 147-182) a section in a book on language change and is possibly the most prominent (cross-linguistically oriented) account of the diachronic development of clefts to date. It is often cited as proposing clefting as an "universally available" syntactic construction (p. 56). The account on the development of "highlighting" constructions from clefts as historical source constructions is part of a chapter discussing the tendency of complex clauses to become mono-clausal ("processes that simplify biclausal structures"). The other phenomena examined in that chapter along clefts are quotation constructions (developing from complex sentences involving superordinate verbs of saying as matrix predicates) and clause fusion (i.e. the development of monoclausal constructions involving an auxiliary and a main verb). The part on cleft as a source of focus-marking constructions discusses Breton (Celtic), East Cushitic (the "focus marker" *baa* in Somali), Japanese (the *Kakari-Musubi* construction), interrogative sentences in French and Mingrelian (Kartvelian), and a cleft-like topic-marking construction in Laz (Kartvelian). Harris & Campbell (1995: 152) propose a change process in three stages. In the first stage the cleft construction is clearly biclausal and corresponds to the following schema:

(52) cleft schema (Harris & Campbell 1995: 152)
{NP/PP} copula – subordinate clause

The second and third stage posited by Harris & Campbell (1995: 151-168) correspond to respectively the weakly and strongly grammaticalized focus marking constructions in the account by Heine & Reh (1984: 181). In the third stage, the construction in question is clearly monoclausal. In-between constructions exhibit mixed characteristics. Harris & Campbell (1995: 151-168) use the data of a small number of languages to illustrate their argument but the typological profile of the selected languages is varied and the diachronic processes they exhibit are very different from one another.

Note, however, that the status of the subordinate clause in the clefts of the initial stage of their proposed development path is not discussed at any length. It is thus not clear what position Harris & Campbell (1995: 151-168) take with respect to the discussion concerning the relationship between clefts and specificational sentences (i.e. whether they regard as clefts only those constructions whose status as equational/specificational sentences is more or less clear). The authors make a reference in passing to this issue but note that they do not consider these "controversial aspects of the structure of clefts" as relevant to their discussion (p. 152).

Creissels (2021) builds on Harris & Campbell (1995: 151-168) and elaborates on some points not covered there. In contrast to Harris & Campbell (1995: 151-168), Creissels does not focus on the final (clearly monoclausal) outcome of grammaticalization processes. Rather, he focuses on the intermediate stages in the development from clefts from biclausal to monoclausal constructions. (As mentioned in the beginning of this chapter, the author is very explicit with respect to what may be described as a cleft and proposes a definition identical to that adopted in this dissertation.) Creissel's (2021) account concentrates on cross-linguistic variation of clefting patterns, concentrating on constructions that may be (more or less) convincingly described as clefts, albeit corresponding to basic specificational sentences to different degrees. He proposes the term 'plain cleft' to describe constructions where the correspondence to specificational sentences is complete and 'grammaticalized cleft' to describe constructions deviating from these. The "grammaticalization" patterns discussed in Creissels (2021) are illustrated mainly with examples from a few West African Languages (and the Bantu language Tswana), and from French and Spanish, but the argumentation is intended to apply cross-linguistically. He discusses (like Harris & Campbell 1995) the

(non-)obligatoriness of copulas and subordination markers, as well as the flagging of the clefted element in accordance to its notional role in the cleft sentence (see Section 2.3.3 in Chapter 2). A particularly interesting contribution in Creissels (2021) is the discussion of constructions involving "focus marking" particles (also used as copulas) which, however, do not reflect a bipartite structure in terms of word order. He argues that constructions of this kind may be only indirectly related to clefts. He illustrates this point with the discussion of use of a particle identical to the copula in Ivorian Jola (Mande) where the particle in question is postposed to a "focused" constituent but the basic word order pattern of the sentence remains unaffected (Creissels 2021, Section 5.2). As pointed out earlier Creissels (2021) does not focus on cases that are taken to be representative of strongly grammaticalized systems (i.e. advanced stages of re-analysis) as Harris & Campbell (1995) do. Harris & Campbell (1995: 159f) consider constituent re-ordering as part of the process of language change from mono-clausal to bi-clausal structures (which they argue to be the case e.g. in Somali).

1.3.3 "Clefts" vs. "pseudo-clefts": Lehmann (1984)

Another account of clefting cross-linguistically is found in Lehmann (1984: 358-363), part of which is presented with some modifications in Lehmann's (n.d.) characterization of clefting discussed in Section 1.1.2. Lehmann's (1984: 358-363) account is relatively brief. It is a (sub)section of a chapter in a typological study on relative clauses with respect to functional sentence perspective (information structure). In this chapter, focalization (clefting) is treated alongside with topicalization and rhematization.

The account in Lehmann (1984: 358-363) builds on an opposition between "clefts" and "pseudo-clefts", a distinction he illustrates with a few examples from diverse languages and characterizes to a large extent in terms of word order. In this account, "clefts" are characterized as constructions in which the clefted constituent precedes the cleft clause. Constructions in which the clefted constituent follows the cleft clause are described as "pseudo-clefts". Both are argued to be bipartite/biclausal. Lehmann proposes a functional distinction between "clefts" and "pseudo-clefts", the former conveying "insistence" and the latter "suspense" (Lehmann 1984: 359).[10] This position is revised in Lehmann (n.d.), noting that the relative order between clefted constituent and cleft clause is not to be understood as a definitional property distinguishing the two construction types but pointing out that it does hold as a cross-linguistic tendency. As discussed in section 1.1.2, the fundamental difference between "clefts" and "pseudo-clefts" in Lehmann's later (n.d.) account is to be found in the properties of the cleft clause. Both constructions involve the subordination of a proposition and the focalization of a constituent that involves its predicate status in a matrix copular clause. The difference is that while "pseudo-clefts" clearly exhibit the structure of a copular sentence with the cleft clause and the clefted constituent as its terms, in "clefts" the "extra-focal" clause does not have any syntactic function in the main sentence. There is a noteworthy remark in Lehmann (1984) concerning the status of "pseudo-clefts" not found in Lehmann's later account. Lehmann (1984: 360-361) argues that "pseudo-clefts" are not merely equational sentences but are rather "constructed according to the pattern of equational

[10] "An qualitativ verschiedenen Formen der Fokussierung können wir zwei unterscheiden, das Insistieren und die Suspension. Beim Insistieren wird der Fokus eher an den Anfang des Satzes gestellt […] Bei der Suspension wird der Fokus aufgespart, also möglichst nah an den Schluß des Satzes gestellt. […] Insistieren wird der Sprecher, wenn er glaubt, daß der Hörer Zweifel an der Richtigkeit des in den Fokus gestellten Elements hat, oder auch, wenn er den Restsatz als bekannt voraussetzt und besonders auf die Bedeutung der Fokuskonstituente in diesem Zusammenhang aufmerksam machen will. Eine Äußerung mit Suspension wird der Sprecher verwenden, wenn er an etwas anknüpft, was ins Redeuniversum eingeführt ist, und in den dadurch gegebenen Zusammenhang etwas stellen will, was er für einen Knüller hält, was wie eine Bombe einschlagen soll." (Lehmann 1984: 359)

sentences" (p. 361).[11] Lehmann argues that two kinds of evidence support this claim. First, copular sentences with a subordinate clause as a subject and a nominal (or non-verbal) main predicate need not be focusing/contrastive. Second, he points out that some pseudo-cleft constructions involve "focusing" (i.e. clefting) of constituents which could not function as terms in equational sentences. What he describes are, respectively, predicational copular sentences with a free relative clause as a subject (discussed in Section 1.2.2) and adjunct clefts (Section 1.2.1). With respect to "cleft" (i.e. "non-pseudo-cleft") sentences, Lehmann (1984) argues that cleft clauses in these constructions are in some respects similar to relative clauses (and to other nominalized clauses) but should not be considered identical to them or derived from them—although this may indeed be the case in particular languages (p. 363). Note that—in my view—out of eight examples from diverse languages presented in his account as "clefts" only two (Italian and Persian) would qualify as "clefts" in his later (Lehmann n.d.) account. Four (Nahuatl, Cahuilla (Southern and Northern Uto-Aztecan respectively), Swahili (Bantu), and Latin) could equally well (disregarding word order) be described as "pseudo-clefts". Two of the examples presented in Lehmann (1984: 361) (a Yucatec Maya existential-presentational sentence and a Mandarin Chinese *shì…de* construction) should—considering my definition of clefts—not be described as clefts at all.

1.3.4 Cleftability: Luo (2009)

As mentioned at the beginning of this section, Luo (2009) is the only monograph-length cross-linguistic investigation on clefts I am aware of. It considers data from about 50 genetically and areally diverse languages (although some major discussions rely almost exclusively on English data and others focus mainly on the Mandarin Chinese *shì…de* construction, which does not qualify as a cleft by the criteria followed in the present dissertation).[12] The subject of Luo (2009) is cleftability, that is, the possibility for a constituent to be clefted. Luo (2009) discusses three principles that can be used to account for cleftability cross-linguistically. The author investigates the relationship between cleftability and the well-known relativization accessibility hierarchy (Keenan & Comrie 1977, 79), the "nouniness" principle, and the principle of "thematicity". There is a progression in terms of adequacy. In the end, the "thematicity" hierarchy shown in (53) below (in descending order) is argued to account for cleftability most effectively.

[11] Lehmann discusses the following example:

(i) German (Lehmann 1984: 361)
Wohin ich 'mal möchte, ist nach Mexiko.
'Where I'd like to go some day is to Mexico.'

He argues that the constituent following the copula (a prepositional phrase) would be impossible as a term in a nominal sentence in German and adds: "We can specify that pseudo-clefts are construed *following the pattern of equations* in such a way that a constituent appears in the predicate [position], which can be thought of filling an empty slot in the suspense clause." ("Wir können mithin präzisieren, daß Pseudospaltsätze *nach dem Muster von Gleichsetzungen* konstruiert werden, und zwar derart, daß im Prädikat eine Konstituente erscheint, die als Besetzer der Leerstelle im Suspensionssatz zu denken ist.") (Lehmann 1984: 361; emphasis mine)

[12] There is an extensive literature on the Mandarin *shì…de* construction. It is frequently referred to as a 'cleft'. Almost none of the accounts I am aware of, however, attempt to describe this construction in terms that would correspond to the definition of 'clefts' adopted in this dissertation. A notable exception is Zhan & Traugott (2015), who argue for an historical link between equational sentences and the *shì…de* construction. Note that, unfortunately, I am not able to access the relevant literature on the subject published in Mandarin.

(53) Thematicity hierarchy (adapted from Luo 2009: 118)
subject
adjunct (temporal/locative)
direct object
indirect object
oblique nominal phrase
genitive nominal phrase
adjunct (adverbial other than temporal/locative)
predicative (nominal/adjectival)
verb phrase

Luo (2009) claims that the more (potentially) "thematic" a constituent is, the more likely it is that it can be clefted (p. 124). Drawing from Chafe (1976) and Dik (1980), Luo (2009) defines a theme as "that which specifies a spatial, temporal, or individual framework with respect of which the subsequent predication is presented as relevant" (p. 163). "Thematicity" in turn is the potential of an element to serve as a theme. The farther up a type is in the continuum represented in (53) above, the more languages it can be clefted in, and the more easily it can be clefted in the language in question. Note that this hierarchy—apart from cleftability—is also claimed to account for "thematicity" generally. The claim is based on the frequency of the respective types functioning as "topical themes" in English texts (p. 127).[13]

It is important to note that Luo (2009) operates with a definition of clefts that explicitly excludes so-called "pseudo-clefts". For Luo (2009) a "cleft sentence is a construction in which a particular constituent is marked by means of a non-equative syntactic device and/or morphological device for the purpose of focus, contrast, or emphasis" (p.11). He clarifies that "[an] equative syntactic device is one like the pseudo-cleft construction, where the pre- and post-copular constituents can be reversed without affecting grammaticality. The term 'non-equative' is used here to properly exclude such constructions" (p. 16). It is not at all clear to me why a cross-linguistic account of cleftability should exclude so-called "pseudo-clefts". In the case of Luo's (2009) investigation this decision is (as far as I can see) not discussed at all. Concerning the data discussed in Luo (2009), this decision is problematic even in Luo's own terms, as a good number of the constructions discussed might in fact be considered to involve "pseudo-clefts".

1.4 Chapter overview

The rest of this dissertation is organized as follows: In Chapter 2 I will continue the discussion of the notion of specificational sentences begun in Section 1.2.2 of this chapter. I argue that specificational sentences should be considered as (a type of) equational sentences. The latter are viewed as being fundamentally distinct (though often formally identical) from nominal-predicate sentences. In 2.1 I will discuss basic notions relevant to the understanding of specificational sentences. In Sections 2.2 and 2.3, I will turn to examine cross-linguistic data. This is done in two major steps. In 2.2, I discuss cross-linguistic data which may be viewed as reflecting a distinction between equational and nominal-predicate sentences. The idea that an opposition between equational and. nominal predicate sentences is a useful notion for the description and comparison of linguistic structures is found in the two major comparative studies on non-verbal predication to date, Hengeveld (1992) and Stassen (1997). I will begin the the discussion in 2.2.1, drawing mainly on an account by Dryer (2007), who proposes three major cross-linguistically recurrent characteristics distinguishing equational sentences from nominal predicate ones. The first difference

[13] A lengthy report from the *Toronto Star* and chapters 3-6 of *Alice in Wonderland.*

concerns word order flexibility. Word order tends to be (more) rigid in nominal-predicate sentences and flexible in equational ones. The second difference concerns what I will call 'copularization patterns'. Here, the idea is that the marking of subject-predicate (or subject-predicate-like) relations in equational as nominal predicate patterns may differ. To put it in more concrete terms: In some languages, a copular element may be present (or obligatorily so) in equational sentences but not in nominal-predicate sentences. In other languages, equational and nominal-predicate sentences may involve different copulas.[14] The third cross-linguistically recurring distinction between equational and nominal-predicate sentences concerns the formal difference between nominal predicates and terms in equational sentences. The latter tend to be coded in a similar way as expressions functioning as arguments in a sentence and the former do not. (The relevant difference here frequently boils down to the contrast between the use of bare vs. determiner-marked expressions.). The aforementioned distinctions will be relevant in the discussion of clefts throughout the dissertation. In 2.2.2 I will discuss linguistic data reported to reflect the difference between nominal predicate and equational sentences (including specificational ones, including clefts). Here, I will focus on data from Thai (Tai-Kadai) and Akan/Twi (Atlantic Congo, Kwa), two genetically and areally unrelated languages for which the differences between nominal predicates and equational sentences are reported to be particularly clearly reflected in copularization patterns.

In Section 2.3, I will discuss particular aspects of specificational sentences. I will begin by addressing the problem concerning the referential status of the VARIABLE expression. I argue that specificational sentences may be best understood as a sub-type of equational sentences, where two referential expressions occur in a subject-predicate-like relationship. Unlike other equational sentences (identity-statements in a narrow sense), however, specificational sentences involve an asymmetry regarding the referential status of the terms involved. I regard both the VARIABLE and the VALUE expressions as referential, but only the VALUE expression as fully referential. In Section 2.3.1, I will examine linguistic reflexes of the "weakly referential" status of the VARIABLE expression, beginning with a discussion of data from (major) European languages in 2.3.1.1. The phenomenon reflecting the special referential status of the VARIABLE expression concerns the use of sortally non-congruent pronominal forms in anaphoric configurations (where a pronominal form normally referring to non-humans may refer back to a variable expression with a human referent). In 2.3.1.2, I will turn to an arguably related phenomenon reported in the literature for Modern Hebrew. In this (heavily Europeanized) Semitic language, two sets of demonstratives may be used in a copula-like fashion. The choice between a copula-like element of one or the other set is reported to distinguish specificational sentences from predicational ones (but also from identity statements). The forms from the demonstrative series used for the latter types also function as personal pronouns in the language; the one used in specificational sentences does not. In 2.3.1.3 I will discuss data from Korean and Japanese. In these languages, cleft clauses may also exhibit sortal incongruencies of the type discussed in 2.3.1.1 and 2.3.1.2. A pro-noun "heading" a fully referential oriented clausal nominalization (i.e. a relative clause construction) must, if referring to a human, be a sortally congruent noun. In cleft clauses this is not the case.

In 2.3.2, I will discuss the problem concerning the distinction between specificational and nominal-predicate sentences in languages reported to regularly distinguish (true) nominal predicates from (referential) nominal expressions (in the sense discussed in 2.2.1). The problem here concerns the VALUE expression in cleft (or cleft-like) sentences in which the VARIABLE expression is coded like a (referential) nominal expression but the term in predicate position is not coded like a referential expression. This discussion points to the difficulty involved in distinguishing specificational sentences from (pragmatically equivalent) nominal-

[14] The notion of copula adopted in this dissertation will be clarified in 2.2.2.

predicate sentences on formal grounds. I will focus on accounts on Upper Nicola Okanagan (Northern Interior Salish, Salishan) by Lyon (2013) and Rapa Nui (Austronesian, Eastern Polynesian) by Kieviet (2017). Before concluding, in section 2.3.3, I will briefly discuss the role of tense-aspect and case congruence reported in the literature as a factor distinguishing (specificational) clefts from nominal-predicate sentences involving oriented clausal nominalizations as subjects (sometimes referred to in the literature as "predicational clefts").

Chapter 3 is dedicated to the discussion of the notion of nominalization and the examination of different strategies involved across languages in the formation of oriented clausal nominalizations. After briefly elaborating on the notion of nominalization in 3.1, I sketch up a classification of oriented clausal nominalization strategies in 3.2. The strategies identified are related to better established classifications (e.g. "headless", "light-headed", and "headed" relative clause constructions). The aim of the classification presented here is to provide a more fine-grained overview of possible nominalization strategies on the one hand. On the other, the classification I propose aims at a shift in perspective. Much of the literature on relative clause constructions tends to concentrate on oriented subordinate clauses as modifying constructions forming part of nominal expressions. For the purposes of this dissertation, the relevant issue is the status of these constructions as nominal expressions (which may, on their own, occur as terms in equational sentences). In Sub-sections 3.2.1 to 3.2.7, I will present a number of nominalization strategies not involving (major) changes in the internal structure of the subordinate clause (with respect to that of main-clause patterns in the respective languages). Note that the strategies identified are not (necessarily) mutually exclusive, though I will attempt to illustrate them in something close to a "pure" form. The classification I present must also be viewed as a rough sketch. It is only based on information on a limited number of languages, and it is based on data from languages from which information on clefts is also available. Beside the limitations concerning the sources, it is important to consider that some of the categories proposed may be difficult to delimit (see also Epps (2012) for an account arguing for gradual distinctions between different relative clause construction types). Section 3.3 is dedicated to languages in which participant-oriented clausal nominalization involves nominal verb forms. In 3.3.1 to 3.3.3, I will discuss data from languages in which oriented clausal nominalizations involve specialized ("nominal") forms of the subordinate verb. The discussion focuses on three genetically and areally diverse languages, Turkish, based mainly on the accounts in Kornfilt (1997) and Erguvanlı (1984), Yaqui (Sonoran, Uto-Aztecan), drawing mainly on Álvarez González (2012), and Ingush (Nakh-Daghestanian), as described In Nichols (1991). An interesting contrast may be observed between the two former languages and the latter with respect to the orientation of clausal nominalization. While in Turkish and Yaqui specialized suffixes mark the orientation of the nominalized clause towards particular role/functions, oriented nominalizations in Ingush are not inherently but contextually oriented (in the sense of Shagal 2019). That is, in the case of Ingush, an oriented nominalization may denote/refer to almost any participant involved in the state of affairs described in the subordinate clause. In Section 3.3.4, I will discuss data from Tagalog (Central Philippine, Austronesian). In this language, participant-oriented clausal nominalization does not involve the use of specialized verbal forms. It has been suggested in the literature, however, that the basic sentence constitution pattern in this language (and in Philippine languages generally) follows a nominal predication pattern. Under this view, Tagalog verbs may may be regarded as essentially nominal predicates and the rich voice system as forming a paradigm of inherently oriented participles.

Before concluding Chapter 3, I will discuss in 3.4 the conditions concerning the use of the noun-anchor/support nominalization strategy (introduced in 3.2.2). The discussion in this section is crucial to further clarify and motivate the understanding of the notion of nominalization proposed in this dissertation.

Chapter 4 is dedicated to the discussion of word-order patterns in clefts (and specificational sentences generally). The case studies presented in this chapter are distributed in two main sections. Section 4.1 presents case studies involving a number of genetically diverse languages: Mandarin Chinese, Akan/Twi (Atlantic-Congo, Kwa), Hausa (Afro-asiatic, Chadic), Korean, and Amharic (Afro-Asiatic, Semitic).This section can be regarded as a continuation of Section 2.2 in Chapter 2 and investigates word-order patterns in specificational sentences in contrast to word-order patterns possible in nominal predicate sentences.

In Section 4.2 I discuss cleft constructions in languages for which left- (or right) dislocation is reported to occur regularly in cleft sentences. The discussion in this section will concentrate on three languages for which dislocation is reported to constitute the main pattern available at least in some kinds of cleft constructions. It is difficult in many cases to distinguish dislocated constructions from constructions involving non-canonical word order (see e.g. Matić et al. 2016 for a recent cross-linguistically oriented account). I take the presence of an element which may be described as a resumptive pronoun (in a somewhat loose sense) as the relevant criterion to distinguish dislocated constructions. The languages discussed are Hindi, based on a description of relative clause constructions by Junghare (1996), Lithuanian, drawing on an account by Mikulskas (2016), Hungarian, for which I draw on an account on specificational sentence by Matić (2007) and one on clefting by Hartmann et al (2013), and finally Nahuatl (Uto-Aztecan, Nahuan). The data discussed for the latter language are taken from Launey's (1986) description of Classical Nahuatl (the written language of the 16th and 17th centuries presumably reflecting acrolectal varieties spoken in the area around Mexico City), and from an account on copular sentences in Sasaki (2014) for a present-day variety spoken in Ixquihuican, Mexico. Arguably related patterns are discussed elsewhere in the dissertation. Notably, the Modern Hebrew data discussed in 2.3.1.2 and the French *ce qu-…c'est…* clefting pattern discussed in 5.4.2 in Chapter 5 (compare also, with respect to Nahuatl, the discussion on Upper Nicola Okanagan in 2.3.2.1 in Chapter 2).

In Chapter 5 I take up the discussion (begun in 1.1.2 of the present chapter) concerning the problematic status of constructions for which the term 'cleft' was originally coined (in Jespersen 1969[1937]). Namely, *it*-clefts in English and similar constructions in other languages. The chapter starts in 5.1 with a brief discussion of the properties of these constructions using English data. The aim of this discussion is to briefly recapitulate what precisely makes these constructions problematic as clefts (i.e. with respect to the definition of clefts I adopt in the present dissertation). Section 5.2 follows with an overview of similar constructions in Germanic and Romance. In Sections 5.3 and 5.4 I discuss further the English *it*-cleft and the French *c'est…qu-* cleft. The discussion in these sections focuses on two alternative approaches to an analysis of the structure of the constructions in question. The first approach, which I refer to as the 'discontinuous constituent' analysis, explains the structure of the *it*-cleft by analyzing the cleft pronoun and the cleft clause as a discontinuous relative clause construction. This approach corresponds to Jespersen's (1949[1927]) original analysis, discussed in Section 1.1.3 of the present chapter. Under the second approach, the cleft clause is regarded in principle as an independent nominal expression anticipated by (cataphorically linked with) the cleft pronoun. I refer to this analysis as the 'extraposition' approach. The two approaches will be discussed in association with accounts on the English *it*-cleft and the French *c'est…qu-…* cleft respectively because the different approaches have been associated with the respective constructions in the literature, going (at least) back to the accounts by the Danish scholars Jespersen (1949[1927]), and Sandfeld

(1965[1936]), for English and French respectively.[15] Most of the data discussed in Chapter 5 involves Germanic and Romance languages but constructions (to some extent) similar to the *it*-cleft occur outside these branches of Indo-European as well and, in fact, outside of Indo-European. In 5.5, I will briefly discuss a number of constructions from languages outside of Germanic and Romance which do not straightforwardly conform to the definition adopted in this dissertation. Chapter 6 presents the general conclusions of the dissertation.

[15] It may seem somewhat ironic that the discontinuous constituent analysis should be associated with the English construction and the extraposition analysis with the French one, given that the pronoun *it* does not regularly function as a "head" (or "light head") in relative clause constructions in Modern English (e.g. **it that I see*) and the demonstrative *ce* regularly introduces free relatives in French (e.g. *ce que je vois*).

2 Specificational sentences

In the previous chapter I defined clefts as a particular type of specificational sentences. The purpose of the present chapter is to further clarify my understanding of the notion of specificational sentence and to discuss the distinction between specificational sentences and other kinds of "nominal" sentences. That is, sentences involving two expressions denoting the same kinds of entities in a subject-predicate or subject-predicate-like relation to each other. The chapter is organized as follows. In Section 2.1, I will clarify my understanding of the notion of specification, which boils down to the use of one nominal expression to specify the reference of another. I refer to the expression whose reference is specified as the VARIABLE and the expression used to specify its reference as the VALUE. [16] I follow the view that specificational sentences in principle involve two (referential) expressions and may be considered a particular type of equational sentences. Equational sentences may be distinguished from nominal-predicate sentences (in a narrow sense) in that a nominal predicate (unlike e.g. a VALUE expression in a specificational sentence) is non-referential. Specificational sentences may be distinguished from another type of equational sentences, often referred to as equative, which I refer to as identity statements (mainly to avoid potential confusions the use of terms with a shape as similar as 'equational' and 'equative' could propitiate). The difference between identity statements and specificational sentences is that only in the latter an expression is used to specify the reference of the other, whereas an identity statement is used to clarify that two referential expressions refer to the same entity. In principle, the distinctions just mentioned concern the communicative aim of expressions (not necessarily their form). In some languages, however, different types of sentences are formally distinct, as will be discussed in Section 2.2 and 2.3. The discussion in Section 2.2 concentrates on the distinction between equational sentences and nominal-predicate sentences. Special emphasis will be given to copularization patterns distinguishing the two types, which will be discussed in 2.2.2. In Section 2.3, I will discuss properties of specificational sentences. Sub-section 2.3.1 is dedicated to the discussion of the special referential status of the VARIABLE expression and linguistic reflexes of this status. Though in principle referential, the VARIABLE expression does not behave as a fully referential expression. In 2.3.2, I discuss the problem concerning the distinction between nominal predicate and specificational sentences on formal grounds in languages for which a systematic formal distinction between (non-referential) nominal predicates and (referential) nominal expressions is reported. This problem concerns the VALUE expression in cleft sentences and the problem concerning a formally motivated distinction between clefts and nominal-predicate sentences with oriented nominalizations as subject terms. In Section 2.3.3, I will briefly discuss role of tense-aspect and case congruence in distinguishing clefts from predicational sentences. The chapter concludes with a summary in Section 2.4.

2.1 Notions

2.1.1 Specification

Specificational sentences can be defined as sentences consisting of two nominal expressions in a subject-predicate-like relationship. The notions of subject and predicate, however, are problematic to describe the expressions involved in specificational sentences. I will use the terms VARIABLE expression and VALUE expression. These expressions refer to the same entity and one of them (the VALUE) is used to specify the

[16] For an account on the history of the terms VARIABLE and VALUE see Davidse & Van Praet (2019). The authors note that the use of the terms in (more or less) the sense discussed above is inspired in mathematical logic. A converse use of the terms in made by Halliday (1967), who uses the terms in a sense inspired in a semiotic model of semantics.

reference of the other (the VARIABLE). What makes a sentence specificational is the relationship between both expressions. An oft-quoted example used to illustrate a specificational relationship between two expressions is the following:

(1) English (www)
The winner is John.

In principle, both expressions could be used as arguments in a verbal sentence (e.g. *The winner went home* or *John went home*). The point of the sentence in (1) is to clarify the reference of the pre-copular expression (here, a definite description) by means of the post-copular one (here, a proper name). Note that in my understanding of the notion of specification, both proper names and descriptive nominal expressions can function as either VALUE or VARIABLE. Consider the example in (2). Imagine the sentence is used to clarify who the person being referred to as *John* is.

(2) English (constructed)
John is the man I was telling you about yesterday/ the man who gave you the keys.

In some accounts, such sentences are referred to as "descriptively identifying" (e.g. Declerck 1988). The label "identificational" (Higgins 1979) is sometimes used in a similar sense, though the latter term is often taken to describe sentences involving objects of direct perception (e.g. *That's my cousin Max, and that's my aunt Helga*, pointing at people or pictures of people; see discussion in section 2.3.1). I follow the principle that there is no fundamental difference between a sentence such as that in (2) and what is usually considered to be a specificational sentence (i.e. a sentence where the VARIABLE expression involves descriptive content). The reason I am mentioning this at this point in the discussion is to make it clear from the outset that my understanding of the notion of specification concerns relations between two nominal expressions in a sentence. The kind of expressions involved is important but not definitory.

2.1.2 Note on reference

I characterize a referential nominal expression an expression (which may consist of a single word or a group of words) that can be used as an argument. Such an expression is necessarily (i.e. by definition) referential but its referential status may vary. It may, but need not, refer to a specific entity. Consider the following examples:

(3) English (Davidse & Van Praet 2019: 9)
Last night, I was woken by a cat meowing outside my door.

(4) English (Davidse & Van Praet 2019: 9)
A cat can see in the dark.

(5) English (Davidse & Van Praet 2019: 9)
Grace wants a cat, preferably a Persian.

In the sense I understand the notion of reference, the expression *a cat* is used referentially in all of the sentences in (3), (4), and (5). In (3) it refers to a specific cat. The status of the nominal expression can be described as specific-referential. The speaker may not have seen the cat and may later not be able to recognize the cat referred to and, in this sense, would not know, as it were, exactly what cat she or he is talking about. Nonetheless, a particular cat is being referred to. In (4), the speaker does not refer to a particular cat but, to put it in the terms of Davidse & Van Praet (2019: 10), to a "representative instance of the category". In principle, the speaker is referring to the class of 'cats'. Note that referring to the entity in

the plural would make little difference: *Cats can see in the dark*. This kind of expression can be described as generic-referential. In the sentence in (5), reference is made, in the terms used by Davidse & Van Praet (2019: 10), to an "arbitrary instance of the type 'cat'". In this case, the status of the nominal expression is non-specific-referential. In all three cases the entity being referred to is a discourse referent (one may go on referring to it).

It seems that the referential status of the nominal expression (in the sense just illustrated) depends on the type of predication in which it occurs. For instance, the sentence in (4) allows a non-specific interpretation of the nominal expression *a cat* because it is the object argument of a desiderative verb. The non-specific status is confirmed by the addendum *preferably a Persian.* Without this further remark, the referential status of the nominal expression might be ambiguous (cf. *Grace wants a cat she saw in the animal shelter*). Note that the definite vs. indefinite opposition is orthogonal to the one concerning specificity (cf. *Grace wants the furriest cat she can find, preferably a Persian*). A nominal expression may be thus marked definite if the individual being referred to is necessarily one of a kind, as in the case of a nominal expression involving a superlative adjective.

Now, referential nominal expressions may be used in sentences where they are not arguments in a verbal sentence. This is the case, for instance, in the specificational sentences presented in (1) and (2). I noted that I consider both nominal expressions in a sentence such as *The winner is John* as (referential) nominal expressions. This is problematic, and I will discuss why this is so further on, but first let me point out that some expressions which may be described as nominal are arguably not used referentially. Consider the following example:

(6) English (Davidse & Van Praet 2019: 9)
Mary Stuart was incumbent queen of Scotland and dowager queen of France.

Note that the expressions in post-copular position in (6) could not be used as arguments of a verbal sentence in English. They may be described as nominal predicates. They are not referential nominal expressions as they are not used to refer but to ascribe properties to the referential expression in subject position. An issue that will come up repeatedly is that sentences involving nominal predicates may be often pragmatically equivalent to specificational sentences (arguably, the sentence in (6) could be used to answer the question *Who was Mary Stuart?*). Another related problem I will discuss further on, but which I wish to address right now, is that often there may be little or no formal difference between (non-referential) nominal predicates and (referential) nominal expressions. In English, for instance, most (singular, countable) noun phrases require the use of a determiner. Bare nominal predicates (of the kind illustrated in example (6)) are mostly restricted to expressions denoting "one-member roles" (Davidse & Van Praet 2019: 8), such as *queen, president, director, owner* and so on. Furthermore, they only allow the use of some types of modifiers (excluding qualitative ones). Thus, in sentences such as the following, the use of a determiner is required for the nominal predicate expressions:

(7) English (constructed)
a. *John is a good cook.*

b. *John is the best cook.*

In their likely reading, sentences in (7) are used to assert that John cooks well or that his cooking is better than anybody else's. In this case, the post-copular expressions *a good cook* and *the best cook* may be considered nominal predicates (and not referential nominal expressions). The sentences ascribe to the

referent of the subject expression (here, *John*) the properties associated with the (type of) entities denoted by the nominal predicate. The expressions have the same form as (referential) nominal expressions in the language, but they are not being used as such.

Before beginning with the discussion of sentence types in the following subsection, a brief clarification is in order: I have been discussing my understanding of the notion of reference in terms of reference to individuals and the examples in (1)-(7) all involve countable common nouns or proper names. But nominal expressions can of course be used to refer to many kinds of entities. I will continue to discuss relations between nominal expressions using examples where these (potentially) refer to individual persons for ease of exposition. But note that when I argue that specificational sentences involve a relationship between two referential expressions I do not imply that these must necessarily refer to persons or physical objects.

2.1.3 Sentence types

Specificational sentences are often discussed in the literature in the context of other sentence types involving two nominal expressions (or a referential nominal expression and a nominal predicate) in a subject-predicate-like configuration. In English, these sentences can be all described as copular sentences. (All involve two expressions flanking the copular verb *be*.) There are many classifications of copular sentences (and arguments to expand, or reduce the taxonomies; see e.g. discussions in Keizer 1992; Mikkelsen 2005). A commonly used classification consists of three types. This is the classification I follow. It is illustrated in the following examples, using oft-cited examples from the literature:

(8) English
Mary is a linguist. (predicational)

(9) English
The winner is John. (specificational)

(10) English
Cicero is Tully. (equative/identity statement)

I will begin by discussing the sentences of the types illustrated in (8) and (10). In the predicational sentence in (8), the pre-copular expression is a proper name and may be said to refer to a particular individual. The expression in post-copular position may be described as a nominal predicate, which is not used to refer. It has, however, the form of an expression which could in principle be used referentially (e.g. *A linguist was looking for you.*). The sentence in (10) involves proper names in pre- and post-copular positions. Both may be described as referential. The point of the sentence is to assert that both expressions refer to the same entity.

Now, there are many ways to characterize the status of the post-copular expressions in (8) and (10). Grossly simplifying different accounts found in the literature: The post-copular expression *a linguist* in (8) could be said not to refer to (or denote) an individual but rather to a class (i.e. the class 'linguist') and the relation between the pre- and post-copular expressions may be described as class-membership or class-inclusion (i.e. Mary is a member of the class 'linguist'). Or, the relation between the subject expression and the nominal predicate may be described as property-ascription. A person called Mary is then ascribed the property of being a linguist (i.e. somewhat as in the sentence *Mary is tall*, where the idea of ascribing

membership status to Mary in the class of tall entities would be less attractive).[17] Likewise, it is in principle possible to say with respect to the relation between the expressions involved in (10) that the referent of the pre-copular expression (Cicero) is assigned membership in the (one-member) class (of entities called) 'Tully' or that Cicero has the property of being Tully (or, is assigned all the properties associated with the entity denoted by the proper name). I will follow the position that the post-copular expression in (8) on the one hand, and (9) and (10) on the other, may be conceived of as having a fundamentally different referential status. In (8), the post-copular expression is used to ascribe the properties associated with the denotation of the noun *linguist* but an individual (particular or otherwise) is not being referred to. In the sentences in (9) and (10), the post copular expressions *John* and *Tully* refer to a particular person (much like the pre-copular expression *Cicero* in (10) does).

The status of pre-copular expressions in specificational sentences (such as *the winner* in (9)) is much less agreed upon in the literature. It is argued in some accounts in the literature that a specificational sentence such as that in (9) can be more or less regarded as involving in a relation between two expressions such as the predicational one in (8), but reversed. The basic idea is that copular sentences (excluding perhaps equative/ identity statement sentences) generally involve two expressions with different referential status. One is a referential expression and the other one is a predicate. Or, one expression is individual-denoting and the other property-denoting (or class-denoting). Or, one is (in some sense) more referential and the other one less so. In a predicational sentence, the more referential expression is in subject (or topic) position. In a specificational sentence, the relation is inverted. Consider the sentences in (11). (Please assume that in both sentences the proper name *John* is used to refer to a particular person whose identity is taken for granted).

(11) English (constructed)

a. *John is the best cook*. (predicational)

b. *The best cook is <u>John</u>*. (specificational)

In an "inverse predication" account of specificational sentences, the assumption is made that the referential status of the expression *the best cook* in both (11)(a) and (b) is essentially the same but that the terms are inverted.[18] As suggested earlier, however, the position I follow is that, in principle, a specificational sentence involves two referential expressions (somewhat like an equative/ identity statement sentence). In this sense, both specificational sentences and identity statements may be considered equational sentences. The point of a specificational sentence is not, however, (only) to assert that two nominal expressions refer to the same entity. As pointed out above, I understand the notion of specification as a particular relation between two expressions in a sentence where one (the VALUE expression) is used to clarify the reference of the other (the VARIABLE expression). Having said this, I must point out that there are good arguments (and problems) associated with either position. In the following I will shortly discuss an account supporting the "inverse predication" approach.

[17] Expressions involving nouns used predicatively invite a description in terms of class-inclusion, some kinds of adjectives less so. In some languages, some adjective-like words may be treated much like nouns while others may be treated much as (stative) verbs.

[18] Note that the inversion of relations is not (only) meant in terms of word order. In English, the order of constituents in a specificational sentence such as (11)(b) is in principle reversible, but a specificational sentence with a VALUE-VARIABLE order such as *<u>John</u> is the best cook* is often assumed to require a special prosodic pattern (the pre-copular expression being accented) to allow a specificational interpretation (but see Van Praet 2019). In context, the pragmatic difference would be clear. A specificational relation would require someone being 'the best cook' to be somehow at issue.

2.1.4 Inverse predication approach to specification

Patten (2016) argues for an "inverse predication" analysis of specificational sentences. The basic idea is that a predicational (nominal) sentence is construed (or conceptualized) placing a referent as the point of departure and assigning class-membership to this referent. In a specificational sentence, the perspective is shifted, and the point of departure is a class of entities. Specificational sentences are, in Patten's account, about (exhaustively) listing the members of a class.

A problem for "inverse predication" accounts of specificational sentences is that while in specificational sentences word order may easily be reversed, this is often not the case with predicational sentences. (Note that the discussion now concerns word order patterns of specificational and predicational sentences in English.) Consider the following example:

(12) English (Patten 2016: 93)
a. *Carla is a waitress*.

b. *#A waitress is Carla*.

Arguably, the most readily available reading for the sentence in (12)(a) is a predicational one. That is, one can easily imagine the sentence being used to provide information concerning Carla's occupation. In such a case, the expression *a waitress* is not used to refer (or, in any case, not to refer to an individual). Patten (2016) argues that the problem with sentences such as (12)(b) can be reconciled with an "inverse predication" approach. As pointed out above, Patten (2016: 79) assumes a particular kind of predicational relation in sentences involving nominal (as opposed e.g. to adjectival) predicates, namely class-inclusion. She cites Croft (1991: 69) who describes this kind of relation as "the relation between token to its subsuming type". In this sense, Carla is conceived of as a member (or token) of the class (or type) 'waitress'. The distinction between class inclusion and property ascription is important in Patten's account. The author argues that "the act of specification can only occur where a class inclusion relation is present" (p. 79).[19] Patten regards specificational relations as involving the (exhaustive) listing of the members included in a class. The problem with a sentence such as (12)(b), according to Patten, would be that the list of members belonging to class of waitresses would (obviously) not be exhausted by one individual. Consider now the following sentences:

(13) English (Patten 2016: 79)
a. *Diane and Carla were the waitresses*.

b. *The waitresses were Diane and Carla*.

In Patten's (2016) account, the relation holding between the pre- and post-copular expressions in (13) is described as class inclusion in the same sense as that holding between the expressions in (12). The difference would be that in (13) the class is narrowed down (the class would include e.g. all waitresses working at a particular bar). Having narrowed down the class, an (exhaustive) listing of its members is possible and the sentence in (13)(b) is felicitous (unlike the one in (12)(b)). The idea that the (exhaustive) listing of the members of a class is the best way to describe a specificational relation would be supported, according to Patten, by the contrast found in cases such as the following:

[19] This implies that a predicational relation such as that holding in a sentence such as Carla is lazy cannot be inverted. Patten does not address this issue further (discussing e.g. sentences such as *The lazy one is Carla*).

(14) English (Patten 2016: 93)

a. *There are several waitresses in the Boston Barmaid contest.* [*An especially efficient waitress*] *is* <u>*Diane Chambers*</u>.

b. *There are several waitresses in the Boston Barmaid contest. ?*[*An efficient waitress*] *is* <u>*Diane Chambers*</u>.

Patten accounts for the contrast with respect to the felicity judgements assigned to the sentences (14)(a) and (b) arguing that "the adverb *especially* enables a uniqueness interpretation" (p. 94). That is, Patten argues, the waitress in question "could well be the most efficient waitress" (p. 94), thus narrowing the membership of the relevant class to a single individual. A problem with Patten's (2016) account pointed out in Davidse & Van Praet (2019) is that Patten (2016) downplays the fact that the use of an indefinite article in sentences such as those in (14) carries the implicature that the individual to which it refers is not the only (relevant) individual to which the properties apply. Thus, the exhaustive listing explanation is problematic.

There is a further problem, in my view, with Patten's (2016) account. Though the author does not directly discuss the use of bare (i.e. determiner-less) nouns used as nominal predicates, she does, however, mention (if I understand correctly) that the relation holding between these and the subject is one of class inclusion (p. 80, 95). If this is the case, the contrast concerning the acceptability arguably found in the following sentences remains unexplained:

(15) English (adapted from Moltmann 2013: 51-52)

a. *John is* (*the*) *mayor of Cambridge*.

b. ?(*The*) *mayor of Cambridge is John.*

I assume that the sentence in (15)(a), under any reading, is fully acceptable to English speakers. I will also assume that in the absence of the article, only a predicational reading is possible. Arguably, in the absence of the article, the sentence in (15)(b) is problematic (perhaps even unacceptable). Assuming that a class-inclusion relation should in principle hold, and that the class in question may be said to include a single member, it is not clear, following Patten's (2006) account, why a variant of (15)(b) involving a determiner-less nominal in pre-copular position should not be as acceptable as one with a determiner. Perhaps the restriction against sentences such as the determiner-less variant of (15)(b) can be stated in terms of a general dispreference in the language for expressions other than referential nominal expressions in pre-copular (canonically subject) position. Now, arguably, the restriction against determiner-less expressions in pre-copular positions does not automatically rule out the possibility of assigning the same (referential) status to an expression which is required to be determiner-marked in pre-copular but not (necessarily) in post-copular position.

2.1.5 Quasi specification

There are constructions in which a theme-rheme relation between a (referential) subject expression and a (non-referential) predicate expression (nominal or not) is inverted, but differing from specificational sentences in that they do not involve two referential expressions. Consider the following German sentence:

(16) German (Hartmann & Heycock 2020: 12)

Ärztin	*ist*	(*nur*)	*die*	*Johanna*.
doctor.SG.F	be.3SG.PRS	only	ART.SG.F	Johana

'Only Johanna is a doctor.'

Hartmann & Heycock (2020) argue that "it is clear that [(16)] is not a specificational sentence, rather the initial nominal is actually a predicate […] a bare (determiner-less) nominal of a type that typically cannot appear in subject position […] just like e.g., mayor or indeed doctor in English" (p. 13). Compare the following (generally unacceptable) sentence:

(17) German (Hartmann & Heycock 2020: 13)
Ärztin ist nie reich.
doctor.SG.F be.3SG.PRS never rich
'Doctor is never rich.'

For the authors, a specificational sentence must involve two nominal expressions which may in principle function as arguments. This is in principle the position I follow.

Note that in German not only expressions involving nouns (i.e. expressions which may be described in some sense to denote a class of individuals) but also expressions denoting properties may occur in pre-copular position. Consider the following sentence:

(18) German (www)
Fertig ist nur der Eingangsbereich.
ready be.3SG.PRS only ART.SG.M entrance_area
'Only the entrance is ready.'

The sentence in (18) may be described in some sense as listing all the things that are ready (in the context of the utterance) but would perhaps not be best described as involving a class-inclusion relationship. Note that the adverb *nur* 'only' is not necessary for the reading in question. A rising accent on the predicative adjective *fertig* 'ready' is sufficient.

In some way, sentences such as (16) and (18) may be described as having a specificational sense or use. Perhaps a good term for such sentences is the term 'quasi-specificational', suggested in Lauwers & Tobback (2020) for copular constructions with pre-posed adjectival predicates in (written) French. Consider the following example:

(19) French (Lauwers & Tobback 2020: 471)
(In 1840, the AIBL awarded a first medal to Jollois[…], for a thesis on […])
Plus importante encore a été l'oeuvre de Th. Vacquer […]
more imortant.SG.F still have.3SG.PRS be.3SG.PST ART_work of Th. Vacquer
'Even more important was the work of Th. Vacquer…'

The authors note that such constructions are pragmatically very closely related to clefts:

(20) French (Lauwers & Tobback 2020: 482)
Ce qui a été plus important,
DEM.SG.M who have.3SG.PRS be.3SG.PST more important.SG.M
c'était l'oeuvre de Th. Vacquer
DEM.SG.M_be.3SG.IPFV ART_work of Th. Vacquer
'What was more important was the work of Th. Vacquer…'

Quasi-specificational sentences in the sense of Lauwers & Tobback (2020) may be said to involve a specificational relation in the sense that the pre-copular expression implies an entity with the relevant

properties and that the nominal expression in post-copular position is felt to specify the identity of this entity (possibly) to the exclusion or in contrast to other possible entities.

There are other patterns reported to trigger similar pragmatic effects. Lahousse (2007) reports VOS word order in Romance to trigger such effects. The following sentence is described to trigger an "exhaustive identification reading":

(21) French (Lahousse 2007: 385)

Receveront	*un*	*bulletin*	*de*	*vote*	*les*	*étudiants*	*et*
receive.3PL.FUT	ART.INDF.SG.M	ballot	of	vote	ART.PL	student.PL	and

le	*personnel*	*académique*.
ART.SG.M	staff	academic.SG

'Students as well as academic staff will receive a ballot paper.'

Lahousse & Lamiroy (2012) note that in French, sentences such as (21) (which, as the authors seem to suggest, may be more or less restricted to particular registers such formal administrative writing) permit only an exhaustive specification reading. That is, the French sentence, unlike its English translation, very strongly implicates that only the students and the staff will receive a ballot paper, to the exclusion of any other potential recipients. The authors note that VOS sentences without the contrastive-exhaustive intention are in principle unacceptable. Lahousse (2007) calls VOS in Romance specificational (and suggests a bipartite analysis of such sentences) but Lahousse & Lamiroy (2012) drop the term and do not pursue a bipartite analysis.

The purpose of discussing accounts such as that in Lauwers & Tobback (2020) and Lahousse (2007) is to clarify that there may be many constructions which may trigger similar pragmatic effects as those commonly associated with specificational sentences (including clefts) but do not correspond to my understanding of the notion of specificational sentence. In my view, furthermore, the recognition of these patterns makes a characterization of specificational sentences in terms of notions such as "exhaustive listing" and "inverse predication" problematic as I see no principled way to distinguish quasi-specificational from specificational sentences. I therefore adhere to the (admittedly, not unproblematic) view that it may be better to regard specificational sentences as sentences involving two referential nominal expressions in a particular relation to each other. (And thus, to view specificational sentences as being a type of equational sentences.)

2.1.6 Specificational relations

I will now illustrate my understanding of the different possible relations in copular sentences comparing different possible readings (uses) of constructed sentences.

(22) English (constructed)

a. *John's neighbor is Mary's cousin*.

b. *Mary's cousin is John's neighbor*.

The sentences in (22) allow different readings. Under a predicational reading, the pre-copular expression refers to a particular individual and the post-copular one ascribes a property to its referent (it does not refer to an individual). Under an equational reading, both the pre-copular and the post-copular expressions refer to individuals. Here, there are two possibilities, an identity statement reading and a specificational one. Under an identity statement reading, both the pre- and post-copular expressions are intended to refer to particular individuals. The speaker assumes that the hearer associates these expressions with particular individuals (or is ready to take this for granted) but is not aware that they refer to the same individual. In a

specificational sentence, both expressions in principle also refer to two specific individuals. The difference with respect to identity statements is that the point of a specificational sentence is to specify the reference of an expression (the VARIABLE expression) by means of another (the VALUE).

I will now clarify how the different readings may come about. I will use (constructed) mini-discourse contextualizations. Some of these are based on question-answer pairs. I do not assume that questions (explicit or otherwise) are necessary, nor do I assume that full sentences are necessarily the most natural way to answer the relevant the questions in the examples. This is just an expository device frequently used in the literature which I happen to find useful (i.e. intuitively accessible).

Under a predicational reading, the sentences in (22) could be used to answer the following questions:

(23) English (constructed)
a. *How is John's neighbor related to Mary*?
John's neighbor is Mary's cousin. (alternative: *She's her cousin*.)

b. *Does Mary's cousin live close to John?*
Mary's cousin is John's neighbor. (alternative: *She's his neighbor*.)

The point here is that the post-copular expressions in (23)(a) and (b) are used to ascribe a property (a kinship relation with someone and the property of living in the immediate vicinity of someone, respectively). Obviously, the pre-copular expressions in the examples entail such properties as well but they are also used to refer to particular individuals. By contrast, the post-copular expressions (under a predicational reading) arguably do not. A good answer to the question in (23)(b) could be for instance *She lives right next door from him* or something similar. It seems that ascription of kinship relations in English generally requires nominal predicates which are formally indistinguishable from (referential) nominal expressions. I will turn now to equational sentences starting with identity statements.

An identity relation would hold, for instance, in the answer part of the mini-discourse in (24).

(24) English (constructed)
I heard John's neighbor and Mary's cousin were going to join us.

Mary's Cousin is John's neighbor. / John's Neighbor is Mary's cousin.
(alternative: *They're the same person. / They're one and the same*.)

In the case at hand, the answer sentences would possibly involve a specific prosodic realization, with the copula accented, possibly so because of a corrective intention. This need, however, not be the case. The crucial point is that in (24), the post-copular expressions are used to refer (in this case, to particular individuals) and not to ascribe a property to the referent of the pre-copular one. It is not clear how frequently equational sentences are actually used to convey the point that two expressions refer to the same individual in naturally occurring speech. Perhaps constructions similar to the suggested alternative answer in (24) (in parentheses) are preferred.[20] I will now illustrate specificational relations between two expressions.

[20] There are cross-linguistically oriented studies in the literature on "equative" constructions (e.g. Treis & Vanhove (eds) 2017). In these studies, however, the notion "equative" concerns constructions asserting that two referents may be ascribed a property to the same degree (e.g. *John is as tall as Mary*). This is not what I have in mind but rather sentences of the sort suggested as alternatives to the equational sentences in (28).

Imagine someone has been referring to some individual as *John's neighbor* in a conversation. In a previous conversation, the speaker had referred to the same person as *Mary's cousin*. At some point, he might utter the following sentence:

(25) English (constructed)
I heard John's neighbor doesn't like cats.
By the way, John's neighbor is Mary's cousin.
You know, the one I was talking about the other day. Some of you may have have met her.

In this case, the speaker is arguably not (mainly) conveying information regarding the kinship relation between John's neighbor and Mary. The expression *Mary's cousin* is used to refer to a particular individual. This would make the sentence *John's neighbor is Mary's cousin* in (25) different from a predicational one like the sentences in (23). The difference to an identity statement (illustrated in (24)) would be that, in the case of the specificational sentence in (25), the sentence is used not only to state that two expressions refer to the same individual but, additionally, to clarify the reference of one expression (in the case at hand, *John's neighbor*) using another expression assumed to be more accessible, familiar, or informative to the addressee. In the case at hand, the expression *Mary's cousin* may be more familiar to the addressee only by virtue of having been introduced earlier. Now, admittedly, the difference argued to hold between the cases illustrated in (24) and (25) is tenuous. The choice of examples is deliberate, however. The point I am trying to make is that the difference between an identity statement is a matter of relative familiarity or informativeness between two nominal expressions. As soon as there is an asymmetry between two terms in an equational sentence, the relation between the terms is a specificational one. As suggested earlier, often, copular clauses involving two proper names (cf. *Cicero is Tully* in (10) above) are adduced as examples for identity statements in the literature. But consider the following (oft cited) sentence:

(26) English (www)
Samuel Clemens is Mark Twain.

The sentence in (26) may be described as equational. In principle, the sentences assert that two expressions refer to the same individual. An identity statement reading is in principle possible. It is conceivable, however, that someone may be familiar with the writer mentioned in (26) but only know (or remember) his pen name (i.e. *Mark Twain*). In this case, the sentence in (26) could be uttered as an answer to the question *Who is Samuel Clemens?* Used in this way, the sentence may be described as specificational. The civil name (*Samuel Clemens*) is a VARIABLE expression and the pen name (*Mark Twain*) the VALUE. The property that distinguishes the VARIABLE from the VALUE is the (assumed) degree of familiarity (or informativeness) the addressee has with one expression relative to the other.

2.2 Linguistic reflexes of the predicational-equational distinction

2.2.1 Formal distinctions between equational and nominal-predicate sentences

Although in many languages equational and nominal-predicate sentences may be coded more or less identically, in some languages these kinds of sentences may be formally distinguished. Dryer (2007) mentions three cross-linguistically recurrent possible differences between the two sentence types.

- word order patterns
- copularization patterns

- formal distinction between nominal predicates and nominal expressions

The first difference concerns typically word order restrictions occurring in one type of sentences and not in the other, rather than two distinctive word order patterns. In predicational sentences, word order may be restricted while word order in equational sentences is flexible.[21] Second, copularization patterns may be different. That is, an equational sentence and a sentence involving a nominal predicate may involve the use of different copulas or only one of these sentence types may require a copula while the other is construed by means of juxtaposition of expressions.[22] Third, the form of the expressions involved may be different. That is, an equational sentence may involve two nominal expressions (formally identical to those occurring as arguments in a regular sentence) while in a predicational sentence the nominal predicate may be coded differently. For instance, in many languages, referential nominal expressions require (or more or less consistently involve) marking with articles (or determiners of some sort) while (non-referential) nominal predicates lack such marking. In the following I will discuss data from different languages to illustrate these distinctions. I will start by briefly discussing the formal distinction between nominal predicates and nominal expressions (the discussion will be continued in 2.3.2).

In some languages, nominal expressions (typically with the exception of pronominal forms and sometimes also proper names) require some kind of marker (e.g. an article) to be used as arguments (or adjuncts). Expressions lacking this marking may normally only be used as predicates. This is reported to be the case, for instance, in Philippine languages. In these languages, non-pronominal arguments (and adjuncts) typically require the use of an article and the use of bare nouns is restricted to nominal predicates occurring in (clause-initial) predicate position. But note that in these languages an article-marked expression may occur in predicate position, in juxtaposition to another (article-marked) nominal expression (these languages lack copulas). Thus, nominal-predicate sentences and equational ones can be distinguished by the presence or absence of an article preceding a nominal in predicate position. Consider the following examples from Cebuano.[23]

(27) Cebuano (Dryer 2007: 235)

a. *ang* *duktur* *ang* *babayi*
TOPIC doctor TOPIC woman
'The woman is the doctor.'

b. *duktur* *ang* *babayi*
doctor TOPIC woman
'The woman is a doctor.'

According to Dryer (2007: 235), the example in (27)(a), where the expression in predicate position is introduced by the article *ang*, can be described as an equational sentence and the one in (27)(b) as a predicational one. In the case of (27)(a), the sentence would be used to assert that the referents of the two

[21] Obviously, identity statements should permit word order permutation. (Though only if the expressions involved are of the same kind; some languages may exhibit specific restrictions when involving e.g. proper names and pronominal forms; see e.g. Kaufman 2018 on Tagalog.) In the case of specificational sentences this is different, it may be the case that the order VARIABLE-VALUE (or VALUE-VARIABLE) in a particular language be fixed. I will discuss word order in more detail in Chapter 4.

[22] Stassen (2007: 120) argues that if a language employs zero-copula in only one type of sentences it is in equational ones.

[23] Dryer (2007) glosses *ang* as TOPIC. The element in question is more commonly referred to as a nominative article.

nominal expressions involved are identical and in (27)(b) the nominal predicate *duktur* 'doctor' would ascribe a property to the referent of the (article-marked) nominal expression following it.

Now, as mentioned in 2.1.2, in English the use of a bare nominal and a nominal expression bearing an article (or determiner) may also distinguish nominal predicates from (referential) nominal. In English, however, the use of bare (determiner-less) nouns is rather restricted and many expressions require an article, whether used referentially or not. (Note, for instance, that the English translation of the Cebuano sentence in (27)(b) involves a nominal predicate introduced by the indefinite article.) Thus, the presence (or absence) of an article in English does not necessarily distinguish the status of an expression as a (referential) expression in an equational sentence from that of a nominal predicate. This holds also for languages in which the use of bare nominals as predicates is less restricted than it is in English. Consider the following German examples:

(28) German (constructed)

a. *Martin ist Arzt.*
Martin is doctor
'Martin is a doctor.'

b. *Martin ist ein Arzt.*
Martin is a doctor
'Martin is a doctor.'

The sentence in (28)(a), with a bare noun denoting an occupation in post-copular position, is unambiguously predicational. It is possibly preferred to the sentence in (28)(b) if the point of the sentence is simply to provide information concerning Martin's occupation. However, a predicational reading of (28)(b) is also perfectly possible and, at least in isolation, arguably more salient than an equational one (an identity statement reading may be impossible). Consider, however, the following sentences:

(29) German (constructed)

a. *Martin ist ein sehr kompetenter Arzt.*
Martin is a very competent doctor
'Martin is a very competent doctor.'

b. ?*Martin ist sehr kompetenter Arzt.*
Martin is very competent doctor
'Martin is a very competent doctor.'

Arguably, the variant in (29)(a) may be used to ascribe a property to Martin (and an equational reading may be as difficult as in the case of (28)(b)). The determiner-less variant in (29)(b), however, though ruled out (similar sentences do occur), seems to be much less preferred. The relevant generalization seems to be that expressions involving (singular) nouns and adjectival modifiers in German normally require a determiner. If no modifier is used, the absence of a determiner (which is otherwise possible for nominal predicates, at least with nouns denoting occupations) may distinguish a property-ascribing use from a referential one.[24] If a bare nominal is not available (or strongly dispreferred) as an option, an expression with the form of a regular nominal expression is used and it can be used either ascriptively or referentially. Note that a

[24] Sentences such as *Martin ist Arzt* and *Martin ist ein Arzt* (both in an arguably predicational reading) may have different meanings and behaviors. See e.g. Geist (2014) on bare predicates in German, Zobel (2017, 2021) on the sensitivity of natural language to the distinction between "class nouns" and "role nouns". I will ignore these (interesting) distinctions and assume that, for the purposes of the present discussion, they can be subsumed as *properties* in one category.

distinction between equational and nominal-predicate sentences may be problematic (and particularly so in clefts) even in languages for which a clear formal distinction is reported to exist. As this problem concerns the distinction between specificational from predicational sentences (rather than the equational vs predicational distinction more generally), I will discuss the relevant data in section 2.3.2. I will turn now to data from two languages for which it is reported that the choice of copulas may distinguish equational from predicational sentences.

2.2.2 Equational vs. predicational copulas

Before discussing the relevant data, a few short remarks on the notion of copula are in order. I use the term copula to describe a support element used to mark a subject-predicate (or a subject-predicate-like) relation between two expressions (much in the same sense as the term 'connector' is used in Adamou & Costaouec (2010)). A copula in this sense may be characterized as a marker used in association with an element to mark its predicate (or predicate-like) status. Copulas may behave morphosyntactically as verbs as in the case of English *be* and its cognates in many Indo-European languages. Copulas may behave in some respects like verbs but not in others (as reported in the literature e.g. for Mandarin *shì*). In some languages, copulas may share no properties with verbs at all. As suggested earlier, in some languages, different copulas mark different relations. Some languages distinguish stage-level from individual-level predicates. Many languages distinguish locative copulas (usually related to verbs expressing position or posture). In the discussion that will follow, the relevant distinction will be the one between different copulas used in equational and predicational sentences.

It is important to keep in mind that elements described as copulas are often polyfunctional. The elements in question may be used as copulas in some contexts but may also be used in contexts not involving a subject-predicate(-like) relation between two expressions. For example, forms from a pronominal paradigm may be used as copulas (or in a copula-like manner). Conjunctions (otherwise serving to link or introduce clauses) may also be used to link nominal expressions in a copula-like way. Accounts may differ in the way these elements are analyzed. Thus, for example, a form described as a pronominal copula (in the sense just alluded to) in one account may be analyzed as a resumptive pronoun or as an instance of argument doubling in another (see discussion on Hebrew in 2.3.1.2 below). Similarly, a form may be analyzed as a "focus marker" in one context and as a copula in another. Or, the element in question may be argued to function first and foremost as a "focus marker", which may be used a copula-like manner in particular environments.

The following discussion concentrates on two languages, Akan/Twi and Thai, exhibiting a rich inventory of copulas, two of which, respectively, are reported to distinguish nominal-predicate sentences from equational ones. The characterization in the literature is more consistent for Thai (though see remarks at the end of the corresponding section). In the case of Akan, there is a tendency in the literature to associate the element used in equational sentences with a "focus marker".

2.2.2.1 Thai copulas

Thai (Tai-Kadai) exhibits a rich copular system. Wongwattana (2015) distinguishes five different copulas in the language. Or rather, five different elements that may be used as copulas, as none of these are exclusively used as copulas. I will concentrate on the contrast between the use of the copulas *pen* and *khɨ:*. The former is described in Iwasaki & Ingaphirom (2009: 221) as a "semi-verbal" and the latter as a "linker". Neither seem to behave as regular verbs in the language. Wongwattana (2015) points out that *pen* may be used (more or less) as a verb with the meaning 'be alive'. In some Thai dialects (and in Lao) *khɨ:* is also reported to be used in a verb-like way with the meaning 'resemble'. Iwasaki & Ingaphirom (2009) point out

the frequent use of *khɨ:* as a linker (conjunction or sentence introducer) with the meanings 'that is', 'because' and 'I mean'. Kuno & Wongkhomthong (1981) point out that "the difference in meaning between *pen* and *khɨ:* [in their use as copulas] is so subtle that native speakers cannot pinpoint what it is" (p. 66). Jaratjarungkiat (2019: 4) notes that "Thai people may be unable to differentiate" sentences involving one or the other form and that "even the dictionary defines /khɯɯ/ as /pen/ (Royal Institute of Thailand, 2013: 752)". Kuno & Wongkhomthong (1981) argue that the choice between the two copulas (none of which seems to be, strictly speaking, obligatory), can distinguish between "characterizational" and "identificational" sentences (corresponding, respectively, to the predicational vs. equational distinction in the terms I am using in this dissertation). Consider the following examples. Note that either *pen* or *khɨ:* may be used, but the interpretation differs:

(30) Thai (Kuno & Wongkhomthong 1981: 72)

a. *cɔ:n* *pen/ khɨ:* *câwkhɔ̌:ŋ* *ro:ŋri:an* *àme:ríka:.*
John COP owner school America
'John is the owner of the American School.'

b. *cɔ:n* *pen/ khɨ:* *prátha:n* *sàma:khom* *àme:ríka:.*
John COP president association America
'John is the president of the American Association.'

According to Kuno & Wongkhomthong (1981), the variants of the sentences in (30)(a) and (b) with the copula *pen* are interpreted as ascribing a property to John (i.e. the property of presiding over of the American Association and of owning the American School respectively). Arguably, the same effect could be unambiguously achieved in English using a bare nominal in post-copular position (i.e. *John is president of the American Association, John is owner of the American School.* The English translations of (30)(a) and (b), as they stand, are ambiguous (the definite article would signal only that the property associated with the post-copular expression can apply to a unique referent). Kuno & Wongkhomthong (1981) argue that, in the variant involving the copula *khɨ:*, the sentence does not ascribe a property to a referent but asserts that the expressions preceding and following the copula refer to the same person. To illustrate the difference between the interpretations triggered by the choice of the different copulas, the authors note that a sentence involving predicate coordination is not compatible with the copula *khɨ:*.

(31) Thai (Kuno & Wongkhomthong 1981: 72)
cɔ:n *pen/**khɨ:* *tháŋ* *câwkhɔ̌:ŋ* *ro:ŋri:an* *àme:ríka:.*
John COP both owner school America
lɛ́ *prátha:n* *sàma:khom* *àme:ríka:.*
and president association America
'John is both the owner of the American School and the president of the American Association.'

Kuno & Wongkhomthong (1981) argue that the restriction in the choice of copula used "must be due to the fact that the sentence [(31)] presents various qualifications or characterizations of John, and is not amenable to the identificational interpretation" (p. 73).

A further illustration of the distinctive uses of the copulas *pen* and *khɨ:* is presented in the following example (in the variant with *khɨ:*, a cleft):

(32) Thai (Kuno & Wongkhomthong 1981: 73)

[*sìŋ*	*thîː*	*khun*	*chɔ̂ːb*	*mâːg*	*thîː*	*sùd*]	*pen/khɨː*	*sǐː khăːw*	*chây_máy.*
thing	that	you	like	much	that	most	COP	color white	TAG.QUESTION

'The thing that you like best is white, isn't it?'

The intended context for the (presumably) constructed the sentence in (32) would be a guessing game. The participants may ask questions about an undisclosed object to narrow down its possible reference. In the variant involving *pen*, the expression following the copula (*sǐː khăːw* 'white') is understood a property of the referent. That is, the player uttering the sentence is asking whether the color of the undisclosed object is white. In the variant involving *khɨː*, the sentence would be an attempt to finally disclose the identity object in question. That is, the post-copular expression would refer to the color itself (i.e. the abstract entity 'white').

The examples presented in (30), (31), and (32) above illustrate the effects observed in the alterative use of the copulas *pen* and *khɨː*. In some environments, only one of these may be used. Consider, for example, in the equational sentence below, in which two alternative names for the same country are linked by the copula *khɨː* (and the copula *pen* is reported to be inappropriate):

(33) Thai (Wongwattana 2015: 112)

sàʔjăːm	*kʰɨː/*pen*	*pràʔtʰêːt*	*tʰaj*
Siam	COP	country	Thai

'Siam is Thailand.'

In cases such as the sentences below, by contrast, the copula *pen* is used:

(34) Thai (Wongwattana 2015: 107)

a. *kʰăw pen bâː*
s/he COP insane
'She is insane.'

b. *mɛ̂ː kʰɔ̂ːnkʰâːŋ pen jàj*
mother somewhat COP big
'The mother is somewhat big/important.'

c. *tɕìttɕaj raw pen sùk*
mind we COP happy
'We are happy.'

The sentences in (34) are translated into English using adjectival predicates. Wongwattana (2015: 107) describes such expressions as "state verbs that denote non-inherent and non-evaluative qualities" (p. 106). Expressions notionally corresponding to adjectives in English pattern in Thai with verbs. Not all of these are compatible with *pen.*[25] It seems, however, that none of these are compatible with *khɨː*. Apparently, none of these property-denoting words may (by themselves) be used as nominal expressions.

[25] Consider the following example:

(i) Thai (Wongwattana 2015: 107

kʰăw	(**pen*)	*diː/sŭaj/sŭːŋ/kʰăːw.*
s/he	COP	good/pretty/tall/white

'S/he is good/pretty/tall/white.'

It must be pointed out that the contrast between equational and predicational readings is not always as clear-cut as the examples above would suggest. Kuno & Wongkhomthong (1981) note that in a number of cases, the use of *pen,* but not that of *khɨ:,* is possible. These cases involve particular syntactic configurations. The animacy status of the referents of the expressions involved may also play a role (the less human-like a referent of the post-copular expression, the less likely it is that *khɨ:* is used—at least in some environments). On the other hand, Takahashi & Shinzato (2003) argue that, while Kuno & Wongkhomthong's (1981) account of *pen* and *khɨ:* as marking a distinction between equational and predicational sentences is essentially correct, some further refinements may be made. Note, however, that in some accounts (e.g. Ruangjaroon 2007, Tawilapakul 2017) clefts (at least of a "contrastive" type) are associated with the copula *pen,* and not with *khɨ:.* (See also Hedberg & Potter (2010) for an experimental study and the difficulties involved in assessing the predicational vs. equational status of sentences involving *pen* and *khɨ:*.)

2.2.2.2 Akan/Twi copulas

Like Thai, Akan/Twi (Kwa, Niger-Congo) distinguishes copulas with different functions. Ellis & Boadi (1969) offer a detailed account of the use conditions of the different copulas in Akan (see also Ofori 2011, and Quarcoo 2014). I will concentrate on the contrast between the copulas *yɛ* and *ne,* which are reported to consistently distinguish predicational from equational sentences. The copula *yɛ* may be used to form sentences involving predicate expressions denoting different kinds of entities (humans, things, places and times). These must be sortally congruent with the subject expression (i.e. a clause involving a person-denoting expression as a subject and a place-denoting predicate expression would require the locative copula *wo* instead). Ellis & Boadi (1969) note that there is a general restriction concerning word order in sentences involving the copula *yɛ*. In the authors' terms, the "more concrete" or "more specific" of both expressions must occur in pre-copular position. Consider the following examples:

(35) Akan (Ellis & Boadi 1969: 17)

a. *ɔbarima no yɛ ɔsɔfoɔ*
man DEF COP priest
'The man is a priest.'

b. *Kwasi yɛ okuafoɔ*
Kwasi COP farmer
'Kwasi is a farmer.'

c. *meyɛ ɔkyerɛkeyerɛfoɔ*
1SG.COP teacher
'I am a teacher.'

d. *ɛha yɛ sukuu*
here COP school
'Here is a school.' (i.e. 'This place is a school.')

e. *ɛnnɛ yɛ buronya*
today COP Christmas
'Today is Christmas.

Word order reversal of the pre- and post-copular expressions in sentences such as those presented in (35) is reported to be unacceptable. Also note that, according to Ellis & Boadi (1969), pronouns, demonstratives, proper names or a definite-marked nominal expressions (in principle, unambiguously referential expressions) cannot occur in post-copular position in *yɛ* copular sentences (p. 19).

In addition to expressions denoting persons, objects, and locations, the expression following the copula *yɛ* may be a numeral (36) or an adjective (37) (some of which behave morphologically like nouns in the language):

(36) Akan (Ellis & Boadi 1969: 22)
daka *no* *yɛ* *baako*
box DEF COP one
'The box is one' (i.e. 'There is one box.')

(37) Akan (Ellis & Boadi 1969: 22)
a. *dua* *no* *yɛ* *tia*
tree DEF COP short
'The tree is short.'

b. *nnua* *no* *yɛ* *ntia*
PL.tree DEF COP PL.short
'The trees are short.'

c. *nua/nnua* *no* *yɛ* *fɛ*
tree/PL.tree DEF COP beautiful
'The tree(s) is/are beautiful.'

A sentence involving the copula *ne,* in contrast, does not allow adjectives as constituents. Nouns such as *ohiani* 'pauper' may occur. In *ne* sentences, however, such nouns may be better described not as denoting a property (the property of being poor) but a bearer of the property in question. Consider the following example:

(38) Akan (Ellis & Boadi 1969: 23)
ɔne *ohiani*
3SG.COP poor
'He is the poor one.'

Note that the English translation of (38) suggests definiteness. Bombi (2018) reports that some bare nouns may be interpreted as definite in the language.

Unlike sentences involving the copula *yɛ,* sentences involving the copula *ne* do not follow the restrictions discussed above for *yɛ* copular sentences, for which is reported that only one "concrete" or "specific" expression may be involved and must occur in pre-copular position. This is illustrated in the following example, in which the post-copular expression involves the marker *no*:[26]

(39) Akan (Ellis & Boadi 1969: 24)
me *nua* *ne* *ɔbarima* *no*
3SG brother COP man DEF/DEM
'My brother is that man.'

[26] The marker *no* is alternatively translated as *that* and *the* in Ellis & Boadi (1969). In the latter use, *no* is reported to correspond very closely in its use conditions with the definite article in English. For a recent discussion of definiteness marking in Akan see Bombi (2018).

The following examples from Boadi (1974) illustrate the principle in in cleft sentences:[27]

(40) Akan (based on Boadi 1974: 15)

a. [*nea* *ɔbaa* *ha* *no*] *ne/yɛ* *kwasea*
REL.PRO(?) 3SG.come.PST here DEF COP fool
'The one who came here was/is a fool.'

b. [*nea* *ɔbaa* *ha* *no*] *ne/*yɛ* *kwasea* *yi/no*
REL.PRO(?) 3SG.come.PST here DEF COP fool PROX/DIST
'The one who came here was/is this/that fool.'

In a sentence such as (40)(a), where the post-copular expression may obtain a non-referential reading, the use of either copula is possible. The use of unambiguously referential expressions in post-copular position, as in (40)(b), is incompatible with the copula *yɛ*.

It seems that sentences involving the copula *ne* freely allow word order permutation. This is illustrated in the following examples involving clefts from Boadi (1974):

(41) Akan (Boadi 1974: 23)

a. [*onipa* (*ko*) *a/* *nea* *ɔbaa* *ha* *no*] *ne* *me*
man PTCL REL REL.PRO(?) 3SG.come.PST here DEF COP 1SG
'The one who came here was I.'

b. *me* *ne* [*onipa* (*ko*) *a/* *nea* *ɔbaa* *ha* *no*]
1SG COP man PTCL REL REL.PRO(?) 3SG.come.PST here DEF
'I was the one who came here.'

(42) Akan (Boadi 1974: 23)

a. [*ade* *a* *metɔɔe*] *ne* *ataade*
thing REL 1SG.buy.PST COP shirt
'What I bought was a SHIRT.'

b. *Ataade* *ne* [*ade* *a* *metɔɔe*]
shirt COP thing REL 1SG.buy.PST
'A SHIRT was what I bought.'

(43) Akan (Boadi 1974: 35)

a. [*bere* *a* *mebaa* *ha* *no*] *ne* *Ɔpɛpɔn*
time REL 1SG.come.PST here DEF COP January
'The time that I came here was JANUARY.'

b. *Ɔpɛpɔn* *ne* [*bere* *a* *mebaa* *ha* *no*]
January COP time REL 1SG.come.PST here DEF
'JANUARY was the time that I came here.'

[27] The cleft clause in Akan is formed either by a common noun linked to a clause introduced by the particle *a* (glossed here as a relative particle) or by the relative-pronoun-like morphemes *nea* or *deɛ*. The choice subject to dialectal variation (Campbell 2020). The use of the latter strategy is apparently limited to oriented clausal nominalizations involving humans. In cases oriented to non-human entities (e.g. inanimate objects, locations or times) the former strategy is used. (See further discussion in 4.1.2.)

Word order flexibility is, according to Ellis & Boadi (1969: 26), also available for simple (i.e. non-cleft) equational sentences.

Some authors (e.g. Ofori 2011, Quarcoo 2014, Amuzu (2005), Amfo (2010) describe the copula *ne* as equational (or "equative"). In the terms of Amuzu (2005), a post-copular nominal in a sentence with the copula *yɛ* "ascribes to a subject [expression] the properties associated with [the denotatum of a noun or an adjective]" (p. 87) while the copula *ne* "signals that the subject and the complement refer to the same entity" (p. 87). In other accounts, however, sentences involving the copula *ne* are discussed in terms of focus relations. For instance, Boadi (1974: 35) suggests that the morpheme *ne* may be a form conflating the marker *na* and the copula *yɛ*. The former, which also functions as a clause-intruducing conjunction in the language is often described as a focus marker. Ofori (2011) argues, however, that such an analysis is not compatible with morphophonological patterns in the language. Ellis & Boadi (1969: 26) explicitly state that either the pre- or post-copular nominal expression in a *ne* sentence may be "emphasized". This is not the case in constructions involving the focus marker *na*, where the (not necessarily nominal) expression preceding the marker is made prominent (or, perhaps, the clause following it is backgrounded).

2.3 Linguistic reflexes of specification

In the previous section (2.2), I examined data reflecting the difference between nominal predicate and equational sentences (including specificational sentences, which I assume to be best considered as a sub-type of the latter category, together with identity statements). In the present section I will turn to the examination of further phenomena reported in the literature to reflect the distinction between specificational sentences and both identity statements and nominal-predicate sentences.

2.3.1 On the special status of the VARIABLE expression

2.3.1.1 Pronominalization patterns (tag questions and resumptive pronouns)

There is evidence that a VARIABLE expression in a specificational sentence is not treated in the same way as a fully referential expression. A phenomenon frequently pointed out in the literature concerns pronominalization patterns in tag questions in English. Consider the oft cited examples in (44). The sentence in (44)(a) is specificational and the tallest girl in the class is identified as being Mary. In the predicational sentence in (44)(b) a property is ascribed to the tallest girl in the class (whoever she might be).

(44) English (Mikkelsen 2005: 64)
a. *The tallest girl in the class is Molly, isn't* {*it/*she*}*?*

b. *The tallest girl in the class is Swedish, isn't* {*she/*it*}*?*

Obviously, the gender of the referent of the expression *the tallest girl in the class* in (44)(a) and (b) is available in both cases. Yet it is reported in the literature that native speakers of English consistently use the neuter pronoun in tag questions attached to specificational sentences with (singular) human referents while they use a pronoun agreeing in gender with the subject expression in predicational ones. Note that, in principle, proper names may also be treated in an analogous manner:

(45) English (Moltmann 2013: 68)
John Miller is him. Isn't it / he?

Similar effects can be observed in other languages in which tag questions are not regularly formed following the pattern found in English. Consider the following German and French sentences illustrating the contrast

in the availability of the neuter demonstrative used as a resumptive pronoun referring back to a left-dislocated feminine expression denoting a human:

(46) German (Moltmann 2013: 52)

Diese	*Frau,*	*das*	*war*	*meine*	*Tante*	/	**das*	*war*	*schön*.
this.F	woman	that.N	was	my.F	aunt		that.N	was	beautiful

'This woman, that was my aunt/ that was beautiful.'

(47) French (Moltmann 2013: 52)

Cette	*femme,*	*c'était*	*ma*	*tante*	/	**c'était*	*belle*.
this.F	woman	that.N_was	my.F	aunt		that.N_was	beautiful.F

'This woman, that was my aunt/ that was beautiful.'

Though not obligatory, the choice of a neuter demonstrative (instead e.g. of the feminine demonstrative *die* or the feminine personal pronoun *sie* in the case of German) is available in the variant in which a referential nominal expression follows the copula but it is not acceptable if it is an adjective. Moltmann (2013: 62) points out that a similar pattern may be observed in clefts involving a left dislocated cleft clause. Consider the following examples:

(48) German (adapted from Moltmann 2013: 62)

Wen	*ich*	*getroffen*	*habe,*	*das*	*war*	*Hans*.
who.ACC	1SG	met	have	that.N	was	Hans

'Whom I met, that was John.'

(49) French (adapted from Moltmann 2013: 62)

Ce	*qui*	*j'ai*	*rencontré,*	*c'était*	*Jean*.
DEM	who	1SG_have	met	DEM_was	Jean

'That whom I met, that was John.'

Moltmann (2013) compares the patterns illustrated in (46)-(47) and (48)-(49) above to restrictions observed in English regarding the use of *this* and *that* as "presentational" demonstratives (as used in what are often described as "identificational" sentences following Higgins 1979).[28] That is, a human referent may be referred to using the bare demonstratives, but only if a VARIABLE-VALUE relationship holds between the demonstrative and the post-copular expression. Consider the following examples (the relevant context would be a situation involving a man unacquainted to the addressee within sight):

(50) English (Moltmann 2013: 52)

a. **That is mayor of Cambridge*.

b. *That is the mayor of Cambridge*.

The variant involving a (bare) nominal predicate in (50)(a), which arguably ascribes a property to a referent, is unacceptable.[29] By contrast, the variant in (50)(b), involving a referential expression in post-copular position, is fully acceptable. This is so, Moltmann (2013) argues, because in (50)(b) a property is not being assigned to a referent. Rather, an object of perception (referred to by the demonstrative *that*) is being

[28] The difference between identificational and specificational sentences (in accounts making a distinction between both types) is that the VARIABLE expression (i.e., the expression to be identified) refers to an entity within sight of the addressee (e.g. *That is Mary* or *That woman is Mary*). This fact may render a characterization of the relation between to expressions in terms of clarifying who or what the pre-copular expression refers to intuitively awkward (the addressee "knows", in a certain sense, who or what is being referred to).

[29] Except perhaps in the sense of *That (wretched thing) is mayor of Cambridge.*

identified by means of the referential expression *the mayor of Cambridge*. The author suggests that the relation holding between the "presentational" pronoun in an "identificational" sentence such as (50)(b) and a referential expression is analogous to the relation holding between the VARIABLE and VALUE expressions in sentences (more) commonly described as specificational (p. 64). Moltmann (2013: 55) notes that "presentational" pronouns such as *that* as used in (50)(b) are described in some accounts as "property-referring". Recall that, as pointed out in Section 2.1.4 in the first part of this chapter, in some accounts, specification is conceived of as "inverse predication". The principle is the same, the difference being that in some accounts the term specificational is reserved for sentences involving a VARIABLE expression with descriptive content.

Note that Moltmann (2013) explicitly argues that although "identificational sentences are not generally considered special cases of specificational sentences" there is a connection between "identificational" and specificational sentences that "demands a unified analysis" (p.44). Moltmann (2013) argues that it is better to think of "presentational" pronouns (e.g. *that* in contexts we are discussing now) as referring to "property instances". Moltmann (2013) characterizes their referential status in terms of double denotation: The expression denotes at once an object of direct perception and a referent. A similar account is given in Davidse & Van Praet (2019) for the VARIABLE expression in specificational sentences involving VARIABLE expressions with descriptive content. The authors characterize a VARIABLE expression as involving "dual reference" (in reference to Ward & Birner 1995). In the terms of Davidse and Van Praet (2019), the "coded referential meaning is the introduction of a generalized instance (Langacker 1991, 2000), but reference is also implied to an actual instance" (p. 29). In this sense, the VARIABLE expression in a specificational sentence may be considered referential but its status is different from that of a fully referential expression. As suggested in 2.1.3, the VARIABLE expression is (for good reasons) frequently associated with expressions involving descriptive content but it can also be a proper name. Consider the following example:

(51) German (Moltmann 2013: 68)

Maria M,	*das*	*könnte*	*Anna*	*oder*	*aber*	*auch*	*Susanne*	*sein.*
Maria M	that.N	could	Anna	or	but	also	Susanne	be

'Mary M, that could be Ann or else Sue.'

In the following I will discuss data from languages which seem to pattern quite closely with the English, German, and French data presented earlier.

2.3.1.2 Hebrew pronominal copulas

In Modern Hebrew there is a verbal copula, but it is not used in the present tense. In the present tense, a pronominal form occurs in some sentences involving non-verbal predicates. In many accounts these forms are described as copulas. The same forms, however, can be used as free pronominals. Consider the following examples involving left-dislocated nominal expressions:

(52) Modern Hebrew (Shor 2020: 1810)

pnina,	*hi*	*ohevet*	*ledaber.*
Pnina	she	love.PRS.SG.F	talk.INF

'Pnina, she loves to talk.'

(53) Modern Hebrew (Shor 2020: 1810)

pnina,	*hi*	*hayta*	*nora*	*xamuda.*
Pnina	she	be.PST.3SG.F	awfully	cute.F

'Pnina, she was awfully cute.'

Note that the analysis of the copula-like use is not universally agreed upon. Thus, for instance, Shor (2020) argues for a subject-doubling analysis in the use commonly analyzed as copular.

There are two (pronominal) copula paradigms, both inflecting for person and number. I will refer to the elements of the two paradigms as the H-forms (masculine/unspecified singular *hu*, feminine singular *hi*, masculine plural *hem*, feminine plural *hen)* and Z-forms (masculine/unspecified singular *ze*, feminine singular *zot*)[30] respectively. The H-forms are also used as distal demonstratives and person pronouns. The Z-forms are also used as proximal demonstratives. The Z-forms are not (normally) compatible with prototypical (property or relation-denoting) non-verbal predicates. The H-forms are in principle compatible with any non-verbal predicate and may be used in identity statements (but apparently not in specificational sentences). Consider the following examples illustrating the use of the H-forms:

(54) Modern Hebrew (based on Sichel 1997: 296)

a. *rina* *hi* *talmid-a.*
Rina H.3SG.F student-SG.F

b. *rina* *hi* *xaxam-a.*
rina H.3SG.F intelligent-SG.F

c. *rina* *hi* *ba-bayit.*
Rina H.3SG.F at-home

d. *rina* *hi* *xaver-a* *sheli.*
Rina H.3SG.F friend.SG.F mine

e. *rina* *hi* *giveret* *kohen.*
Rina H.3SG.F Mrs. Cohen

'Rina is a student/ intelligent/ at home/ my friend/ Mrs. Cohen.'

Note that the obligatory status of the H-form varies. It seems to be obligatory in a sentence such as (54)(e) where two referential expressions are involved. Not so, apparently, in other cases. In its use with adjectives, it may distinguish a state vs. individual level predication, as reported by Falk (2007). Consider the contrast between (55)(a) and (b).

(55) Modern Hebrew (Falk 2007: 20)

a. *ha-dinosaur* *hu* *šikor*.
the-dinosaur H.SG.M drunk.SG.M
'The dinosaur is a drunkard.'

[30] Concerning Z-forms, Heller (2005) notes that "the masculine plural form *ele* and the feminine plural forms *ele* and *elu* have not been fully grammaticalized as […] copulas, so a sentence containing them is very marginal as a pseudocleft and instead reads as a left dislocation" (p. 206). The author shows the following example (translated as a cleft):

(i) Modern Hebrew (Heller 2005: 206)
ma *še-dekart* *maca* *ele/elu* *hoxaxot* *le-kiyum* *ha-el.*
what that-Descartes found Z(PL) proof.SG.F to-existence the-god
'What Descartes found was proof of God's existence.'

Note that in some accounts the copular status of Z-forms is rejected altogether (but not necessarily that of H-forms).

b. *ha-dinosaur šikor*.
the-dinosaur drunk.SG.M
'the dinosaur is drunk.'

The distribution of Z-forms is different. It is reported that these cannot be normally used with non-referential terms in post-copular position (see exceptions at the end of this section). Thus, the use of the Z-form as a copula is (normally) ruled out with nominal predicates (56)(a), adjectives (b), and locative predicates (c):

(56) Modern Hebrew (based on Sichel 1997: 296)

a. **rina* *zot* *talmid-a.*
rina Z.3SG.F student-SG.F

b. **rina* *zot* *xaxam-a.*
Rina Z.3SG.F intelligent-SG.F

c. **rina* *zot* *ba-bayit.*
Rina Z.3SG.F at-home

d. *rina* *zot* *xaver-a* *sheli.*
Rina Z.3SG.F friend.SG.F mine

e. *rina* *zot* *giveret* *kohen.*
Rina Z.3SG.F Mrs Cohen

'Rina is (*a student)/ (*intelligent)/ (*at home)/ my friend/ Mrs. Cohen.'

It seems that the crucial factor determining the choice between H- and Z-forms is the referential status of the expressions preceding them. Predicational sentences and identity statements, where the pre-copular expression is fully referential require the H-form. In identity statements (57)(a) obligatorily, and in predicational sentences (57)(b) optionally:

(57) Modern Hebrew (Shirtz 2014: 30)

a. *dani* *(*hu*) *mar* *kohen.*
Dani H.3SG.M Mr Cohen
'Dani is Mr. Cohen.'

b. *dani* (*hu*) *nexmad/rofe/al ha-gag.*
Dani H.3SG.M nice/doctor/on the-roof
'Dani is nice/ a doctor/ on the roof.'

The obligatory status of the H-form in (57)(a) and the optionality in (57)(b) resembles a pattern also observed in (Standard) Arabic (see e.g. Moutaouakil 1989). A possible functional explanation would be that the co-occurrence of an indefinite nominal expression (or an adjective, or a locative) with a referential expression may be more readily interpreted as involving a subject-predicate relation.

I will turn now to the distinction between the use of H- and Z-forms as reported in Heller (2005) beginning with a minimal pair involving two proper names in pre-copular position. The sentence in (58)(a), involving a H-form, illustrates a predicational sentence, the one in (58)(b), with a Z-form, is reported in Heller (2005) to obtain a specificational reading:

(58) Modern Hebrew (Heller 2005: 171)

a. *dan* *hu* *ha-more* *šeli* *le-karate.*
Dan H.SG.M the-teacher mine to-karate
'Dan is my karate teacher.'

b. *dan* *ze* *ha-more* *šeli* *le-karate.*
Dan Z.SG.M the-teacher mine to-karate
'Dan is my karate teacher.'

According to Heller (2005), the use of the Z-form in (58)(a) favors a predicational reading (and allows an identity statement reading). Under the predicational reading, the definite description following the H-form (*ha-more šeli le-karate* 'my karate teacher') is used to ascribe a property (that of being the karate teacher of the utterer of the sentence) to an individual. According to Heller, the use of the Z-form is "appropriate as a clarification in a context where the proper name has been used and has turned out not to be known to the addressee – in this context, [(58) (b)] identifies the less familiar referent of the proper name with the more familiar referent of the description" (p. 172). In this sense, (58)(b) can be described as specificational.[31] The relation is not such that the definite description contains a VARIABLE and the proper name provides a VALUE but the other way around. That is, of the point is not revealing the identity of the karate teacher, but rather, revealing the identity of some person called Dan.

Consider now the following sentence, where the choice between a H- and a Z-form in copula function would, according to Heller (2005) distinguish a (specificational) cleft and a sentence involving a predicational sentence involving a (fully referential) free relative as a subject term:

(59) Modern Hebrew (Heller 2005: 74)

[*ma* *še-dekart* *maca*] *hu/ze/zot* *hoxaxa le-kiyum ha-el.*
what that-Descartes found H.SG.M/Z.SG.N/Z.SG.F proof(F) to-existence the-god
'What Descartes found is/was a proof of god's existence.'

In the predicational variant (involving the H-form), the sentence asserts that what Descartes found (whatever it was) can be used to prove the existence of God. In the specificational variant (involving a Z-form, either the default or inclusive masculine, glossed here as neuter), the expression *hoxaxa le ki-yum ha-el* 'a proof of god's existence' is not used to ascribe a property to the referent of the pre-copular expression but is a referential expression clarifying the reference of the VARIABLE pre-copular expression (in this case, a free relative).

Heller notes that a further distinction is possible. The Z-form may be used in the default (formally masculine) gender, or it may agree in gender with the expression following. Heller (2005) argues that the choice between a generic-masculine and a gender-congruent form may signal the distinction between specific and non-specific reference of the VALUE expression:

(60) Modern Hebrew (Heller 2005: 71)

a. [*ma* *še-dan* *mexapes*] *ze* *jirafa vruda.*
what that-Dan seeks Z.SG.N giraffe.F pink.F
'What Dann seeks is a pink giraffe.' (*de dicto* or *de re*)

[31] Or, as noted earlier, "descriptively identifying" (or "identificational") in accounts reserving the term specificational to sentences where the VARIABLE expression involves descriptive content.

b. [*ma* *še-dan* *mexapes*] *zot* *jirafa* *vruda.*
what that-Dan seeks Z.SG.F giraffe.F pink.F
'What Dann seeks is a pink giraffe.' (*de re* only)

If I understand Heller's (2005: 71) argumentation correctly, the use of the default (formally masculine) demonstrative form would allow both a reading involving a specific pink giraffe (albeit one assumed not to be identifiable by the hearer, hence the absence of a definite marker), or an arbitrary (non-specific) pink giraffe. The use of the agreeing demonstrative form allows only a specific reading and precludes a non-specific (*de dicto*) one.

There is also a use of the Z-form noted in the literature which allows a predicational (ascriptive) relation between the pre-copular and post-copular terms. This involves cases where the pre-copular expression is used in cases where "the original denotation of the subject is 'widened' to some contextually supplied property involving the original denotation" (Greenberg 2008: 169). Note that there is no agreement of any sort with the pre-copular expression in (61)(a) below. The form used of the demonstrative element is a default masculine singular regardless of the gender and number of the pre-copular expression.

(61) Modern Hebrew (Shirtz 2014: 2)
a. *yeladim* *ktanim* *ze* *nexmad.*
child.PL.M small.PL.M Z.SG.M nice.SG.M
'(Raising/having) small children is nice.'

b. *yeladim* *ktanim* *hem* *nexmadim.*
child.PL.M small.PL.M H.PL.M nice.PL.M
'Small children are nice.'

The relation between the (widened) denotatum of the expression in sentence-initial position and the property-denoting expression following the Z-form in (61)(a) is predicative but the subject expression refers by metonymy to some state of affairs involving the type of entities denoted by the nominal expression.

2.3.1.3 Japanese/Korean coding of the cleft clause

The coding possibilities of cleft clauses in Korean and Japanese cleft sentences may be characterized as reflecting the special referential status of the VARIABLE expression. The basic pattern in both languages involves a sentence with a copula in clause-final position preceded by two nominal expressions, one of which involves an oriented nominalization (the basic word order in both languages is SOV). There are two alternative ways in which the cleft clause may be formed.

In Korean, an oriented nominalization consists of a clause with a verb taking a suffix signaling its dependent status. Verbs thus marked are described as adnominal forms. The clause is followed by a noun, which, in the general case, must be sortally compatible with the kind of entity the construction as a whole denotes (i.e. a headed relative clause construction with a head noun). In cleft clauses, there is a choice between a noun with a very general meaning (*kes* 'thing') and a noun corresponding to the kind of entity the clausal nominalization denotes (e.g. *salam* 'person'In relative clause constructions, the noun *kes* is often described as a "formal noun". It can also, however, be used as a common noun meaning 'thing'. The choice between the relevant forms is illustrated in the following examples involving an interrogative and a declarative cleft:

(62) Korean (Kim 2016: 91)

a. [*i* *seysang-eyse* *ceyil* *alumtaw-un* *salam/kes*]-*un* *mwukwu-ci?*
this world-LOC most beautiful-ADN person/KES-TOP who-Q
'Who is the most beautiful (person) in the world?

b. [*ku* *il-ul* *ha-l* *swu* *iss-nun* *salam/kes*]-*un* *ne-ppwun-i-ta.*
the work-ACC do-? can exist-ADN person/KES-TOP you-just-COP-DECL
'the person/one who can do the work is just (=only) you.'

In principle, both the common noun *salam* 'person' or *kes* may be used. Note, however, that the choice is only possible in this configuration. In predicational sentences (i.e. involving either verbal predicates or non-referential nominal predicates) a noun sortally congruent to the denotatum of the relative clause construction must be used. Thus, in a predicational sentence, a relative clause construction denoting a non-human entity is compatible with *kes* (63)(a) but not one denoting a human (63)(b).

(63) Korean (Kim & Sells 2013: 110)

a. [*John-i* *sa-n* *kes*]-*un* *acwu* *pissa-ta.*
John-NOM buy-ADN KES-TOP very expensive-DECL
'What John bought is very expensive.'

b. *[*John-i* *kyelhon* *ha-n* *kes*]-*un* *alumtap-ta.*
John-NOM marry do-ADN KES-TOP beautiful-DECL
'The one who John married is beautiful.'

A remark often found in the literature is that a sentence such as (63)(b) would be pragmatically not unlike an English sentence such as *What John married is beautiful* with a (possibly) derogatory effect resulting, as it were, from referring to a person as a thing. This is not the case in specificational sentences, where a *kes*-marked relative clause construction can be felicitously used without triggering any such effects. Consider the following sentences.

(64) Korean (Kim & Sells 2013: 111)

a. [*John-i* *kyelhon* *ha-n* *kes*]-*un* *Mina-i-ta.*
John-NOM marry do-ADN KES-TOP Mina-COP-DECL
'The one who John married is mina.'

b. [*John-i* *kyelhon* *ha-n* *kes*]-*un* *uysa-i-ta.*
John-NOM marry do-ADN KES-TOP doctor-COP-DECL
'The one who John married is a doctor.'

c. *[*John-i* *kyelhon* *ha-n* *kes*]-*un* *uysa-ka* *toy-ess-a.*
John-NOM marry do-ADN KES-TOP doctor-NOM become-PST-DECL
(intended) 'The one who John married became a doctor.'

Kim & Sells (2013: 111) argue that, while a sentence such as (64)(b), where the immediately pre-copular expression is interpreted as indefinite might be taken to involve a predicational relation, the sentence is in fact specificational (like the one in (64)(a)). Korean, unlike English, does not obligatorily signal the definite vs. indefinite status and a bare noun can be interpreted either way. But even under an indefinite interpretation, the sentence can be argued to express a specificational relation, and not a predicational one (as in (64)(c)), where the use of *kes* heading the relative clause construction, would, in the authors' terms, lead to "total unacceptability". Perhaps a better English translation for (64)(b) would be *What John married was a doctor* (with no derogatory implications at all), as the translation chosen by the authors, lacking tense

congruence between the verb in the relative clause construction and the copula, hardly allows a specificational reading in English.[32]

A particularity of the *kes* vs. common noun alternation is that it is only possible in a configuration which may be the basic one in clefts in the language. That is, with the VARIABLE expression preceding the VALUE. In the reverse order, the use of *kes* is not compatible with human referents:

(65) Korean (Kim 2016: 19)

**Mini-ka*	[*John-i*	*kyelhon*	*ha-n*	*kes*]-*i-ta*.
Mini-NOM	John-NOM	marry	do-ADN	KES-COP-DECL

(Intended: 'Mini is who John married.')

The sentence in (65), with a noun compatible with the referent in question (perhaps *salam* 'person' or some noun with a meaning similar to 'woman') could in principle allow a number of different readings: predicational (ascribing the property of having married John to Mimi), identity statement, or specificational (with either expression as VARIABLE or VALUE; see discussion on word order in Chapter 4). With *kes*, the sentence is unacceptable under any intended reading.

The conditions under which a similar alternation can take place in Japanese seem to be virtually identical to those in Korean. The basic structure of a cleft sentence is very similar, the major difference being that in (modern standard varieties of) Japanese, specialized suffixes marking the subordinate (adnominal) status of predicates are restricted to a limited class of words (roughly corresponding to some adjectives in English). Another difference is that the element *no*, with which a common noun following a subordinate (adnominal) clause may alternate, is not used as a common noun. (It resembles the common noun *mono* 'thing' but the (historical) relation between *mono* and *no* is apparently not clear.) Otherwise, what was said about Korean above applies to Japanese as well.[33] Harada (2018: 49) discusses the relevant patterns. Consider the contrast between a verbal sentence involving a relative clause construction as an oblique argument (66)(a) and a cleft (66)(b).

(66) Japanese (Murasugi 1991: 149-150; cited in Harada 2018: 49)

a.

Taroo-wa	[*asoko-de*	*tabe-te-orare-ru*	*hito/*no*]-*to*	*hanasi-o*	*si-ta*
Taro-TOP	there-at	eat-STAT-HON-NPST	person/*no*-with	talk-ACC	do-PST

'Taro talked to the person/one who is eating there.'

b.

[*asoko-de*	*tabe-te-orare-ru*	*no*]-*wa*	*Tanaka*	*sensee*	*desu*
there-at	eat-STAT-HON-NPST	*no*-TOP	Tanaka	professor	COP.HON

'The one who is eating there is Prof. Tanaka.'

---

[32] The translation suggested with a relative clause construction introduced by *what* (instead of *the* one) seems to be fully acceptable in English. It seems, however that the person referred to may not be a subject. Consider the following sentences from Ross (1995: 30):

(i) *What she married was a cosmonaut.*

(ii) **What married her was a cosmonaut.*

Apparently, such restrictions do not apply in Korean.

[33] Save for some particularities concerning the use of *no* in immediately pre-copular position, which is, interestingly, reported to allow only referential readings (in contrast e.g. to a common noun such as *mono* 'thing', which allows both referential readings and non-referential ones (see Komagata 1996 for details).

Harada (2018) argues that the fully referential expression (in square brackets) marked by the postposition (or case marker) *to* in (66)(a) cannot be formed with *no* but only with a common noun such as *hito* 'person', compatible with a human referent. In principle, a referential nominal expression formed with *no* is possible, but such a construction would imply either a derogatory intention, or perhaps familiarity (possibly expressing endearment), which would in turn be incompatible with the honorific form of the verb in the subordinate clause. Given the VARIABLE status of the expression in (66)(b), no such conflict is observed. A similar account is given in Seraku (2011). The author points out (referring to Kuroda 1992) that relative clause constructions involving the marker *no* may have derogatory implications or express familiarity (p. 221). This effect, however, is absent in clefts. Consider the contrast in the following sentences:

(67) Japanese (Seraku 2011: 221)

a. [*nai-ta* *no*] *wa* *nige-ta*
cry-PST *no* TOP run-PST
'The person who cried ran away.'

b. [*nai-ta* *no*] *wa* *Tom* *da*
cry-PST *no* TOP Tom COP
'It is Tom that cried.*'*

Since there are no honorific forms involved, there is no clash anywhere so both sentences are in principle acceptable. But Seraku points out that the verbal sentence in (67)(a) can be said to express familiarity or endearment (p. 221). To judge form Seraku's account this could be paraphrased perhaps as *the poor thing that was crying ran away* or something of the sort. By contrast, in the cleft sentence in (67)(b) there is no connotation of either familiarity or disdain. Consider now the sentences in (68) below. They are variations of the sentences in (67)(a) and (b) where the free relative is left-dislocated and picked up anaphorically by a pronoun:

(68) Japanese (Seraku 2011: 222)

a. [*nai-ta* *no*] **sore/kare* *wa* *nige-ta*
cry-PST *no* it/he TOP run-PST
'The one who cried, he ran.'

b. [*nai-ta* *no*], *sore/*kare* *wa* *Tom* *da*
cry-PST *no* it/he TOP Tom COP
'The one who cried, it was Tom.'

According to Seraku (2011: 222), the demonstrative *sore* cannot be used to refer back to the dislocated expression involving *no* in the verbal sentence (68)(a) and the masculine personal pronoun *kare* cannot be used in the specificational sentence (68)(b). In both cases, the left-dislocated expression *nai-ta no* 'the one who cried' is associated with a male person, but only in the specificational configuration it is referred back to using the neuter/inanimate demonstrative in a manner reminiscent of the patterns in the German and French examples in (46) and (47), where the use of the pronominal forms *das* and *ce* were reported to be compatible with the post-copular expression 'my aunt' (in *that's my aunt*) but not with 'beautiful' (in *that's beautiful*, referring to a woman).

2.3.2 Problems in distinguishing clefts from nominal-predicate sentences

In this section I will address the problem concerning constructions in which a sentence resembling a cleft involves an expression in predicative position which, however, resembles formally a nominal predicate

rather than a (referential) nominal expression (and thus does not have the form which would be expected of a term in an equational sentence).

As pointed out in 2.2.1, (non-referential) nominal predicates may be (in some languages) formally distinct from (referential) nominal expressions. In a language marking this distinction consistently, the distinction between a nominal-predicate sentence and an equational one should be straightforward. The distinction was illustrated in examples (27) repeated below for convenience as (69):

(69) Cebuano (Dryer 2007: 235)

c. *Ang* *duktur* *ang* *babayi.*
TOPIC doctor TOPIC woman
'The woman is the doctor.'

d. *Duktur* *ang* *babayi.*
doctor TOPIC woman
'The woman is a doctor.'

The assumption is that (69)(a) is an equational sentence and (b) can only be a predicational sentence where the bare noun *duktur* 'doctor' is used to ascribe a property to the subject expression. Consider now the following sentences:[34]

(70) Cebuano (Dryer 2007: 235)

a. *Nag-tawag* *ang* *babayi* *nakuq.*
SBJ.FOCUS.DUR-call TOPIC woman 1.SG.NONTOPIC
'the woman was calling me.'

b. *Babayi* [*ang* *nag-tawag* *nakuq*].
woman TOPIC SBJ.FOCUS.DUR-call 1SG.NONTOPIC
'the one who was calling me was a woman.'

The sentence in (70)(a) is a canonical verbal sentence in the language. The verbal predicate occurs in canonical clause-initial position and is followed by its arguments, which (except for pronouns such as *nakuq*, and demonstratives) must be introduced by an article (as in *ang babayi*). Now, a nominal expression introduced by an article (as well as a pronominal form) may occur in predicate position, like *ang duktur* does in (69)(a). In this case, there is an equational sentence. Assuming the generalization holds, the expression in predicate position in (70)(b) must (since it is a bare noun) be a (non-referential) nominal predicate. If this is indeed the case, (70)(b) is not an equational sentence (and therefore not a specificational one either). Thus, in principle, the sentence in (70)(b) should be taken to convey the information that, whoever was calling the utterer of this sentence was female. Note that *babayi* can be freely used as a modifier in the language in this sense: *babayi nga iro* 'female dog', where the word for 'woman' (or 'female') and the word for 'dog' (or 'canine') are linked by the particle *nga* forming a modifier-modified expression. Now, some sentences cited in the literature would rather seem to convey a specificational relation. Consider the following example taken from Kaufman's (2018) account on clefting in western Austronesian languages:

[34] Dryer (2007) glosses the prefix *nag* as "subject focus". Note that the term "subject focus" corresponds to the more commonly used term 'actor voice' (see discussion on Tagalog in 3.3.4) The forms glossed NONTOPIC are more commonly referred to as 'genitive' in the relevant literature.

(71) Cebuano (Constantino 1965; cited in Kaufman 2018: 220)

maŋga	[*aŋ*	*ginka:ʔun*	*han*	*bataʔ*]
mango	NOM	eat.PV.PFV	GEN	child

'It was the mango that the child ate.'

The English translation of (71) arguably does not suggest a predicational relation. This problem is not addressed in Kaufman's (2018) account, where the specificational vs. predicational opposition is not treated in terms of a referential vs. non-referential opposition but rather in terms of relative referentiality (i.e. "more referential" vs. "less referential" (p. 212)).[35] Unfortunately, I am not aware of any accounts directly addressing the problem concerning the status of sentences such as (71) in Cebuano or in Philippine languages generally. It may well be the case that sentences involving bare nouns in predicate position are ambiguous, allowing both predicational and specificational readings.

There are accounts in the literature reporting a rather clear-cut distinction between nominal predicates and referential nominal expressions which, however, turn out not to hold (or not very clearly) in particular contexts. Clefts (or cleft-like configurations) seem to be particularly problematic in this respect. I will focus on accounts for two languages. Upper Nicola Okanagan (Lyon 2013) and Rapa Nui (Kieviet 2017).

2.3.2.1 Upper Nicola Okanagan

The problem to be discussed for Upper Nicola Okanagan (Southern Interior Salish, Salishan), based on the account in Lyon (2013), is one that concerns Salish languages generally (see Kroeber 1999 for an overview) and for other typologically similar languages such as Kwak'wala (Northern Wakashan), discussed in Littell (2012). In many of these languages referential expressions are regularly introduced by determiners and nominal predicates lack these markers. In sentences in which a nominal occurs in predicate position, it is sometimes difficult to decide for a predicational or specificational analysis on the basis of the presence (or absence) of an article (putatively) marking an expression as referential. In principle, the problem concerns any language in which (referential) nominal expressions are (reported to be) consistently marked in a way distinct to (non-referential) nominal predicates and where both types of expressions may occur in predicate position, like the Philippine languages discussed in the previous section (this is not always the case, see discussion on Nahuatl in 4.2.4 in Chapter 4). In some Salishan languages such as Nicola Okanagan, the use of articles is reported to mark neither definiteness nor specificity (but only referentiality). Thus, constructions described as equational sentences may be (and are sometimes described as) pragmatically equivalent to nominal-predicate sentences. I will begin the discussion of Lyon's (2013) account with a few remarks about features of the language relevant to the understanding of the problem at hand.

Argument (and oblique) nominal expressions in Upper Nicola Okanagan (except for proper names, pronouns, demonstratives and possessees) are reported to consistently be marked by an article. There are two articles in the language: *iʔ* and *t*. The discussion will concern the use of the former only as expressions marked with the article *t* (described as oblique article) cannot occur in predicate position. Nominal predicates, by contrast, do not involve the use of an article. According to Lyon (2013), the article *iʔ* does not mark either definiteness or specificity. A nominal expression (which may involve a lexical noun or a nominalization) preceded by the article *iʔ* may receive a definite interpretation (either in the sense of

[35] Kaufman (2018) explicitly associates the article ang with definiteness rather than referentiality (or even specificity) "The precise semantics of *ang* has been debated in the literature. […] I maintain that definiteness, rather than specificity, best captures the pragmatic function of *ang."* (p. 219). Here, the author refers to the use of *ang* in Tagalog, not in Cebuano. However, the example (71) is also shown in Tagalog (and other Philippine languages) and translated as 'It was the mango which the child ate'.

referring to a unique or a familiar referent). It may also receive an indefinite-specific interpretation (referring to a particular entity not assumed to be identifiable by the addressee), a non-specific interpretation, or generic one. Lyon (2013) notes, however, that the article *iʔ* "carries an implicature of uniqueness/maximality" (p. 142f) (which the article *t* does not) but that this implicature can be easily cancelled.

Lyon (2013) makes a distinction between equation (corresponding to equational sentences in my terminology) and direct predication (i.e. sentences involving nominal predicates). In the former case, two expressions introduced by the article *iʔ* (save the exceptions mentioned earlier) occur in a subject-predicate-like relationship and in the other, the predicate expression is not introduced by an article. Sentences with (bare) nominal predicates are characterized as being used to ascribe a property to a referent. The contrast is illustrated in the following examples involving two alternative answers to a question:

(72) Upper Nicola Okanagan (Lyon 2013: 252)

a. *swit iʔ səxʷ- k̓ʷúl̓-łxʷ-əm ha Spike k̓əm Wilford?*
what DET OCC-make-house-MID YNQ Spike CONJ Wilford
'Who is the carpenter, Spike or Wilford?

b. *Wilford iʔ səxʷ- k̓ʷúl̓-łxʷ-əm*
Wilford DET OCC-make-house-MID
'Wilford is the carpenter.'

c. *#Wilford səxʷ-k̓ʷúl̓-łxʷ-əm*
Wilford OCC-make-house-MID
'Wilford is a carpenter.'

The variant in (72)(c) would, according to Lyon (2013), be judged odd. This, presumably, in more or less the same sense that its English translation would be judged inappropriate as an answer to the English translation of (a). That is, ascribing the property to a referent would not directly answer the question at hand. A brief remark concerning word order is relevant here. In the equational sentence in (72)(b), word order is fixed due to the kind of expressions involved. In equational sentences, a proper name (or a demonstrative) necessarily precedes a lexical nominal expression. In a predicational sentence such as (72)(c), word order is in principle free. Inverting the order of the subject (*Wilford*) and the (determiner-less) nominal predicate would still yield a predicational sentence (see e.g. examples (73)(c-d) below).

Now, apparently, a question which may be assumed to be more appropriately answered with a predicational sentence, as in the case of (73)(c) or (d) below, may be felicitously answered using an equational sentence as in (b):

(73) Upper Nicola Okanagan (Lyon 2013: 255)

a. *stiṁ Mary aʔ c-k̓ʷúl̓-st-sʔ*
what Mary DET CUST-make-CAUS-3SG.ERG
'What does Mary do for work?

b. *Mary iʔ səxʷ-ṁaʔ- ṁáyaʔ-m*
Mary DET OCC-RED-teach-MID
'Mary is a teacher'

c. *Mary* *səxʷ-ṁaʔ- ṁáyaʔ-m*
Mary OCC-RED-teach-MID
'Mary is a teacher'

d. *səxʷ-ṁaʔ- ṁáyaʔ-m* *Mary*
OCC-RED-teach-MID Mary
'Mary is a teacher'

Lyon (2013) argues that the felicity of a sentence such as the equational one in (73)(b) involves a "pseudo-predicational reading [which is] pragmatically equivalent to a direct predication […] yet semantically equative" (p. 254). It is not completely clear, however, what "semantic equation" entails. As mentioned earlier, Lyon notes that the use of the article *iʔ* allows definite, indefinite, non-specific and generic interpretations. Lyons shows that the following sentences can be translated into English using a definite and an indefinite article:

(74) Upper Nicola Okanagan (Lyon 2013: 237)
a. [*ixíʔ*] [*iʔ* *pəptwínaxʷ*]
DEM DET old.lady
'She is an/the old lady.'

b. [*axáʔ*] [*iʔ* *ṗínaʔ*]
DEM DET birch.bark.basket
'This is a/the basket.'

c. [*John*] [*iʔ* *səxʷ-mrím-əm*]
John DET OCC-medicine-MID
'John is a/the doctor.'

d. [*iʔ* *sqəltmíxʷ*] [*iʔ* *səxʷ-píx̌-əm*]
DET man DET OCC-hunt-MID
'The man is a/the hunter.'

Without the article *iʔ*, the expressions in second position in (74)(a-d) would, according to Lyon, unambiguously be used to ascribe a property to the referent of the expression in first position. In the case of (74)(d), for instance, the sentence could be used to assert that a man's role or occupation is that of a hunter. Consider now Lyon's account of what a sentence such as that in (74)(d) could (and could not) mean. Compare the example in (75). (Note that in principle both present or past interpretations are possible.)

(75) Upper Nicola Okanagan (Lyon 2015: 238)
iʔ səxʷ-píx̌-əm iʔ sqəltmíxʷ
DET OCC-hunt-MID DET man

(i) 'The hunter is a man.'

(ii) #'The hunter was a man.'

(iii) *'The man is a hunter.'

(iv) *'The man was a hunter.'

To begin, Lyon notes that in sentences involving two article-marked nominal expressions word order is in principle flexible unless there is an asymmetry in the referential status of the expressions involved. In this case, the "more referential" one must precede. Thus, the readings suggested in the English translations of

(iii) and (iv) are ruled out. Now the question arises as to how the relation between the "more referential" and the "less referential" expression might be interpreted. The author notes that the (odd) variant in (ii) "is only interpretable under the somewhat unusual reading that the referent of the DP *iʔ* s*əx*^{w}*píx̌əm* 'the hunter' was, but is no longer 'a man'" (p. 239). This explanation could of course be described as "pseudo-predicational". Perhaps, however, the "less referential" expression is not used to refer at all, but to ascribe a property to a referent. The case here is arguably more difficult to reconcile with a notion of equation than the intuitively more plausible use of a (possibly) equational sentence to answer a question concerning a referent's occupation, as in (73)(b) above. Now, apparently, the use of s*əx*^{w}*píx̌əm* 'hunter' as a determiner-less nominal predicate is possible. As far as I can see, Lyon does not point out what the difference would be between a pseudo-predicational and a predicational one beyond the fact that in the former case an equational and a (pseudo-)predicational reading is possible and in the other only a predicational one. I will now turn to clefts, where, in a sense, the problem is the converse. That is, there are sentences with a (bare) nominal predicate that seem to be used specificationally.

Lyon (2015) distinguishes clefts (which he considers a type of equational sentences) from sentences involving a nominal predicate and a relative clause construction as a subject. Clefts in Upper Nicola Okanagan, as described in Lyon (2013), consist of two expressions preceded by the article *iʔ* (again, unless the clefted constituent is a demonstrative, a proper name or an expression marked as possessed). Either of the terms (typically the clefted constituent) may be (optionally) preceded by the demonstrative *ixiʔ*. Lyon glosses this demonstrative as CLEFT when preceding the VALUE expression in a cleft sentence. Note, however, that the demonstrative *ixiʔ* may be used in a similar way in simple equational sentences (i.e. it is not only used in cleft sentences). The following example illustrates the construction:

(76) Upper Nicola Okanagan (Lyon 2013: 298)
Context: I saw a deer and a bear on a stroll through the woods today, you did not see the bear but rather the deer.

(*ixiʔ*)	*iʔ*	*sɬ̓aʔcínəm*	*iʔ*	*wik-ənt-x*^{w}
CLEFT	DET	deer	DET	see-DIR-2SG.ERG

'It's the deer you saw.'

Consider now the following example illustrating the use of a construction involving a determiner-less nominal followed by a free relative clause:

(77) Upper Nicola Okanagan (Lyon 2013: 310)

a.

stiṁ	[*iʔ*	*wik-ənt-x*^{w}]
what	DET	see-DIR-2SG.ERG

'What did you see?'

b.

x̌^{w}*ɬ̓iʔ*	[*iʔ*	*wik-ən*]
mountain.goat	DET	see-[DIR]-1SG.ERG

'A mountain goat is what I saw.'

The question arises as to whether the relation holding between the two expressions is necessarily predicational (ascribing a property to a referent) which happens to be pragmatically acceptable (i.e. involving, as it were, a "pseudo-equational" reading as the English translation suggests). It must be pointed out that (as far as I can see) all the examples described as "cleft" sentences in Lyon's account (2015) involve definite referents. I will turn to this issue now.

As mentioned above, arguments in verbal sentences require a determiner irrespective of their referential status (i.e. definite-specific, indefinite-specific, non-specific, or generic). Thus, expressions (possibly) referring in the same way as the expression *x̌ʷλ̓iʔ* 'goat' in (77)(b) must be introduced by an article in a sentence such as the following:

(78) Upper Nicola Okanagan (Lyon 2013: 138)

wik-ən	*iʔ*	*x̌ʷλ̓iʔ*	*l*	*nɬq̓ɬməlx,*	
see-[DIR]-1SG.ERG	DET	mountain.goat	LOC	Quilchena	
uɬ	*wik-ən*	*iʔ*	*x̌ʷλ̓iʔ*	*l*	*spáx̌mən*
CONJ	see-[DIR]-1SG.ERG	DET	mountain.goat	LOC	Spáx̌mən

'I saw a mountain goat in Quilchena, and I saw a mountaing goat in Spáx̌mən (Douglas Lake).'

The author adduces the sentence in (78) to illustrate how the strong "implicature of maximality/uniqueness" carried by expressions marked by the article *iʔ* may be cancelled. In the case at hand, given the implausibility of the referents of the nominal expressions referring to a single animal due to the distance between the sighting sites involved. To clarify: the use of definite articles in the English translation would be unacceptable unless the mountain goat referred to was a single, perhaps discursively salient mountain goat.

Now, it is imaginable that the absence of the article in a "bare cleft" (in the sense of Kroeber 1999) construction such as (77) above is used contrastively to avoid a definite reading. It would not require an article because it would not be an argument of a predicate, which would be clear given the nominal status of the expression with which it co-occurs. In other words, an expression with a referential status comparable to that of the nominal expressions in (78) would not require an article in a context such as (77). In the former construction, the article would mark referentiality (or argument status), in the latter, it would be used to mark definiteness. A similar state of affairs will be described for Rapa Nui in the following section.

2.3.2.2 Rapa Nui

In Rapa Nui, as in Eastern Polynesian languages generally, equational sentences and nominal-predicate sentences are distinguished by the use of two different morphemes preceding the expression in predicate position.[36] In Rapa Nui, the marker *ko* is used in equational sentences and the marker *he* marks nominal predicates. Kieviet (2017) describes *ko* as a "prominence" marker and *he* as a "predicative determiner". In a certain sense, the marker *ko* could be described as a copula (though perhaps *he* less so). In a way, the alternation between *ko* and *he* could be described as distinct copularization patterns distinguishing equational and nominal-predicate sentences. Note that the use of *ko* and *he* is not restricted to equational and specificational sentences.[37]

It is important to note that argument expressions in Rapa Nui are generally introduced by a determiner. There is one general article *te,* used with common nouns, and a personal article *a* used with proper names. It seems that the use of the article *te* is in principle obligatory (or, omissible only under certain more or less well-defined conditions). According to Kieviet (2017), the article *te* in Rapa Nui may be best described as a marker of referentiality (somewhat like the article or determiner *iʔ* in Okanagan Salish discussed earlier). It, does not mark either definiteness or specificity. Definiteness, however, may be marked in the language.

[36] Basic word order in Eastern Polynesian languages is predicate-initial. In Rapa Nui, this pattern may also be considered as basic or neutral, but word order is to some extent flexible.

[37] Note also that elements with the shape *ko* and *he* in verbal clauses are analyzed (glossed) also as aspect markers in Kieviet (2017).

A definite vs. indefinite distinction may be signaled by combining the general article *te* with a postnominal demonstrative for definiteness, or with the numeral *e tahi* 'one' for indefiniteness respectively (Kieviet 2017).

The following examples illustrate the use of the "prominence marker" *ko* in a simple equational sentence:

(79) Rapa Nui (Kieviet 2017: 453)

Te	*me'e*	*o*	*te*	*pā'ena*	*'uta*	*ko*	*tō'oku*	*māmā*	*era.*
ART	thing	of	ART	side	inland	PROM	POSS.1SG.INAL	mother	DIST

'That (person) on the inland side is my mother.'

In the sentence in (79), a (complex) nominal expression involving the article *te*, a common noun (*me'e* 'thing') and a further nominal expression introduced by a preposition is followed by a nominal expression introduced by the marker *ko*. Note that the expression following *ko* involves an *o*-class possessive pronoun of the *t*-series. The first distinction corresponds roughly to that between alienable and inalienable possession (the *o*-class marking the latter). The second distinction concerns the referential status of the pronoun. The forms of the *t*-series may be analyzed as combinations of the general article *te*, a possession marker and a personal pronoun. Thus, in principle, the string following the marker *ko* is a full-fledged referential expression. Note the presence of the distal demonstrative *era* marking the expression as definite. Kieviet (2017: 453) notes that *ko*-marked nominal expressions in equational sentences are always marked with a demonstrative, and that they are all preceded by the article *te* or an equivalent form (in the author's terms, by a *t*-determiner).

The example in (80) below illustrates a *ko*-marked cleft sentence. Such sentences exhibit the same basic characteristics as simple equational sentences. (As far as I can see, however, in all the examples in Kieviet (2017), the *ko*-marked expression occurs in sentence-initial position).

(80) Rapa Nui (Kieviet 2017: 461)

Ko	*te*	*nūna'a*	*era*	*'a*	*'Ōrare*	[*te*	*nūna'a*	*i*	*rē*].
PROM	ART	group	DIST	of.AL	Orare	ART	group	PFV	win

'(in a report about a music contest:) Orare's group was the group that won.'

In principle, the only difference between a simple equational/specificational sentence and a cleft sentence in Rapa Nui is that the latter involves a nominalized clause. In the language, oriented clausal nominalization requires a supporting noun such as *nūna'a* 'group' in (80) or, very frequently, *me'e* 'thing' (see discussion in Section 3.4 in Chapter 3).

I will turn now to the discussion of nominal-predicate sentences in Rapa Nui. The example in (81) illustrates the use of the "predicative determiner" *he* in nominal-predicate sentences. The sentence illustrates two simple nominal-predicate clauses in coordination:

(81) Rapa Nui (Kieviet 2017: 252)

He	*taŋata*	*tau*	*manu*	*era,*	*he*	*poki*	*'a*	*Uho*	*tau*	*manu*	*era.*
PRED	person	DEM	bird	DIST	PRED	child	of. AL	Uho	DEM	bird	DIST

'That bird was a human being, that bird was Uho's child.'

In this case, the expressions *he taŋata* 'a human being' and *he poki 'a Uho* 'Uho's child' are not used referentially but ascribe a property to the referential expressions introduced by the (*t*-series) demonstrative *tau* and marked definite by the distal demonstrative *era* (i.e. *tau manu era* 'that bird'). Note that the

expressions introduced by *he* lack an article (or a *t*-series determiner). These apparently do not co-occur with the "predicative determiner" (hence Kieviet's (2017) analysis of the morpheme as a determiner).

Note that nominal predicates introduced by *he* may be marked much like nominal expressions (except for the absence of an article or a *t*-series determiner). Consider the example in (82) below, where the noun *científico* is followed by the numeral 'one'. As mentioned above, this is a regular pattern which may be used to (overtly) mark the indefinite (non-uniquely identifiable) status of a nominal expression.

(82) Rapa Nui (KIeviet 2017: 252)

A	*Thor*	*Heyerdahl*	*he*	*científico*	*e*	*tahi*
PROP	Thor	Heyerdahl	PRED	scientist	NUM	one

'Thor Heyerdahl was a scientist.'

In his description of the language, Kieviet (2017) describes some constructions which are construed on the model of nominal-predicate sentences as clefts. Consider the second of the sentences in example (83). (The first sentence is a verbal sentence with a subject expression in sentence-initial position. In this sentence, an element with the shape *ko* is analyzed as an aspect marker.)

(83) Rapa Nui (Kieviet 2017: 461)

Te	*ŋā*	*poki*	*ko*	*'ite*	*'ana*	*o*	*ruŋa*	*i*	*te*	*me'e*	*o*
ART	PL	child	PRF	know	CONT	of	above	at	ART	thing	of

ta'ato'a	*o*	*te*	*naonao*
all	of	ART	mosquito

'The children know everything about the mosquitoes,

'*e*	*he*	*pa'ari*	[*te*	*me'e*	*i*	*ta'e*	*'ite*].
and	PRED	adult	ART	thing	PFV	CONNEG	know

and the adults are the ones who don't know.'

In the second sentence in (83), the VALUE expression (i.e. *he pa'ari* 'adults') is coded like a nominal predicate. The relation between this expression and the nominalized clause (i.e. 'the ones who don't know', lit. 'the things that don't know') is arguably not ascriptive. Now, Kieviet (2017) states that "identifying [*ko*] clauses are used when the predicate refers to a unique individual which is accessible to the hearer; in other cases, classifying [*he*] clauses are used." (p. 461). This, however, does not fit well with the general account concerning the marking of referentiality distinctions discussed by the author.

Kieviet (2017) very explicitly argues that expressions marked with the article *te* (as well as other *t*-series determiners) are in principle referential (but neither necessarily definite nor even specific). On the other hand, the use of the "predicate determiner" *he* is mainly associated with (non-referential) nominal predicates. Kieviet (2017: 242) notes that in some other Polynesian languages (e.g. in the Tongic and Samoic branches) cognates of *he* may be used as indefinite articles. This seems not to be the case in Eastern Polynesian generally.[38] In Rapa Nui, *he* cannot be used to mark an argument in a clause (whatever its specificity status). The use of the marker *he* to mark indefiniteness in argument expressions is reported by Kieviet (2017) to be ruled out in Rapa Nui, as illustrated in the following examples:

[38] See a similar point for Tahitian *e*, described as a predicative particle in Lazard & Peltzer (2000: 37). The authors comment also on the "severe constraints" on the use of *he* as an article in Maori and Hawaiian.

(84) Rapa Nui (Kieviet 2017: 237)

*He	oho	he	taŋata	ki	te	hare
NTR	go	PRED	man	to	ART	house

'A man went home.'

(85) Rapa Nui (Kieviet 2017: 237)

*Ko	tike'a	'ā	a	au	(i)	he	honu.
PRF	tike'a	CONT	PROP	1SG	ACC	PRED	turtle

'I have seen a turtle.'

Note, however, that *he* is not only used as a marker of nominal predicates in the language. It is also used marking expressions in some contexts such as following (some) prepositions, introducing free standing nominals, following negation markers, and introducing nominal expressions in constructions involving predicates with the meanings 'become' and 'exist' (Kieviet 2017: 242ff). In these cases, expressions introduced by *he* are not used to ascribe properties (but they are not used as regular arguments either).

Topicalization is a context in the language in which the use of *he* seems to regularly show a parallel behavior to that discussed with respect of the cleft in example (83) above. Kieviet (2017) notes that in topicalization, "[o]ccasionally, a preverbal subject is introduced by the nominal predicate marker *he*." The author notes that "[t]his is somewhat surprising, as *he* normally introduces non-referential noun phrases and is limited to nominal predicates and other non-argument NPs" (p. 410). Topicalization, as defined by Kieviet (2017: 407), involves the placement of an expression in pre-verbal position and overt marking with the markers *ko* and *he.* The following sentence illustrates topicalization involving definite expressions.

(86) Rapa Nui (Kieviet 2017: 407)

[*Ko*	*tū*	*hoi*	*era*]	*i*	*e'a*	*haka'ou*	*ki*	*ruŋa*.
PROM	DEM	horse	DIST	PFV	go_up	again	to	above

'(The horse and the boy fell.) The horse stood up again.

[*Ko*	*tū*	*poki*	*era*	*i*	*hiŋa*	*era*	*ki*	*raro*]	*he*	*rerehu*	*rō*	*atu*	*'ai*
PROM	DEM	child	DIST	PFV	fall	DIST	to	below	NTR	faint	EMPH	away	SUBS

The boy who had fallen down fainted.

Topicalization of non-specific (e.g. generic) and indefinite referents follows a similar pattern, but the marker *he* is used:

(87) Rapa Nui (Kieviet 2017: 411)

He	*nu'u*	*pa'ari*	*'ina*	*e*	*ko*	*aŋa*	*tahaŋa*	*nō*	*i*	*a*	*koe*
PRED	people	adult	NEG	IPFV	NEG.IPFV	do	aimlessly	just	at	PROP	2SG

i	*te*	*aŋa*	*i*	*mana'u*
ACC	ART	work	PFV	think

'Grown-ups don't simply do the work they think of (i.e. without preparation).'

(88) Rapa Nui (Kieviet 2017: 411)

He	*taŋata*	*he*	*oho*	*he*	*ruku*	*i*	*te*	*ika*	*mo*	*te*	*hora*	*kai*
PRED	man	NTR	go	NTR	drive	ACC	ART	fish	for	ART	time	eat

'(The women sat down to cook the food.) The men went diving for fish for lunch.'

The sentences in (87) and (88) nicely illustrate the contrast between the use of *he* introducing non-specific and indefinite topicalized expressions (*he nu'u pa'ari* 'the adults', and *he taŋata* 'the men' respectively) in

contrast to those with (arguably) comparable referential status (i.e. generic/non-specific and indefinite) in argument position (*te aŋa* 'the work' in (87) and *te ika* 'fish' in (88)).

Kiviet (2017) notes that

> *Ko*-marked topicalisations refer to individuated entities (a single referent or a clearly defined group) which are accessible to the hearer; in other words, the exact referent of the *ko*-marked subject has been introduced in the preceding context. By contrast, the *he*-marked subjects in the examples above are not accessible as individuated entities. Even though the hearer can infer their existence from the context, they have not been mentioned as such. […] The difference between *ko* and *he* in marking topicalised subjects is reminiscent of the use of *ko* and *he* with nominal predicates […]: in both cases, *ko* marks an accessible, individuated entity, while in other cases *he* is used. (Kieviet 2017: 412)

It seems that parallels can be drawn to solve the problematic status of "bare clefts" in Salish and Philippine languages discussed earlier (and other languages, see discussion in 4.2.4 in Chapter 4 on Nahuatl). Perhaps one should only be wary of (possibly simplistic) conclusions with respect to what linguistic signs convey, especially if they are used in different contexts.

2.3.3 On tense-aspect congruence and case marking on the clefted constituent

It has been noted in the literature that, in some languages, the availability of a (specificational) cleft interpretation may depend on the compatibility of tense-aspect of the verb in the cleft clause (or, to put it more neutrally, of the verb in the term involving a clausal nominalization) and that of the copula. Consider the following examples:

(89) English (Akamajian 1970: 169)

a. [*What you are holding in your hand*] *is* *a small brown butterfly with spots on its wings*.

b. [*What you are holding in your hand*] *was a small brown butterfly with spots on its wings*.

Akamajian (1970) remarks that "only [(89)(a)] informs us as to precisely what is being held; [(89)(b)] does not specify what is being held, but merely indicates that the object in question once had certain properties" (p. 169). Thus, only (89)(a) may receive a specificational interpretation. The sentence in (89)(b) may not. It is unambiguously predicational.[39] Note, however, that tense-aspect congruence is not an absolute requirement even in languages in which it plays a role. Whatever the tense of the verb in the cleft clause, a present tense copula allows a specificational reading (Declerck 1988: 81-83; Den Dikken 2017: 219).

Case marking may trigger similar effects. It has been pointed out in the literature that case-marking of the VALUE expression corresponding to that required by the verb in the relative clause construction permits a specificational reading only. Consider the following German examples. The sentence in (90)(a), involving an accusative-marked VALUE expression, is reported not to allow a predicational reading, while the sentence in (90)(b) does.

(90) German (Iatridou & Varlokosta 1988: 6; cited in Den Dikken 2017: 219)

a.	[*Was*	*Hans*	*essen*	*wollte*]	*war*	*einen*	*Apfel*
	what	Hans	eat	wanted	was	an.ACC	apple
b.	[*Was*	*Hans*	*essen*	*wollte*]	*war*	*ein*	*Apfel*
	what	Hans	eat	wanted	was	an.NOM	apple

[39] Some accounts distinguish specificational and predicational clefts. As was pointed out in 1.2.2 in Chapter 1, in this dissertation I do not call the latter clefts.

'What Hans wanted to eat was an apple.'

Den Dikken (2017: 219) notes that it is important to clarify that the restriction in interpretation does not apply in both directions. That is, while the sentence in (90)(a) does not allow a predicational reading, the one in (b) does not preclude a specificational one (note, however, that Iatridou & Varlokosta (1988: 6) seem to suggest it does). Den Dikken (2017: 33) points out that languages may vary with respect to the very possibility of case-marking variation in analogous configurations and mentions that in Dutch, a post-copular expression is consistently marked nominative (on this issue, see also Section 4.2.2 on Lithuanian in Chapter 4).

In Spanish, case marking (by means of a preposition) on both the VARIABLE and VALUE expression seems to be the norm (see discussion in 5.2.2 in Chapter 5). Consider the examples culled from the internet in (91), involving differential object marking by means of the preposition *a*.

(91) Spanish (www)

a. [*Al* *que* *viste*] *fue* *a* *mi* *gemelo*.
DOM.SG.M REL see.2SG.PST be.3SG.PST DOM.SG.M 1SG.POSS twin.SG.M
'The one you saw was my twin brother.'

b. [*Al* *que* *viste*] *es* *a* *su* *gemelo*.
DOM.SG.M REL see.2SG.PST be.3SG.PRES DOM.SG.M 3SG.POSS twin.SG.M
'The one you saw is his twin brother.'

Note that, as illustrated in example (91)(b), tense-aspect congruence may be neutralized. (The verb in the cleft sentence is in the past tense but the copula in the present.) As in the case of the German sentence in (90)(a), both of the sentences in (91) may only obtain a specificational interpretation. In some languages, case may be only exceptionally marked. Seraku (2013), for instance, reports that case marking of clefted constituents in Japanese only occurs under special circumstances. Clearly, the tense-aspect and case patterns as described above apply in principle to languages with verbal copulas marking tense-aspect distinctions (and case marking for nominal expressions). Note that in some accounts (e.g. Creissels 2021), case marking of the VALUE expression is taken to reflect a process of language change through which clefts tend to become distinct from specificational sentences (see related discussions in 1.3.2. in Chapter 1, 4.2.2 in Chapter 4, and throughout Chapter 5).

2.4 Summary

The main purpose of this chapter was to clarify the notion of 'specificational sentence'. A specificational sentence was characterized as a type of equational sentence. An equational sentence was in turn characterized as a sentence consisting of two nominal expressions in a subject-predicate-like relation. It was proposed that equational sentences may be regarded as fundamentally different from nominal-predicate sentences. The crucial difference is that in an equational sentence both of the terms involved are used referentially. A nominal predicate, though possibly formally identical to a referential nominal expression is not used to refer but to ascribe a property to the referent of a nominal expression. My understanding of the notion of reference is made explicit in 2.1.2.

Two types of equational sentences were distinguished: specificational sentences and identity statements. In specificational sentences, one nominal expression (the VALUE expression) is used to specify the reference of another (the VARIABLE expression). In identity statement sentences two co-referential nominal expressions occur in a subject-predicate-like relation but none is used to specify the reference of the other.

Rather, the communicative point of an identity statement is to assert that two nominal expressions refer to the same referent without there being asymmetry in terms of referentiality between the terms.

Though nominal-predicate sentences and equational ones may be formally identical in some languages, in some cases the two types may be formally distinct. Three cross-linguistically recurrent differences between the two sentence types (pointed out in Dryer 2007) were discussed. The distinguishing features are the following:

- word-order patterns
- copularization patterns
- formal distinction between nominal predicates and nominal expressions

In nominal-predicate sentences word order may be rigid, while word order in equational sentences may be flexible. Nominal predicate and equational sentences may differ with respect to the presence of a marker (copula) signaling a predicate-subject (or predicate-subject-like) relationship between two terms in a sentence. It may be the case that a copula is present (or obligatory) in equational sentences but not in nominal-predicate sentences. In some languages, different markers may be used in nominal predicate and in equational sentences (see discussion on Thai and Akan/Twi in 2.2.2.1 and 2.2.2.2 respectively, and Modern Hebrew in 2.3.1.2; see also details of the discussion on Upper Nicola Okanagan, and Rapa Nui in 2.3.2.1 and 2.3.2.2). In some languages, nominal predicates may also be formally distinct from referential nominal expressions. Referential nominal expressions may (obligatorily or optionally) involve determiners or other adnominal markers while nominal predicates may lack such marking. As discussed in 2.3.2.1 and 2.3.2.2 with respect to data from Upper Nicola Okanagan and Rapa Nui, the distinction between nominal-predicate sentences and equational sentences (especially of the specificational variety) on formal grounds may be problematic, as markers which may signal referentiality in some environments (e.g. in association with arguments in verbal sentences) may mark definiteness in equational sentences.

Though, as argued in 2.1.6, the distinction between specificational sentences and identity statements is regarded as being in principle a matter of communicative intent, linguistic phenomena have been pointed out in the literature which may be considered as reflexes of (or evidence for) the special referential status of the VARIABLE expression in specificational sentences. I follow the view that the VARIABLE expression is best regarded as referential (but see discussion in 2.1.4 on an "inverse-predication" approach to specificational sentences). The referential status of the VARIABLE expression, however, is not the same as that of the VALUE, which may be described as fully referential. There are phenomena that point out to the special referential status of variable expressions. Pronominal forms anaphorically linked with VARIABLE expressions pattern differently from those associated with fully referential nominal expressions. The data examined included tag questions and resumptive pronouns in left dislocation constructions in English, German and French (2.3.1.1), patterns involved in the copula-like use of pronominal forms in Modern Hebrew (2.3.1.2), and patterns involved in the formation of oriented clausal nominalizations in Korean and Japanese (2.3.1.3). In all these cases a contrast between the treatment of fully referential expressions and VARIABLE expresions can be observed. The common denominator is that while fully referential expressions require associated pronominal forms exhibiting sortal congruence (e.g. distinguishing forms acceptable for human vs. non-human referents), VARIABLE expressions allow (and in some cases require) the use of pronominal forms neutralizing the relevant distinctions.

In Section 2.3.3, I briefly discussed the role of tense-aspect and case congruence in distinguishing clefts from nominal-predicate sentences involving oriented clausal nominalizations as subject terms. I mentioned in passing that the latter phenomenon is associated in some accounts with historical processes through which clefts tend to diverge from simple specificational sentences. This divergence may be regarded as symptomatic of a process of sentence simplification through which a bi-clausal structure develops into a mono-clausal one, as argued in Harris & Campbell (1995), mentioned in Section 1.3.2 in Chapter 1.

3 Oriented clausal nominalizations

The purpose of this chapter is to clarify the notion of oriented nominalization, expanding on the definition provided in the introduction: a nominal expression describing a referent in terms of a state of affairs it is involved in. The chapter is organized as follows: In Section 3.1, I will briefly discuss the notion of clausal nominalization as understood in the present investigation, distinguishing clausal vs. word-level nominalization on the one hand, and oriented vs. non-oriented nominalization (the latter sometimes referred to as action- or event-oriented), on the other. In Section 3.2, I will present an overview of strategies involved in oriented clausal nominalization. The enumeration of strategies I present builds upon the strategies identified in the literature for the formation of adnominal (modifying) relative clauses and strategies involved in clausal nominalization generally. In Section 3.3, I will focus on one particular strategy, namely, oriented clausal nominalization involving nominal verbal forms in the nominalized clause. In Section 3.3.4, I will discuss the case of Tagalog, a language for which it has been suggested in the literature that its basic sentence constitution pattern follows a nominal pattern. In Section 3.4, I will elaborate further upon a nominalization strategy discussed in 3.2.2, which I propose to call noun anchoring/support (and more or less boils down to the distinction between what is often discussed in terms of "headed" vs. "headless" relative clause constructions). The discussion of the noun anchoring/support strategy is particularly relevant to clarify the understanding of the notion of nominalization as understood in this dissertation. Keep in mind that, as pointed out in 1.1.3 in Chapter 1, specificational sentences featuring a VARIABLE expression involving "headed" relative clause constructions are very often not regarded as clefts. In Section 3.4, I argue against this position. The chapter concludes with a summary in Section 3.5.

3.1 Basic notions

The notion of nominalization can be understood—in the most general sense—as an operation through which an expression can be used as a nominal one. The term 'nominalization' may be used either to describe such an operation or its output (i.e. a nominal expression). The kind of nominalization which concerns us in the present discussion is clausal nominalization, that is, the morphosyntactic operation through which a clause may be used as a nominal expression. For our purposes, a clause is an expression consisting of a (morphosyntactic) predicate (typically, but not necessarily, a verb) and its arguments (and adjuncts). The notion of clausal nominalization corresponds roughly to "grammatical" nominalization (Shibatani 2019) or "syntactic" nominalization (Lehmann 2020a). This type of nominalization is to be distinguished from "derivational" or "lexical" nominalization, which would refer to an operation through which a nominal expression is derived from a lexical root or base. Instead of the terms mentioned above, I suggest the terms 'clausal' and 'word-level' nominalization.

Nominalizations (clausal and word-level alike) may be described as oriented or non-oriented.[40] An oriented nominalization is one in which the nominalized expression expresses not the action, process or state described in a nominalized expression, but an entity involved in it (a participant). By contrast, non-oriented nominalizations express an action, process, or state and, correspondingly, are sometimes referred to as action- or event-oriented nominalizations (in contrast to participant-oriented nominalizations).

The terminology used to refer to clausal nominalizations (on which I will concentrate in the following) varies in the literature. Oriented nominalizations correspond to what is commonly referred to as relative

[40] The concept and the term (*Gerichtetheit* 'orientation') is attributed by Lehmann (2020b) to Kaznelson (1974: 221-229). Note however that in Kaznelson's account the notion of nominalization does not play a major role.

clause constructions. Non-oriented nominalizations roughly correspond to what is commonly described as complement clauses, clausal arguments, or substantive clauses. All these terms are prone to cause confusion and may be used in rather different senses. For the purposes of the present discussion, it is crucial to clarify the notions of 'relative clause' and 'oriented nominalization'.

Relative clauses are typically described in the literature as subordinate clauses whose function is to restrict the reference of a nominal expression. They are regarded as modifiers within a phrasal or syntagmatic expression whose head is a noun. Understood in this sense, a relative clause is not viewed as a nominal expression but rather as part of one. One may, following Dryer (2013), shift perspective and consider a relative clause construction as a single syntagmatic unit involving both a noun (if present) and a relative clause as two sub-constituents of a larger construction. The noun provides sortal information about the referent (i.e. what kind of entity the nominal expression refers to) and the relative clause describes a state of affairs in which the referent is involved. Understood in this sense, the notion of 'relative clause construction' is in principle identical to that of 'oriented clausal nominalization'. In the following section I will discuss strategies used in different languages to form such constructions.

3.2 Nominalization strategies

To a considerable extent, the strategies illustrated in the following overlap with those described for relative clauses (e.g. in Lehmann 1984, Hendery 2012) and with those found in accounts of (clausal) nominalization (e.g. Shibatani 2019). The strategies are the following:

i. Zero/minimal marking
ii. Noun anchoring/support
iii. Relative pronoun
iv. Nominal classifier
v. Article
vi. Article/ general subordinator
vii. Relative marker
viii. Nominal verbal form

As will become clear, in some cases the distinctions between strategies are not clear-cut but gradient. The strategies to be discussed below are not to be understood in any way as mutually exclusive nor must they be the only strategy available in a given language. Many languages exhibit one or more of the following strategies (alone or in combination). I will, however, attempt to illustrate each of the strategies identified in something close to a pure form. Strategies i through vii will be presented in the following sub-sections. The cross-linguistic variation observed in the use of nominal verbal forms (Strategy viii) will be discussed in Section 3.3.

3.2.1 Zero/minimal marking

A nominalization may involve no overt (or only minimal) marking. In this case, a nominalized clause may be distinguished from an independent clause by virtue of functioning as an argument of a superordinate predicate. Consider the Tilapa Otomí (Otomian, Otomanguean) examples in (1) and (2). The first one is a cleft, the second one is a sentence in which an oriented nominalization (also in square brackets) is the subject argument of an intransitive verbal clause.

(1) Tilapa Otomí (Palancar 2021: 149)

keh=a	*a*	*Papa*	*Boniphasio*	[*bi=kha*]=*'a*
COP.AS=CL	DEF.SG	Pope	Boniphace	[3]COMPL=do[3OBJ]=3SG

'It was Pope Boniphace that did it."
(Lit. "Pope Boniphace was who did it.')

(2) Tilapa Otomí (Palancar 2021: 148)

hin=á	*za*	[*rati*	*kha*]
NEG=ICMPL.STAT	be.good	[3]ICMPL	do

'What she does is not good.'

The structure of the clauses marked in square brackets in both cases is virtually identical to that of an independent clause. In the sentence in (1), a clitic pronoun (=*'a*), cross-indexing the referent subject argument, marks the cleft clause (not the verb itself) as a nominal predicate. (In Otomí, Palancar argues, the cleft clause is formally not the subject, but the complement argument of the copula.) In the sentence in (2), the nominalized clause follows a verbal predicate. The construction is almost identical to two adjacent independent clauses. It can be, according to Palancar (2021: 147), distinguished from two contiguous independent clauses on prosodic grounds and by the (apparently not obligatory) presence of subject-cross-referencing enclitics (marking the entire nominalized clause as a term).

3.2.2 Noun anchoring /support

A relative clause construction may require a noun as anchor or support (i.e. forming a "headed" relative clause construction). This pattern is illustrated in the cleft sentences in examples (3) and (4) from Santa María Peñoles Mixtec (Mixtecan, Otomanguean), a language (distantly) related to Otomí.

(3) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 133)

na^{1}ma^{2}ni^{3}=i^{3}	*kuu^{23}*	[*xi'^{3}na^{1}*	*nu^{2}koo^{21}*]
godmother=1SG	RLS.be.ICMPL	granny	RLS.sit.ICMPL

'La abuelita que está sentada es mi madrina.'
'The old lady who is sitting is my godmother.'

(4) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 164)

xi^{2}chi^{3} aa^{12}	*kuu^{23}*	[*xi^{2}chi^{3}*	*ni^{3}-ndua^{21}*]
girl DIST	RLS.be.ICMPL	girl	COMPL-RLS.fall

'La muchacha que se cayó es esa muchacha.'
'The girl who fell is that girl.'

As in Otomí, the subordinate clause exhibits the same basic structure as an independent clause. The constructions following the copula (in square brackets) have the same shape as independent clauses with a fronted subject argument (the canonical/unmarked word order in the language is VSO).

3.2.3 Relative pronoun

Another strategy which may be involved in the formation of oriented clausal nominalizations involves the use of specialized pronominal forms corresponding to the referent expressed by the nominalized clause. Frequently, they are identical or very similar to other pronominal forms in the language, especially interrogative or indefinite pronouns. In some cases, they may develop specialized forms (cf. e.g. Greek and Hausa examples at the beginning of 3.3 and in 3.4 respectively). This pattern (also available, with some restrictions, in English) is illustrated in the following Latin example:

(5) Latin (Plautus' Mercator, ca. 200 BC; cited in Löfstedt 1966: 261)

Non	*ego*	*sum*	[*qui*	*te*	*dudum*	*conduxi*]
NEG	1SG.NOM	be.1SG.PRS	PRO.REL.M.SG.NOM	2.SG.ACC	before	hire.1SG

'It's not me who just hired you.'

The structure of the nominalized clause in (5) is in principle identical to that of an independent clause. However, the relative pronoun cannot be omitted and occurs regularly at the left edge of the construction. The relative pronoun is marked for the gender, number, and the case corresponding to the role/function of the referent in the subordinate clause. Note that the relative pronoun *qui* is in the nominative not because the relative clause construction is a constituent of a copular sentence but because it corresponds to the agent of the verb *conduco* 'bind, hire'. If corresponding to another role/function, it would take the appropriate case. A relative clause such as that in (5) may occur in combinaton with a noun or a demonstrative marked in the case corresponding to the role/function of the referent in the superordinate clause (Pinkster 2012).

As noted above, relative pronouns in Latin and many other languages are related to interrogative (and/or indefinite) pronouns. This pattern is common in Europe but is also attested in languages throughout the world (though it is often presumed to be borrowed from European languages, or develop through language contact; see e.g. Mithun 2012). Another common pattern involves relative pronouns related to demonstratives (see Hendery 2012 and references there).

In some languages in which relative pronouns otherwise do not (normally) occur, relative pronouns (or related strategies) may be used in particular configurations only. In Tilapa Otomí, for example, the use of interrogative words as a relative pronouns is reported in relative clause constructions with locative reference. The example in (6) is a cleft. In the example in (7), the location-oriented nominalization occurs in combination (it is not clear whether in apposition to, or forming a constituent) with a noun preceded by a demonstrative and a plural marker:

(6) Tilapa Otomí (Palancar 2021: 149)

keh=a=gwa	[*'abwu*	*bi=zo*]
COP.AS=CL=here	WHERE	[3]COMPL=SS/fall

'It's here where he fell.'
(Lit. 'Here is where he fell.')

(7) Tilapa Otomí (Palancar 2021: 149)

ra=kha	*yi*	*i*	*bañü*	[*'abwu*	*gugu*	*hi=^{g}wi*]
[3]ICMPL=exist	DEM.PL	PL	bath	WHERE	1.COMPL.IRR	bathe=DU[.INCL]

'These are the baths where we may have a bath.'

A similar strategy to the one observed in the Tilapa Otomí examples in (6) and (7) is used to orient a clause towards a location in Santa María Peñoles Mixtec as well. Ramírez Pérez (2014: 72) explicitly points out that this strategy is used exclusively in this case. Otherwise, an antecedent noun or a nominal classifier is required. In Santa María Peñoles Mixtec, however, the support element in question is not an interrogative word (as is the case with Tilapa Otomí *'abwu* 'where'), but a relational noun (*nuu*^{2} 'face', 'in front of'). The following sentences illustrate the use of the relational noun used much like a relative pronoun in (8) (hence the author's glossing), and the preposition-like use of the same element in (9).

(8) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 69)

ni³-x-i²ni²=i³	*be'ee²³*	[*nuu²*	*ni³-ka¹tu*	*ndi̶i̶²³*]
COMPL-RLS-see=1sg	house	PRO.REL.LOC	COMPL-RLS.lie	dead.person

'Vi la casa donde estuvo tirado el difunto.'
'I saw the house where the dead man was lying.'

(9) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 71)

ni³-xi³-nu²ni³=n	*nuu²*	*be'ee²³*
COMPL-RLS-stand=2SG	RN.in	house

'Estuviste parado en la casa.'
'You stood in the house.'

Ramírez Pérez (2014: 71) notes that *nuu²* is the only relational noun (out of seven in the language) that allows a relative-pronoun-like use. Judging from the translations provided by the author, this is the relational noun with the most general sense. Unfortunately, it is not clear from the author's account whether a construction such as the clause in square brackets in example (8) may occur without a noun such as *be'ee²³* 'house' preceding it. I will return to the general nominalization pattern in Santa María Peñoles Mixtec in the following sub-section.

3.2.4 Nominal classifier

In some languages, nominalized clauses may be marked as such by nominal classifiers, which also occcur in basic nominal expressions. This strategy is illustrated in the following examples from Yoloxóchitl Mixtec, a Mixtecan (Oto-Manguean) language related to Santa María Peñoles Mixtec discussed above. Consider the cleft sentences in (10) and (11) below. Note in (10) that the nominal classifier *tí* used for animals is used both preceding the noun meaning 'horse' and introducing the nominalized clause following the copula. In the sentence in (11), the classifier used corresponds to the class corresponding to the referent of the noun *ñu'ú* 'land' preceding the copula:

(10) Yoloxóchitl Mixtec (Guadalupe Joaquina 2007: 146)

tí	*kʷayu*	*kú*	[*tí*	*xínu*]
CLF.animal	horse	COP	CLF.animal	PROG.run

'Es el caballo el que está corriendo.'
'The horse is the one that's running.'

(11) Yoloxóchitl Mixtec (Guadalupe Joaquina 2007: 166)

ñu'ú	*kú*	[*yà*	*kúni*	*satá*	*na*]
land	COP	CLF.thing	PROG.want	POT.buy	3PL.HUM

'Terreno es lo que quieren comprar.'
'Land is what they want to buy.'

The structure of the subordinate clauses in (10) and (11) can be distinguished from that of independent clauses by the presence of a classifier on the left edge of the clause and by the obligatory omission ("gapping") of an expression denoting the entity the nominalization refers to. Compare the cleft sentence in (11) and verbal sentence in example (12).

(12) Yoloxóchitl Mixtec (Guadalupe Joaquina 2007: 166)
ñu'ú *satá* *nà* *yolo*
land POT.buy CLF.people Yolo
'Los (de) Yolo comprarán el terreno.'
'The ones from Yolo will buy the land'.

In Yoloxóchitl Mixtec, a nominal classifier can combine with a wide range of expressions (including nouns, clauses, and demonstratives) forming nominal expressions. In (12), for instance, a classifier and a place name form a nominal expression referring the people from the place in question. Note that classifiers in the language, though similar in shape to nouns with related meanings, are bound forms. Also note that in this language, apparently, a clause may not be directly attached to a noun to form a relative clause construction. Consider the examples in (13): (a) is a canonical sentence; (b) and (c) illustrate relative clause constructions. Judging from the examples in the account in Guadalupe Joaquina (2007: 201), relative clause constructions in combination with nouns appear to consistently involve the apposition of (in principle autonomous) nominal expressions.

(13) Yoloxóchitl Mixtec (Guadalupe Joaquina 2007: 140)

a. *nì* *jìini* *tà* *tauni* *îi* *isu*
COMPL COMPL.see CLF.man sir one deer
'El señor vio un venado.'
'The man saw a deer.'

b. [*tà* *tauni*] [*tà* *nì* *jìini* *îi* *isu*]
CLF.man sir CLF.man COMPL COMPL.see one deer
'el señor que vio un venado'
'the man who saw a deer'

c. [*îi* *isu*] [*tí* *nì* *jìini* *tà* *tauni*]
one deer CLF.animal COMPL COMPL.see CLF.man sir
'un venado que el señor vio.'
'a deer that the man saw.'

Guadalupe Joaquina (2007) reports that the use of classifiers in Yoloxóchitl Mixtec is "obligatory in nominal phrases when their nucleus is a proper name and a countable noun, [and] if the noun phrase functions as a subject" they are "not so obligatory [in expressions involving] non-countable nouns or if they function as objects" (p. 63).[41] Apparently, their use is obligatory in oriented clausal nominalizations.

Classifiers are also used in Santa María Peñoles Mixtec, though the inventory reported in Ramírez Pérez (2014) is smaller than that reported in Guadalupe Joaquina (2007) for Yoloxóchitl Mixtec. To judge from the examples in Ramírez Pérez (2007), classifiers are less regularly used in Santa María Peñoles Mixtec. Apparently, they are most consistently used in association with proper names. Though perhaps not as regularly as in Yoloxóchitl Mixtec, classifiers may also be used to mark clausal nominalizations in Santa María Peñoles. Consider the cleft sentence in example (14):

[41] "Los clasificadores nominales son obligatorios en las [frases nominales] cuando el núcleo de éstas es un nombre propio y contable, además si funge como sujeto de la oración. No son tan obligatorios si se trata de nombres no contables o si fungen como objetos." Guadalupe Joaquina (2007: 63)

(14) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 133)

de'e[23]	*taa*[1]	*Cha*[1]*na*[1]	*kuu*[23]	[tee[23]	ni[3]-ka[1]kwe[23]tu[3]]
son	CLF.woman	Susana	RLS.be.ICMPL	CLF.man	COMPL-RLS.pray

'El que rezó es el hijo de Susana.'
'The one who prayed is Susana's son.'

There seems to be language internal variation with respect to the use of classifiers and anchor/support nouns in nominalizations in Santa María Peñoles Mixtec. Compare the sentence in (14) above, where the classifier *tee*[23] (which also functions as a common noun with the meaning 'man' in the language) with the one in (15) below, where the common noun *to'o*[12] 'people, person' is used instead. Note that the nominal expression in sentence-initial position does not involve a common or proper noun. It consists of the classifier *tee*[23], an adjective, and a distal demonstrative.

(15) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 133)

tee[23]	*li'*[3]*li*[3]	*aa*[12]	*kuu*[23]	[*to'o*[12]	*ni*[3]*-ndua*[32]]
CLF.man	small	DIST	RLS.be.ICMPL	person	COMPL-RLS.fall

'La persona que se cayó es ese niño.'
'The person who fell down is that small boy.'

3.2.5 Article

A cross-linguistically more common strategy involves the marking of nominal (and nominalized) expressions by means of articles, that is, bound forms marking nominal expressions often signaling distinctions in referential status. These markers may also mark categories such as number and gender but convey less sortal information than nominal classifiers. This is the case for instance in Tahitian, where all nominal expressions are consistently introduced by an article. Consider the cleft sentence in (16). Both the clefted constituent and the cleft clause are introduced by the article *te.*

(16) Tahitian (Lazard & Peltzer 2000: 85)

'o	*te*	*taote*	[*te-i*	*haere*	*mai*]
ID	ART	doctor	ART-ASP	go	DIR

'C'est le docteur qui est venu.' (lit. 'Celui qui est venu est le docteur.')
'It's the doctor who came.' (lit.'The one who came is the doctor.')

Though limited with respect to TAM marking possibilities (following a "strict" subordination pattern, in the terms of Lazard & Peltzer 2000: 46), the clausal nominalization introduced by the article *te* in the sentence in (16) resembles very closely an independent clause, except for the obligatory omission ("gapping") of the expression denoting the referent towards which the nominalization is oriented.

3.2.6 Article/ general subordinator

It is sometimes difficult to distinguish an article from a general subordinator in languages in which one element (or a member of a paradigm of similar elements) may be used to introduce not only nominal expressions but practically all kinds of arguments and adjuncts. This is for instance the case in Nahuatl. In this language, the particle *in* (related to the proximal/non-distal demonstrative *īn*(*in*)) introduces simple

nominal expressions, nominalized clauses (oriented and non-oriented alike), but also all kinds of subordinate clauses (e.g. conditional, or temporal adverbial clauses).[42] Consider the following examples:

(17) Classical Nahuatl (Olmos 1875[1547]: 234; cited in Launey 1986: 406)
[*in* *tlā* *huel* *ti-c-chīhua-z* *in*], *ic* *ti-tlācamach-ō-z*
ART COND well 2SG.SBJ-3SG.OBJ-make-FUT DEM.PROX thus 2SG-obey-PASS-FUT
'If you do this well, you will be obeyed.'

(18) Classical Nahuatl (Carochi 1892[1645]: 501; cited in Launey 1986: 1342)
pācticâ [*in* *ō-mo- tēcac*]
be.well ART ANT-RFL-lie.down.PRF
'He was all well when he went to bed.'

(19) Classical Nahuatl (Codex Chimalpopoca; cited in Launey 1986: 1342)
[*in* *ō-qu-ittac* *tezcatl*], *qu-ìtô*:…
ART ANT-3SG.OBJ-see.PRF mirror 3SG.OBJ-say.PRF
'When he saw the mirror, he said…'

The use of the particle *in* introducing arguments and adjuncts involving simple nominal expressions is not obligatory (cf. example (19), where the nominal expression *tezcatl* 'mirror' occurs without the article preceding it). With subordinate clauses (of any type), its use seems to be obligatory.[43] In Nahuatl, nouns, verbs, and locative expressions may freely function as predicates. Launey (1986) points out that the presence of the article *in* introducing an expression reliably signals that the expression following it does not function as the main predicate in a sentence. Consider the sentences in (20) and (21), involving nouns in (canonically sentence-initial) predicate position. The expression following the article *in* (in square brackets) are oriented clausal nominalizations:

(20) Classical Nahuatl (Sahagún vol. II: 122; cited in Launey 1986: 1315)
zan *nō* *cihuâ* [*in* *tla-napaloâ*]
only also woman.PL ART NSP.OBJ-carry.in.arms.PL
'Also only the women carried (litters) on their shoulders.'
('The ones who carried (litters) on their shoulders were also only the women.')

(21) Classical Nahuatl (Sahagún vol. I: 39; cited in Launey 1986: 1316)
Tzapotēcâ [*in* *huel* *īn-teōuh* *catca*]
Zapotec.PL ART very 3PL.POSS-god be.LOC.PRF
'He was the proper god of the Zapotec'
('The ones whose proper god he was were the Zapotec.')

[42] See Himmelmann (1997: 207-214) on the notion of *Komplementartikel* used there to describe similar uses of articles in Papuan Languages.

[43] For an account on the many factors determining the use of *in* in Classical Nahuatl see Launey (1986: 1433ff). The author argues that with simple nominal expressions, the use of *in* may (to some extent) be associated with definiteness (though, according to Launey, more in the sense of signaling uniqueness/maximality than givenness/familiarity). Launey also discusses the use of *in* in terms of "bracketing" or "sentence partitioning", which may in some cases conflict with (override) definiteness-marking (as is sometimes the case within subordinate clauses, or in a particular sentence pattern, in which argument or adjunct expressions immediately precede the main predicate and are not introduced by *in* regardless of their referential status). Launey reconciles both aspects of the use of *in* invoking the notion of autonomy/autonomization of an expression with respect to a superordinate predicate.

In both cases, the article *in* is the only sign of subordination. The internal structure of the nominalized clauses is otherwise identical to that of an independent clause. Note that the nominalized clause in (21) involves the possessed noun *īn-tēteô* 'their god' in predicate position. In this case, the clausal status of the expression is rendered clear by the presence of the (in principle locative) verbal form *catca* which is used as a copula in the perfect aspect and by the clause-initial modal particle *huel*. In principle, both of the nominalized clauses in (20) and (21) could, in the appropriate context (e.g. in association with a superordinate verb of perception or cognition), be interpreted as non-oriented nominalized clauses (i.e. 'that they carried (litters) on their shoulders' and 'that he was their proper god' respectively).[44]

3.2.7 Relative marker

In some languages, an oriented nominalized clause is marked by elements which, unlike articles or classifiers, are neither regularly used to mark simple nominal expressions, nor have the characteristics of relative pronouns. These elements are frequently referred to as relative markers (also relative particles or relativizers). These markers may be polyfunctional (or historically related with elements marking attributive or genitive relations, or function as clausal conjunctions; for an overview of relative markers cross-linguistically see e.g. Hendery 2012). Often, clauses introduced by such markers cannot function as nominal expressions unless combined with one of the strategies mentioned earlier. For instance, a clause introduced by the general subordinator *que* in Spanish, which is etymologically related to a Latin relative pronoun (cf. example (5) above), requires, when used in relative clause constructions, to be (minimally) introduced by an article to function as a nominal expression. Consider example (22):

(22) Spanish (cf. (23) below)

[*(*El*)	*que*	*evalúa*	*el*	*rendimiento*]	*es*
ART.SG.M	REL	evaluate.3SG.PRS	ART.SG.M	output	be.3SG.PRS
el	*presidente*				
ART.SG.M	president				

'The one who evaluates the output is the president.'

As mentioned earlier (3.2.3), in Latin, an oriented nominalization could be associated with a demonstrative (the historical precursor of the Spanish definite article) but its use was facultative (Pinkster 2012: 384ff).

In some languages, elements often described as relative particles or relative markers may mark a clausal nominalization, which may be used as a nominal expression without further support. This is for instance the case with the particle *yang* in Malay/Indonesian (which may be historically derived from a pronominal form in combination with a linking particle; see Adelaar 1992). The following example illustrates the use of the marker *yang* introducing an (agent-oriented) nominalized clause:

(23) Indonesian (Kaufman 2018: 226)

[*Yang*	*menilai*	*kinerja*	*Menteri*]	*adalah*	*Presiden.*
REL	AV.evaluate	output	minister	COP	president

'(The one) who evaluates the output of a minister is the president.'

[44] Note that the status of the sentences in (20) and (21) as clefts (as opposed to nominal-predicate sentences) is problematic (see discussion on similar constructions in Upper Nicola Okanagan in 2.3.2.1 in Chapter 2). I will further treat Nahuatl data in 4.2.4 in Chapter 4, where I will discuss cleft sentences involving a VALUE expression introduced by *in*. In such cases the (sentence-initial) main predicate slot is occupied by a pronoun. (As mentioned above, expressions introduced by *in* may not occur in main predicate position.)

The particle *yang* is not a general subordinator. It may not be used to introduce complement or adverbial clauses but only relatives. Note, however, that *yang* cannot mark orientation towards a time or a location, in which case the noun anchor strategy involving the common nouns *waktu* 'time' and *tempat* 'place' must be used (Sneddon et al. 2010). Otherwise, the particle *yang* is indifferent to the kind of referent a clausal nominalization is oriented to.

The examples presented in Sub-sections 3.2.1 to 3.2.7 illustrate nominalization strategies not involving major modifications in the internal structure of the nominalized clause. Yet, in many languages, clausal nominalization involves, as mentioned earlier, a specialized (nominal) form of the predicate heading the nominalized clause. This will be the subject of the following section.

3.3 Nominal verbal forms

The use of nominal verbal forms in clausal nominalizations may be the only strategy involved in a given language or it may co-exist with other strategies. Consider the cleft sentences in (24) from Ancient (Koine) Greek (corresponding to two alternative translations from Hebrew).

(24) Ancient Greek (Banti 2013: 48)

a. *Egṓ* *eimi* [*ho* *anabibásas*
1SG.NOM be.1SG.PRS ART.SG.M.NOM. cause.to.ascend.PTCP.SG.AOR.ACT.M.NOM
humâs *eks* *Aigúptou*]
2PL.ACC from Egypt.GEN
'It is I who made you come out of Egypt.' (Judges 6.8 Codex Alexandrinus)

b. *Egṓ* *eimi* [*hòs* *anḗgagon*
1SG.NOM be.1SG.PRS REL.PRO.SG.M.NOM. lead.up.1SG.AOR.ACT
humâs *ek* *gês* *Aigúptou*]
2PL.ACC from land.GEN Egypt.GEN
'It is I who brought you out from the land of Egypt.' (Judges 6.8 Codex Vaticanus)

The predicate *anabibasás* in the cleft clause (marked in square brackets) in (24)(a) is a participle. It can be regarded as a nominal form as it is marked for gender and case. In contrast to the "finite" (or non-nominal) verbal form in (24)(b), it is not marked for person. Both forms however, mark aspect and voice. The clauses are to a considerable extent semantically equivalent. In the sense relevant for the present discussion both clauses are nominalized. The internal structure of the one in (24)(b) is like that of an independent clause except for the form of the relative pronoun *hòs* corresponding to the agent argument. Unlike a pronoun in an independent clause, it does not inflect for person (note that the verb *anḗgagon* is inflected for the first person), it cannot normally be omitted, and its position in the clause is fixed (occurring regularly in clause-initial position). In the case of the cleft clause in (24)(a), the coding of the undergoer argument and the adjunct is identical to that in (24)(b), but the form of the verb is distinct from that normally occurring in an independent clause. Furthermore, the nominalized clause is introduced by an article, like a basic nominal expression. In principle, the participial clause following the article in (24)(a) could be substituted by a (singular masculine) common noun. As for the orientation of the clauses in (24)(a) and (b), note that in the case of (a), nominalization is (inherently) oriented towards the role corresponding to the voice the participle is inflected for. In the case of (24)(b), orientation is marked by the case form of the relative pronoun.

In the following three sections, I will discuss unrelated languages in which word-level nominalization may be considered the main strategy in oriented clausal nominalization. The first two languages, Turkish and Yaqui, illustrate the use of inherently oriented forms. The third language, Ingush, presents an example of contextually oriented nominalization (in the sense of Shagal 2019). That is, where the nominalized forms of the verbs are oriented (i.e. necessarily refer to a participant involved in a state of affairs rather that to the state of affairs itself), but the same forms can be oriented to participants with a wide variety of role/functions. In 3.3.4 I will discuss Tagalog. Unlike the three languages discussed in 3.3.1 to 3.3.3, participant-oriented clausal nominalization in Tagalog does not involve the use of specialized verb forms. This language is interesting in the context of the present discussion because it has been suggested that its basic sentence constitution pattern is in principle nominal and voice-marked verbs may in principle be regarded as inherently-oriented participles.

3.3.1 Turkish: subject-oriented and non-subject oriented participles

Cleft sentences in Modern Standard Turkish consist of a nominal expression taking a predicate marker or a copula plus TAM suffix, and a participial clause (Kornfilt 1997, Erguvanlı 1984). The pattern is illustrated in the following examples:

(25) Turkish (Erguvanlı 1984: 165-166)

a. [*Bu* *mektu-bu* *yaz-an* (*sahis*)] *Ali-dir*.
this letter-ACC write-SBJ.PTCP person Ali-PRED
'Ali is the person/one who wrote this letter.'

b. *Ali-dir* [*bu* *mektu-bu* *yaz-an* (*sahis*)].
Ali-PRED this letter-ACC write-SBJ.PTCP person
'It is Ali who wrote this letter.' 'Ali is the person/one who wrote this letter.'

The canonical word order in Turkish is SOV. The sentence in (22)(a) aligns with this pattern in the sense that the term marked as a predicate occurs sentence-finally. The sentence in (25)(b) reflects a pragmatically marked pattern involving the placement of a constituent in post-predicate position. The clefted constituent in the examples in (25) is marked with the (optional) third person suffix *-DIr*, which is used to mark nominal predicates in the present tense.[45] A participial clause may be used adnominally (e.g. in combination with a noun such as *sahis* 'person', as in the examples in (25)) but may function in principle as a full-fledged nominal expression by itself.

There are two participial suffixes in Turkish. The suffix *-An* is subject-oriented and the suffix *-DIK* is used for non-subjects (the vowels alternate following vowel harmony and voicing of the consonants is conditioned by the environment). Consider the following examples:

(26) Turkish (Kornfilt 1997: 193)

a. [*Sinema-ya* *gid-en*] *Hasan-dı.*
cinema-DAT go-SBJ.PTCP Hasan-PST
'Hasan was the one who went to the movies.'

[45] In other tenses, there is a copular suffix *-y*, which is, however, not realized following consonants. Nominal predicates in Turkish inflect for tense using the same suffixes as verbal predicates (e.g. *-DI* for past tense) (Kornflt 1997).

b. [*Sinema-da* *gör-dük-ler-im*] *öğrenci-ler-im-di.*
cinema-LOC see-OBJ.PTCP-PL-1SG student-PL-1SG-PST
'My students were the ones whom I saw at the movies.

In the sentence in (26)(a) the verb in the participial clause takes the subject (or actor-oriented) participial suffix *-An* (here: *-en*) and is in an equational relationship with a proper name taking the past tense suffix. In the sentence in (26)(b) the participial clause is object- (or, non-actor) oriented, and is marked with the suffix *-DIK* (here, *-dük*). Note the parallels between the participial and the copularized nominal expression: Both take the (nominal) plural suffix (*-lIr*) and the same (first person singular) suffix-(*I*)*m*.

Participial clauses can be used as argument expressions in verbal clauses. In this case, they can take case suffixes corresponding to their role/function in the clause. In the following example the object-oriented nominalization occurs as an object and takes the accusative case suffix.

(27) Turkish (Kornfilt 1997: 63)
[*Adam-ın* *ye-diğ-in*]-*i* *al-dı-m.*
man-GEN eat-OBJ.PTCP-3SG-ACC take-PST-1SG
'I took what the man eats/ate.'

However, when followed by a noun, accusative case marking is suffixed only on the noun (Kornfilt 1997: 57). A participial clause differs from a finite one with respect to the coding of the actor argument. In (27), the notional subject of the participial clause (*adam-ın*) is in the genitive case and not in the nominative as it would be in a main clause. All other arguments and adjuncts in participial clauses, however, are case-marked according to the role/function they would have in a main clause. The following examples further illustrate the contrast between a non-subject-oriented (28)(a) and a subject-oriented (b) participial clauses (as the translations suggest, these are not full sentences but nominal expressions in isolation):

(28) Turkish (Kornfilt 1997: 63)
a. [*adam-ın* *ye-dik-ler-i*]
man-GEN eat-OBJ.PTCP-PL-3SG
'those (things) which the man eats/ate'

b. [*para-yı* *ver-en*]
money-ACC give-SBJ.PTCP
'(the person) who gave the money'

In addition to the participial pattern, Turkish has an alternative strategy to form relative clause constructions, illustrated in (29). Here, the relative clause is introduced with the relative particle *ki*. In this case, the relative clause follows a noun.

(29) Turkish (Kornfilt 1997: 60)
[*Bir* *adam* *ki* *çoçuk-lar-ın-ı* *sev-me-z*] *yalnız* *yaşa-malı-dır.*
a man that child-PL-3SG-ACC love-NEG-AOR alone live-NEG-EP.COP
'A man who does not love his children must live alone.'

Kornfilt (1997: 60) notes that this pattern is a loan from Persian (see Section 5.5.1 on clefts in Modern Persian). The author notes that its use has declined since the language reform in the first half of the 20th century. I have no information concerning its use in cleft sentences. Kornfilt (1997: 63) notes that a clause introduced with *ki* cannot function as a nominal expression without a nominal (or pronominal) expression preceding it.

Note that there is a partial formal overlap in Turkish between non-subject oriented participial clauses and non-oriented clausal nominalizations. Turkish has two suffixes for non-oriented clausal arguments: the suffix *-DIK* (formally identical with the non-subject participle) and the suffix *-mA*. The difference between these two is that the former is used in nominalized clauses denoting facts and the latter for non-factual propositions. Kornfilt (1997: 50) refers to the suffixes as "Factive Nominal" and "Action Nominal" markers respectively. The contrast is illustrated in the following sentences:

(30) Turkish (Kornfilt 1997: 51)

a.	(*Ben*)	[*Ahhmed-in*	*öl-düğ-ün*]-*ü*	*duy-du-m.*
	1.SG.NOM	Ahmed-GEN	die-FACTIVE.NMLZ-3.SG-ACC	hear-PST-1.SG
	'I heard that Ahmed died.'			
b.	(*Ben*)	[*Ahmed-in*	*öl-me-sin*]-*den*	*kork-uyor-du-m.*
	1.SG.NOM	Ahmed-GEN	die-ACTION.NMLZ-3.SG-ABL	fear-PROG-PST-1.SG
	'I was afraid that Ahmet would die.'			

Like oriented nominalized clauses, non-oriented ones such as (30) are regularly marked for the case corresponding to their function/role in the matrix clause. Note that the internal structure of the nominalized clause in these cases is also altered with respect to that of finite clauses. As in non-subject participial clauses, the actor argument of the nominalized clause is marked with the genitive case suffix. A difference between participial clauses and non-oriented clausal nominalizations is that the former can be marked for number (if denoting a plural entity) while the latter cannot.[46]

3.3.2 Yaqui: nominalization suffixes

The strategy used to mark oriented clausal nominalization in Yaqui (Sonoran, Uto-Aztecan) is quite similar to the one observed in Turkish. These languages are not related either genetically or areally but they exhibit some typological affinity. Role/function of arguments and adjuncts is signaled by suffixes and/or postpositions. Basic word order is predicate-final, both in verbal and nonverbal clauses.

Álvarez González (2012) reports that subject-oriented nominalization is marked by the suffix *-me*. The suffix *-'u* is used for orientation towards objects and obliques. Álvarez González (2012: 83) suggests that the origin of the non-subject nominalizing suffix is the allative case marker *-u* of similar form. In Álvarez González (2016), the author describes the suffix as "an old agent nominalizer with temporal/aspectual restrictions (suffix -(')u) has now developed new uses as a patient nominalizer" (p. 107). Suffixes with similar form and function as Yaqui *-me* are attested in other Sonoran languages (cf. Ramos Bierge 2014: 439). There is also a nominalizing suffix -'*Vpo* that marks locative-oriented nominalizations. This suffix corresponds to the locative suffix *-po*. Unfortunately, I do not have information concerning the use of the locative nominalizer in cleft sentences. The examples in (31) illustrate the use of actor (a) and non-actor-oriented (b) nominalized verb forms in cleft sentences:

[46] Kornfilt (1997) notes that

> When the *-DIK*-type of participial form (i.e. the OBJ.PTCP) needs to be used, it is preferred to use it with a plural morpheme (whenever appropriate) in headless relative clauses. The reason is probably because *-DIK* is also the general factive nominalization morpheme, and when it is used without the plural marker the distinction between a headless relative clause and a regular subordinate factive clause is neutralized, leading to confusion and unclarity. The attachment of the plural marker reveals the nominal nature of the relative clause. (Kornfilt 1997: 63)

(31) Yaqui (Álvarez González 2012: 86)

a. [*wa-me* *yabe-m* *tea-ka-me*] *Joan*
DEM-PL key-PL find-PFV-NMLZ.SBJ John
'The one who found those keys is John.'

b. [*Joan-ta* *tea-ka-'u*] *yabe-m-tu-kan*
John-GEN find-PFV-NMLZ.OBJ key-PL-VBZ-IPFV
'What John found were the keys.'

Álverez González (2012: 87) notes that oriented nominalized verbs preserve compatibility with TAM morphology as well as valency/argument structure. However, as in the case of Turkish, the subject in non-subject oriented nominalized clauses takes the genitive case, like a nominal possessor. This can be observed in the sentence in (31)(b) above. Incidentally, note the suffixal copularization (by the verbalizing element *-tu-*) of the clefted constituent resembling the Turkish pattern in (31)(b).

In some languages, there may be a clear distinction between what could be described as participial forms and de-verbal nouns proper (e.g. active participles vs. agent nouns). In others, there may be an overlap. This seems to be the case in Yaqui (and Álvarez González 2016 argues that the latter notion best describes the nature of Yaqui nominalizations). The suffixes *-me* and *-Vpo* are found both in clausal nominalizations and in lexical nominalizations in the language (in the case of *-Vpo*, mostly in toponyms). This is not, however, the case with the suffix *-'u*. The following examples illustrate lexical nouns derived from verbal bases with the suffix *-me*:

(32) Yaqui (Álvarez González 2012: 83)

a. *e'~etbwa-me*
RED~steal-NMLZ
'thief'

b. *majta-wa-me*
teach-PASS-NMLZ
'student'

c. *etbwa-wa-me*
steal-PASS-NMLZ
'theft'

d. *totte-me*
fold-NMLZ
'snail'

Álvarez González (2012: 83) notes that the suffix *-me* is used productively to derive nouns with verbal bases. This is the case in (32)(a) and (b) with a reduplicated (habitual aspect) verb and a passive respectively. The meaning is 'the one who'. Note, however, that the suffix can be also used with a passivized verbal base with an action/result meaning as in (32)(c). The author notes that this action noun interpretation is more common than the agent noun exemplified by (32)(b). Other, more idiosyncratic patterns such as the derivation of animal names (32)(d) are also found in the lexicon.

The marking of oriented clausal nominalizations with the suffixes *-me*, *-'u,* and *-Vpo* on the subordinate verb seems to be fully productive in the language. They can be used in apposition to other nominal expressions. Consider the following examples:[47]

(33) Yaqui (Álvarez González 2012: 89)

Joan	*uka*	*chu'u-ta*	[*enchi*	*ke'e-ka-m*]*-ta*	*mea'a-k*
John	DET.ACC	dog-ACC	2SG.ACC	bite-PFV-NMLZ-ACC	kill-PFV

'John killed the dog that bit you.'

(34) Yaqui (Álvarez González 2012: 73)

Joan	*inika*	*bachi-ta*	[*em*	*jinu-ka-'u*]	*bwa'a-ka*
John	DEM.ACC	corn-ACC	2SG.GEN	buy-PFV- NMLZ	beat-PFV

'John ate this corn that you bought.'

(35) Yaqui (Álvarez González 2012: 78)

inepo	*u-m*	*kari-m*	[*em*	*joa-ka-'apo*]	*jinu-k*
1SG.NOM	DET.PL	house-PL	2SG.GEN	live-PFV- NMLZ	buy-PFV

'I bought the houses where you have lived.'

Note that there is only case marking on the subject-oriented nominalized clause (33), where both the nominal expression *uka chu'u-ta* 'this dog' and the verb in the nominalized clause following it take the accusative suffix *-ta*. The absence of the case marker in the object- and location-oriented clauses in (34) and (35) is problematic, then, for an analysis of -'*u* and -'*Vpo* as nominalizers. Álvarez González (2012: 86) points out, however, that accusative marking is not allowed for possessed arguments (which is formally the case for object-oriented (*-'u*) nominalizations as the notional subject expression is coded as a genitive). The omission of accusative case marking with possessed arguments is illustrated in the following examples:

(36) Yaqui (Álvarez González 2012: 86)

a.

[*in*	*yaa-bae-'u*]	*ne*	*kopta-k*
1SG.GEN	do-DES-NMLZ	1SG.NOM	forget-PFV

'I forgot what I was going to do.'

b.

in	*kabuji*	*ne*	*kopta-k*
1SG.GEN	drum	1SG.NOM	forget-PFV

'I forgot my drum.'

The examples in (36) nicely illustrate the parallel treatment of a -'*u* nominalized clause (a) and a possessed nominal expression (b). Admittedly, the sentence in (36)(a) involves a verb whose English counterpart readily allows non-oriented clausal arguments. But the English translation of (33)(a) suggests an oriented clausal nominalization and the sentence in (36)(b) shows that the verb *kopta* 'forget' may take non-clausal arguments. As in Turkish, there is some formal overlap between oriented and non-oriented nominalizations. I will return to this at the end of this section but first I will discuss further nominal properties of oriented nominalized clauses.

Álvarez González (2012) points out that oriented clausal nominalizations in Yaqui behave like basic nominal expressions in their ability to take determiners:

[47] Glosses in the examples (34) and (35) are modified (REL to NMLZ), in accordance to the glossing used by Álvarez González (2012) in the rest of his article.

(37) Yaqui (Álvarez González 2012: 86)
U-me [*bwa'am-ta joa-me*] *pu'ato-m* *tapejti-po* *joa-k*
DET-PL lunch-ACC do-NMLZ plate-LOC roof.top-LOC put-PFV
'The ones who are cooking, they put the plate on the tapanco.'

(38) Yaqui (Álvarez González 2012: 86)
U [*itom* *nu'upa-ka-'u*] *kaa* *jaleki*
DET 1PL.GEN bring-PFV-NMLZ NEG enough
'What we brought is not enough.'

(39) Yaqui (Álvarez González 2012: 86)
U [*in* *tekipanoa-'apo*] *mekka* *taawa*
DET 1SG.GEN work-NMLZ far be
'(The place) where I work is far.'

As in the examples in (36), the nominalized clauses in examples (37), (38) and (39) function as arguments of the sentence-final predicates. In these cases, the nominalized clauses are introduced by the article-like morpheme *u* glossed in Álvarez González (2012) as a determiner. Following the author's account, the constructions are not to be analyzed, as it were, as "light-headed" relative clause constructions. Rather, the determiner here functions as it would with a basic nominal expression.[48] Consider the sentences in (40) below, where a demonstrative precedes the nominalized clause:

(40) Yaqui (Álvarez González 2012: 86)
a. *Junu'u* [*wekeya-me*] *nakapit*
DEM be.standing-NMLZ deaf
'This one who is standing is deaf.'

b. *Junu'u* [*in* *ta'a-ka-'u*] *nee* *alle-tua*
DEM 1.SG.GEN learn-PFV-NMLZ 1SG.ACC be.happy-CAUS
'What I learned made me happy.'

In these cases, the demonstrative could appear without the nominalized clauses. The demonstrative (*junu'u*) and the nominalized clauses in (40) are in principle two (co-referential but) independent expressions. As for the determiner (*u*) observed in the examples in (37) and (39), they do not function as nominalizing devices. That is, the nominalized clauses are already nominal without a preceding determiner. The use of the determiner seems to mark the referential status of the nominal expressions in question. Interestingly, Álvarez González (2012: 86) explicitly notes the absence of determiners in referring to examples of clefts (such as those in (31) above). The absence of determiners introducing cleft sentences could be related to the weakly

[48] But note that, apparently, the determiner would not be repeated:

(i) Yaqui (Guerrero 2012: 103)
U *jamut* [*Joan-ta* *ili* *uni-ta* *a-u* *bizzua-ka-'u*] *sii-ka*
DET woman.NOM Juan-ACC little child-ACC 3SG-DIR send-PFV-CLM leave-PFV
'The woman to whom Juan sent the child, left.'

This may pose a problem to an analysis in which the clause in square brackets is analyzed as a nominal expression, as in the account by Álvarez González (2012). Guerrero (2012) prefers not to describe these clauses as nominalizations. I will address this issue shortly.

referential status (in the sense discussed in 2.3.1 in Chapter 2) of the VARIABLE expression in specificational sentences.

As mentioned earlier, there is some formal overlap in Yaqui between oriented and non-oriented clausal nominalizations. There are, however, differences with respect to the internal structure of both types of clauses. Oriented clauses involve the gapping of the participants for subjects, direct objects and locatives, as well as the use of resumptive pronouns for indirect objects and obliques (Álvarez González 2012: 80). In non-oriented (i.e. complement) clauses, all arguments may be expressed with full nominal expressions. As in oriented nominalizations, the coding of actors in non-oriented nominalized clauses differs from that observed in main clauses, where actors occur in the (formally unmarked) nominative case. But the coding of actor arguments differs also between oriented and non-oriented nominalized clauses as well. Guerrero (2012) points out that subjects of object-oriented relative clauses are coded as genitives while those of complement clauses are coded as accusatives. A formal contrast, however, occurs only in first- and second-person pronouns as the genitive and accusative are marked identically otherwise. The contrast is illustrated in the sentences in the examples in (41) and (42) below. (Note that Guerrero glosses the suffix *-'u* as a clause-linkage marker.)

(41) Yaqui (Guerrero 2012: 116)

[*Em/*enchi*	*bwika-'u*]	*ne*	*yi'i-ne*	(UG oriented)
2SG.GEN/ACC	sing-CLM	1SG.NOM	dance.POT	

'I will dance whatever you sing.'

(42) Yaqui (Guerrero 2012: 116)

a.

Aurelia	*bicha-k*	[*tajo'o-ta*	*nim/*ne*	*baksia-ka-'u*].	(UG oriented)
Aurelia.NOM	see-PFV	cloth-ACC	1SG.GEN/ACC	wash.PFV-CLM	

'Aurelia saw the clothes I washed.'

b.

Aurelia	*ne*	*bicha-k,*	[*tajo'o-ta*	*ne/*nim*	*baksia-ka-'u*].	(non-oriented)
Aurelia.NOM	1SG.ACC	see-PFV	cloth-ACC	1SG.ACC/GEN	wash-PFV.CLM	

'Aurelia saw me washing clothes.'

A further difference Guerrero points out is that oriented nominalizations may be marked for plural while this is not possible for complement clauses:

(43) Yaqui (Guerrero 2012: 116)

a.

Min	*kaba'i-m*	*bicha-k*	[*Anselmo-ta*	*junu-ka-'u-m*].	
Fermín.NOM	horse-PL	see-PFV	Anselmo-ACC	buy-PFV-CLM-PL	

'Fermín saw the horses that Anselmo bought.'

b.

Min	[*Anselmo-ta*	*kaba'i-m*	*junu-ka-'u*]	*bicha-k.*	
Fermín.NOM	Anselmo-ACC	horse-PL	buy-PFV-CLM	see-PFV	

'Fermín saw that Anselmo bought the horses.'

The oriented nominalized clause in (43)(a) following the verb *bicha-k* 'saw' takes the same plural marker as the nominal expression *kaba'i-m* 'horses', with which it is co-referential (or which it modifies). This is not the case in (43)(b), where, arguably, the nominal expression *kaba'i-m* 'horses' may be best analyzed not as an argument of the main verb *bicha-k* 'saw', but of the subordinate *jinu-ka-'u* 'buy'. In Guerrero's analysis, the accusative-marked proper name *Anselmo-ta* would also be analyzed as an argument of the subordinate verb. The glossing of the suffix *-ta* on the proper name *Anselmo-ta* in (43)(a) as accusative is due to the form of the suffix (a pronoun would take the formally distinct genitive).

Guerrero (2012) points out that a formal overlap between oriented and non-oriented clausal nominalizations can be also observed in clauses involving the suffix *-me*. The following sentences illustrate the contrast between a subject oriented nominalized clause (44)(a) and a complement clause in (b)

(44) Yaqui (Guerrero 2012: 116)

a. *Jipi'ikim misi-ta miiika* [*pa'aku weama-m-ta*]
milk.PL cat-ACC give.PRES outside be.around-CLM-ACC
'Give milk (to) the cat that is outside.'

b. *Nim achai* [*jaibu enchi sika-m-ta*] *te'a-k*
1SG.GEN father already 2SG.ACC go.SG.PFV-CLM-ACC find-PFV
'My father discovered that you already left.'

It is not quite clear whether the clause in (44)(b) may not also be regarded as an oriented nominalization, perhaps in adnominal/appositional relationship with the accusative second person *enchi*. Guerrero notes that the use of *-me* clauses in a manner resembling complement clauses is much more restricted than that of *-'u* marked clauses.

3.3.3 Ingush: contextually-oriented participles

Oriented clausal nominalization in Ingush (Nakh-Daghestanian) is reported in Nichols (2011) to consistently make use of specialized verb forms. There are differences, however, between the nominalization strategies discussed in the previous sections for Turkish and Yaqui and the pattern present in Ingush. To start with, a formal overlap between oriented and non-oriented nominalizations does not occur. Non-oriented de-verbal forms (action nouns or masdars) exist and they are involved in clausal nominalization but they are formally distinct from oriented nominalized forms, which Nichols describes as participles. Furthermore, there is a formal distinction between participles that may be used as modifiers and "nominalized participles", which may be used as nominal expressions (Nichols 2011: 134). The latter may be described as being derived from the former by means of (further) suffixation. Another major difference concerns the orientation of participles in Ingush. Participial forms in Ingush are not inherently but contextually oriented (in the sense of Shagal 2019). That is, the same form may be oriented towards a number of different semantic roles (agent, undergoer, or a spatial or temporal location). This allows for very free clefting relations.

Clefts in Ingush are quite transparently constructed as copular sentences with a contextually oriented clausal nominalization as one of its terms. Equational and nominal-predicate sentences in the language involve a verbal copula (glossed in Nichols 2011 as 'be'). The following examples illustrate cleft sentences in the language. In (45), the cleft clause is oriented towards the actor (a person leading prayer), in (46), towards the undergoer (an animal shot at):

(45) Ingush (Nichols 2011: 309)

Hwo vac, [*yz wa-dieshazh vaaghar*] *so vy ealar*
2SG V.be.NEG 3SG DX-D.read.CVSIM V.sit.PTCP.NMLZ 1SG V.be say.WP
'It's me, not you, that reads it (the Koran) here (i.e. leads prayer)
'The one who sits reading the Koran is me' [KM]

(46) Ingush (Nichols 2011: 630)

[*aaz top tiexaar*] *jer cha jy*
1S.ERG gun strike.PTCP.NNLZ DEM bear J.be.PRS
'It's this bear I shot (at).' ('The one I shot is this bear.')

Participles are formed from verbal present and past stems plus the endings *-a* and *-cha* (for the present direct and oblique), and *-aa/-na* and *-cha* (for the past direct and oblique). To be used as nominal expressions they take the suffix *-r* or form periphrastic constructions with a nominalized participial form of the copula (Nichols 2011: 598).[49] Nichols notes that both formations seem to be semantically and syntactically equivalent. The internal structure of the nominalized clause remains unaltered with respect to case marking of arguments and adjuncts. Like verbal nouns, nominalized participles may be case marked corresponding to their role/function in the sentence. Case, however, is not directly marked but involves the extension *-chuo-* on the verbal base (p. 594). Modifying (non-nominalized) participles are not case marked. Only the head noun modified by the adjectival clause receives case marking, the adjectival participle heading the adjectival clause only receives the oblique suffix *-cha*. Consider the following example:

(47) Ingush (Nichols 2011: 588)

Wanagh	*my*	*qiera,*	[*wa*	__	*t'ehwa*	*doghacha*]	*giirienagh*	*qiera.*
winter.LAT	NEG	fear.IMPF	winter	DAT	after	D.come.PTCP.OBL	autumn.LAT	fear.IMP

'Don't fear winter, fear autumn, which follows ('comes after). (Proverb)

The participial clause (in square brackets) in example (47) cannot function as a nominal expression but only as a modifier. Note that participial forms are also used in periphrastic TAM constructions (in association with inflected forms of the copula, like English participles):

(48) Ingush (Nichols 2011: 261)

a.

Cuo	*gazet*	*diesha*	*xuddy*.
3SG.ERG	newspaper	D.read.PTCP	INFR.FUT.D

'He'll probably read the newspaper. He probably reads/ is reading the paper.'

b.

Yz	*ciga*	*vuoda*	*xugby*.
3SG	there	V.go.PTCP	INFR.FUT.D

'He'll probably go there. He must be going there/ going to go there.'

Nominalized participles express a participant involved in the state of affairs denoted by the verb they derive from. A present or past nominalized participle can have any of the interpretations suggested in the translations of the examples in (49) below. The forms are all in the nominative/unmarked case and inflected for the D gender (corresponding to the citation form).

(49) Ingush (based on Nichols 2011: 225)

a. *diesh-a-r*
D.read-PTCP-NMLZ
'the one who reads'
'the one which is read'

b. *diesh-a-r-azh*
D.read-PTCP-NMLZ-PL
'the ones who read'
'the ones that are read'

[49] Note that masdars/verbal nouns are formed with the infinitive stem and the suffix –(*a*)*r*, yielding similar, but not always identical forms to non-periphrastic nominalized participles.

c. *diish-aa-r*
D.read.PST-PTCP-NMLZ
'the one who has read'
'the one which was read'

d. *diish-aa-r-azh*
D.read.PST-PTCP-NMLZ-PL
'the ones who have read'
'the ones that were read'

Nichols (2011: 225) notes that some verbs ("chiefly […] the verb 'be' and other that can indicate location") can form participles with a locative sense such as '(the place) where…'. These forms have a "nominalized meaning" but involve no "no overt nominalizing morphology". Such forms occur in the singular only. Consider the examples in (50).

(50) Ingush (Nichols 2011: 225)
a. *Muusaa* [*sho dolcha*] *hwa=chy-vierii?*
Musa 2PL D.be.PTCP.OBJ DX=in-V.come.WP=Q
'Did Musa get to your place.' ('to where you are')

b. *Qoana* [*hwo volcha*] *qy juxa voaghargvaac yz.*
tomorrow 1SG V.be.PTCP.OBL any.more again V.com.FUT.V.NEG 3SG
'He won't come back to your place again tomorrow.'

In general, however, participles ("nominalized" or modifying) are contextually oriented. In terms of "relativization potential", this means they are subject to (almost) no restrictions. Nichols notes that "[a]ny case and nearly any syntactic role can be relativized on, as long as the antecedent is identifiable" (p. 589). The principle is illustrated below with modifying participles:

(51) Ingush (Nichols 2011: 591)
a. [*so* __ *jaaxa*] *c'aa*
1SG LOC J.live.PTCP house
'the house I live in'

b. [*hwo* __ *vea*] *shu*
2SG LOC V.born,PTCP year
'the year you were born'

c. *aaz* __ *keaxat jaazdea*] *q'oalam*
1.ERG INS letter write.D.PTCP] pen
'the pen I wrote the letter with'

d. [*so* __ *jixie leattaa*] *c'aa*
2SG DAT beside stand.PTCP house
'the building I stood next to'

e. [*zhwalez* __ *kylgaa carjg tiexaa*] *sag*
dog.ERG DAT beside tooth bite.PTCP person
'the person whose hand the dog bit'

The examples in (51) illustrate the orientation towards locatives (a), instruments (b), objects of prepositions (involving postposition stranding within the subordinate clause) (d), and possessor (e). Nichols notes that when the relativized element is not recoverable, relativization fails:

(52) Ingush (Nichols 2011: 592)
*[*suoga* __ *gour* *jola*] *sag*
1SG.ALL GEN horse J.be.PTCP person
intended: 'the person whose horse I've got'

*[*aaz* __ *kinashjkaazh* *diesha*] *park*
1SG.ERG LOC book.PL D.read.PTCP park
intended: 'the park I read books in'

Apparently, the same flexibility in orientation is found with nominalized participles. Nominalized clauses orient towards any role. The case marking corresponds to that which a basic nominal expression would take in the corresponding environment:

(53) Ingush (Nichols 2011: 595)
Muusaaz [*shiina* *dear*] *juxa-diiqaad.*
Musa.ERG 3SG.RFL.DAT D.do.PTCP.NMLZ back-D.avenge.NW.D
'Musa avenged what had been done to him.'

(54) Ingush (Nichols 2011: 595)
T'aaqqa *yz* *qiitar* [*aaz* *duucachogh*].
then 3SG understand.WP 1SG.ERG D.tell.PTCP.NMLZ.LAT
'The he (suddenly) understood what I was saying.'

(55) Ingush (Nichols 2011: 595)
Ucyga *Maalsag* [*shii* *xaannara* *bolchaazhca*] *wa-xeinuu.*
Ucyga Maalsag 3SG.RFL.GEN time.FOC.ABL B.be.PTCP.PL.NMLZ.INS DX.sit.NW.V
'Ucyga Maalsag sat with people of his own age.' ('with those of his very age')

(56) Ingush (Nichols 2011: 595)
Zhwalii [*shie* *ullachara*] *dwa-juxa* *hweira.*
dog 3SG.RFL lie.PTCP.NMLZ.ABL DX-back move.around.WP
'The dog moved around where it was lying.' (e.g. moved from side to side)

It would be expected that clefts also show this flexibility. Nichols notes with respect to (interrogative) clefts that "[t]here are almost no text examples of (what would be) oblique cases, adverbials, etc., questioned, but they could easily be elicited" (p. 627). An example of a locative (interrogative) cleft is illustrated below:

(57) Ingush (Nichols 2011: 592)
[*Sho* __ *uqaza* *deaxkaar*] *malagha* *pojezd* *jar?*
2PL LOC here D.come.PL.PTCP.NMLZ what train J.be.PST
'What train did you come on?'

To clarify: the English translation could be paraphrased as 'The one in which you came was what train?', where the declarative counterpart would be something like 'The one in which you came was that train'. The following example illustrates a copular sentence in which a nominalized clause is oriented, in Nichol's (1991) terms, towards the "the location where the nominalized verb takes place" (p. 630):

(58) Ingush (Nichols 2011: 630)

[*Handz*	*vai*	*daaxar*]	*dika*	*xa*	*yi*.
now	1PL.INCL	D.live.PTCP.NMLZ	good	time	J.be.PRS

'The times we live in are good.' ('The one we live in is a good time.')

The example is arguably predicational. According to the definition I follow, the sentence is not a cleft, but the principle is the same. As in English, the Ingush copula does not distinguish equational from predicational sentences.

3.3.4 Tagalog: nominal predication as basic sentence pattern

Oriented clausal nominalization in Tagalog involves the marking of a clause as a nominal expression by means of an article. The internal structure of the nominalized clause is in principle identical to a main clause (minus the obligatory omission of the argument toward which the nominalized clause is oriented).[50] It has been argued that the sentence constitution pattern in the language is in principle nominal. Under this analysis, there is in principle no difference between a nominal and a verbal sentence. The only difference between an equational and a basic (verbal or nominal-predicate) sentence is that the former consists of two nominative-marked expressions.

The canonical verbal sentence in Tagalog is predicate initial. All content words may function as predicates or arguments. (Parts of speech may be morphologically distinguished but non-verbal predicates do not require special marking to function as such; there is no copula in the language.) All (non-pronominal) argument and adjunct expressions are obligatorily preceded by articles (or "phrase markers", see Himmelmann 2015). The subject argument (typically occurring in sentence-final position) is preceded by the nominative article *ang* (or the corresponding form *si* for proper names). Note that the notion of 'subject' in Tagalog is to be understood in relation to the voice system. Verbal predicates may alternate between different voices, oriented towards a variety of semantic roles. The subject of a clause involving a voice-marked verbal predicate is the argument with the semantic role selected by the voice the predicate is marked for. Consider the following examples:

(59) Tagalog (Schachter & Otanes 1972: 69)

Bumabasa	[*ng*	*diyaryo*]	[*ang*	*titser*].
<um>RED-basa	ng	diyaryo	ang	titser
<AV>RED-read	GEN	newspaper	NOM	teacher

'The teacher is reading a newspaper.'

(60) Tagalog (Kaufman 2017: 603)

Kinain	[*ng*	*pusa*]	[*ang*	*daga*]
<in>kain-ø	ng	pusa	ang	daga
<RLS>eat-PV	GEN	cat	NOM	rat

'The cat ate the rat.'

[50] I will only discuss oriented clausal nominalizations involving the article-marking strategy. In constructions involving noun anchoring/support, the subordinate clause is linked to the noun by means of the morpheme *-ng/na* occurring between the noun and the subordinate clause. In these constructions, the order between the noun and the modifying clause is variable and the construction as a whole is preceded by an article). A similar pattern (minus an article) is found in constructions involving a demonstrative and a subordinate clause.

(61) Tagalog (Himmelmann 2008: 265)

Iniabot	[*ng*	*manggagamot*]	[*sa*	*sundalo*]	[*ang*	*itlog*].
<in>i-abot	ng	manggagamot	sa	sundalo	ang	itlog
<RLS>CV-reach	GEN	doctor	DAT	soldier	NOM	egg

'The physician handed the egg to the soldier.'

(62) Tagalog (Kaufman 2017: 603)

Kinainan	[*ng*	*pusa*]	[*ng*	*daga*]	[*ang*	*pinggan*].
<in>kain-an	ng	pusa	ng	daga	ang	pinggan
<RLS>eat-LV	GEN	cat	GEN	rat	NOM	plate

'The cat ate the rat on/from the plate.'

Each of the sentences in (59) to (62) feature a verbal predicate marked for a different voice. In each case, the subject is the expression introduced by the nominative article *ang* in sentence-final position. In the actor voice (59), the subject is an agent, in the patent voice (60) a patient, in the conveyance voice (61) a theme, and in the locative voice (62) a location/source.

As mentioned earlier, to function as an argument (or adjunct) a subordinate clause must be introduced by an article. The nominalized clause is necessarily oriented towards the participant selected by the voice marked on the verb. Consider the sentence in (63).

(63) Tagalog (www)

Naka-kita	*din*	*ako*	*ng*	*perlas*	[*sa*	*k<in>ain*	*ko*]*!*
AV.RLS.POT-see	also	1SG.NOM	GEN	pearl	DAT	<RLS>eat.PV	1SG.GEN

'I also found a pearl in what I ate!'

In (63) the main predicate is in the actor voice. The subject is the first-person person pronoun *ako* (pronominal subject expressions behave as clitics within the predicate complex and thus do not tend to occur sentence finally in the presence of other argument and/or adjunct expressions). A nominalized clause (in square brackets) introduced by the dative article *sa* functions as an oblique argument in the sentence and refers to whatever the utterer of the sentence ate (presumably an oyster). Note that the first-person pronoun in the clausal nominalization (where the verb is in the patient voice) occurs in the (non-subject) genitive form, like the main-clause (non-subject) patient argument *ng perlas* 'a pearl'.

A cleft in Tagalog can be defined as a sentence involving a clausal nominalization and another nominal expression in a (specificational) equational relation to each other. An equational sentence in the language consists of two nominal expressions in juxtaposition. Both expressions are introduced by a nominative article (unless one of them is a pronominal form). In principle, any semantic role that can be selected as a subject by a voice-marked verb can be clefted. Compare a canonical verbal sentence in the locative voice in (64)(a) and a cleft in (b)

(64) Tagalog (Schachter & Otanes 1972: 315)

a.

P<in>agsulat-an	*ko*	[*ang*	*desk*	*na*	*ito*].
write<RLS>-LV	1SG	NOM	desk	LK	PROX

'I wrote on this desk.'

b.

Ang desk na ito	[*ang*	*p<in>agsulat-an*	*ko*].			
NOM desk LK PROX	NOM	write<RLS>-LV	1S			

'It's this desk that I wrote on.'

The sentence in (64)(a) above is a canonical sentence in the language, albeit one that marks the location as a subject. Schachter & Otanes (1972: 315) point out that sentences involving the locative voice as a main sentence predicate such as the one illustrated in example (64)(a) are much less frequent than clefts such as that in (b), where the locative-voice verb functions as the predicate within a clausal nominalization.

Note that clefting is possible also with clausal arguments. Consider the following examples:

(65) Tagalog (Ramos & Cena 1990:163)

a. *Ang pag-kanta ni Ismael* [*ang g<um>ulat sa bata*].
NOM GER-sing GEN Ismael NOM <AV>panic DAT child
'What scared the child was Ismael's singing.'

b. *Ang pag-katok niya sa pinto* [*ang b<um>ulahaw kay Fred*].
NOM GER-knock 3.SG.GEN DAT door NOM <AV>disturb DAT Fred
'What disturbed Fred was his knocking on the door.

c. *Ang pag-bilang ni Carla nang mali* [*ang i-k<in>agalit ng customer*].
NOM GER-count GEN Carla so small NOM <RLS>CV-angry GEN customer
'What made the customer angry was Carla's counting incorrectly.'

Ramos & Cena (1990) point out that for a clause such as the ones appearing sentence-initially in (65) to "serve as a predicate […] it requires that its subject must also be a clause" (p. 163). Note that the nominalized clauses in sentence-initial position in the examples (65) are not oriented. The forms of the corresponding predicates (*pag-kanta* 'singing', *pag-katok* 'knocking', *pag-bilang* 'counting'), unlike those in the nominalized clauses in sentence-final position, are not voice marked (and thus not oriented, or oriented towards the state of affairs and not to a participant involved in it). In the sentences in (65) each of the non-oriented clausal nominalizations in sentence-initial position functions as the VALUE expression and corresponds to the notional subject argument of the oriented clausal nominalization occurring sentence-finally (the VARIABLE expression).

It has been suggested in the literature that the sentence constitution pattern in Tagalog basically corresponds to that of nominal predication. The idea is that a basic clause consists of a subject term and a predicate, both of which are in principle nominal in nature. One telling feature of the sentence constitution pattern in the Tagalog pointing in this direction is the fact that non-subject direct arguments of voiced-marked verbs are coded in the same way as possessors in possessor-possessum relations between two nominal expressions (hence the label 'genitive' for non-subject arguments). As illustrated in Section 3.3.1 and 3.3.2, this happens in undergoer-oriented participial sentences in other languages (Turkish and Yaqui) as well, where the agent argument in an object-oriented participial clause takes the genitive case. In those two languages, however, the agent argument of an independent clause is coded as an (unmarked) nominative. That is, main clause morphosyntax in those languages, unlike in Tagalog, does not follow a "nominal" pattern. Another nominal-like feature of Tagalog sentence structure discussed in the literature concerns transitivity. Kaufman (2018; see also Kaufman 2009) argues that sentences in Tagalog, like nominal-predicate sentences, are essentially intransitive and involve a subject-predicate relation between a nominal term (introduced by *ang*) and an (essentially nominal) participial predicate (which, in equational sentences is simply marked as definite/specific by means of the nominative article *ang*). The idea is also articulated e.g. in Sasse (1991) who also describes the Philippine type sentence constitution pattern as basically that of nominal predication. Kaufman's (2018) argument for a basically nominal (and intransitive) structure of the Tagalog verbal sentence is based on the observation that in a (voice-marked) verbal sentence only one argument is required

(the subject). By "required", it is meant that only one argument must be available (though not necessarily overtly expressed, zero-anaphora is possible; more to this below). Consider the following examples from Kaufman (2018):

(66) Tagalog (Kaufman 2018: 211)

a. *K<in>áin* *ni* *Juan* *ang* *tokwa*.
<BEG>eat.PV GEN Juan NOM tofu
'Juan ate the tofu.'

b. *K<um>áin* *ng* *tokwa* *si* *Juan*.
<AV.BEG> GEN tofu NOM Juan
'Juan ate tofu.'

The predicate of the sentence in (66)(a) is in the patient voice and the verb selects the patient participant as the subject, which is introduced by the nominative article *ang* and typically occurs in sentence-final position. The agent argument, in this case a proper name (*ni Juan*), is marked by the genitive article (like the possessor in a possessum-possessor construction). The (undergoer) subject expression (*ang tokwa* 'the tofu') must involve a specific referent. Otherwise, another voice is selected. The sentence in (66)(b), where the predicate is in the actor voice, the subject is the agent argument (*si Juan*) marked by the nominative article and the undergoer (*ng tokwa*) is marked by the genitive article. Note that in this case, as the undergoer is not a specific entity, the use of the patient voice is ruled out. Kaufman (2018: 211) argues that the subject is the only necessary argument in these sentences. Only the subject is necessarily specified (either overtly expressed or recoverable from the context). Lacking a subject (whatever its role) a sentence is not complete. Consider the possible variants of the sentences in (66), shown in (67) and (68).

(67) Tagalog (Kaufman 2018: 211)

a. *K<in>áin* *ni* *Juan* *ang* *tokwa*.
<BEG>eat.PV GEN Juan NOM tofu
'Juan ate the tofu.'

b. *K<in>áin* *ang* *tokwa*.
<BEG>eat.PV NOM tofu
'The tofu was eaten.'

c. %*K<in>áin* *ni* *Juan*.
<BEG>eat.PV GEN Juan
'Juan ate (it).'

(68) Tagalog (Kaufman 2018: 211)

a. *K<um>áin* *ng* *tokwa* *si* *Juan*.
<AV.BEG> GEN tofu NOM Juan
'Juan ate tofu.'

b. %*K<um>áin* *ng* *tokwa*.
<AV.BEG> GEN tofu
'(Sh/e) ate tofu.

c. *K<um>áin* *si* *Juan*.
<AV.BEG> NOM Juan
'Juan ate.

The oddness of the patient voice sentence in (67)(c) and of the actor-voice sentence in (68)(b) is to be understood as applying to cases in which a subject argument (undergoer and actor, respectively) is not recoverable from the context and thus may be omitted.[51] This in contrast to the sentences in (67)(b) and (68)(c), where the absence of a specific referent corresponding to an actor and an undergoer respectively is, according to Kaufman, not in any way problematic. In these cases, the sentences would be, in some sense, notionally transitive (someone must have eaten the tofu in the state of affairs expressed in (67)(b) and John must have eaten something in (68)(c)), but not syntactically.

The idea that all (voiced-marked) sentences in Tagalog consist basically of a subject and a participial predicate plus the syntactic versatility of content-words in the language favors an analysis of all sentences being basically nominal (but not necessarily equational). Thus:

(69) Tagalog (Kaufman 2018: 213)

a. *K<um>a~káin ang laláki.*
<AV>IPFV~eat NOM man
'The man is eating.'

b. *Laláki ang k<um>a~káin.*
man NOM <AV>IPFV~eat
'The eating one is a man.'

Notice the way Kaufman translates the sentences in (69)(a) and (b): The translations are not 'the man is the one eating' or 'the man is the eating one' in (69)(a) and 'the eating one is the man' in (b). To obtain equational interpretations the expressions in sentence-initial position would need to be marked as referential expressions (with the nominative article *ang*). At this point, one turns back to the problem discussed in 2.3.2 in Chapter 2, regarding the question of whether a sentence such as (69)(b) is necessarily predicational (i.e. 'the one who is eating is male') or can be considered a cleft with a non-definite (i.e. not unique or familiar) or non-specific (but referential) expression as the clefted constituent.

3.4 A note on the noun anchoring/support strategy

It was mentioned in Chapter 1 that accounts of clefting in the literature may differ with respect to the status assigned to a sentence such as the following, cited in Collins (1991):

(70) English (Collins 1991: 31)
[*The artist he resembles most closely in spirit] is, I think*, <u>*Watteau*</u>.

It is perhaps clear that this sentence is a specificational sentence (in the sense explained in Chapter 2). In many accounts on clefting, however, sentences such as the one in (70), if discussed at all, are not described

[51] Himmelmann (1999) points out that

> [i]t has been widely assumed that Tagalog allows zero anaphora freely for both actors and undergoers in semantically transitive constructions. [Evidence] strongly suggest that this assumption is wrong for actors of one of the two basic transitive construction types: undergoer-oriented constructions. In these constructions, the actor argument does not appear to be omissible in contexts in which zero anaphora would be pragmatically warranted. (Himmelmann 1999: 231)

The argument may be taken to support Kaufman's (2018) view: If a non-subject expression is omitted, there is simply an intransitive sentence with no oblique or adjunct expressions. Only subjects (whatever their role/function) allow zero anaphora. An undergoer-voice sentence lacking the overtly expressed actor receives an (intransitive) agent-less passive interpretation.

as clefts. Some authors would prefer to describe the sentence in (70) as a simple specificational sentence involving a VARIABLE expression, which, as it were, happens to involve a relative clause as a modifier. As I (hope to) have made it clear, the position adopted in the present investigation is that the sentence is a cleft.

As was pointed out earlier (3.2.2), in some languages the presence of a noun is required in an oriented nominalization. Dryer (2004) notes that in some languages, nominal expressions ("noun phrases" in his terms) may require nouns but in others a noun providing sortal information about the referent may be omitted if this information is recoverable. This is for instance the case in Spanish. Consider the examples in (71), contrasting two nominal expressions, once with a noun and an adjective following a definite article (a), and once without the noun (b):

(71) Spanish (constructed)

a. [*El* *perro* *pardo*] *me* *mordió*.
ART.SG.M dog brownish.SG.M 1SG.ACC bite.3SG.PST
'The brownish dog bit me.'

b. [*El* *pardo*] *me* *mordió*.
ART.SG.M brownish.SG.M 1SG.ACC bite.3SG.PST
'The brownish one bit me.'

If it is sufficiently clear from the context that the referent of the nominal expression in (71)(b) is a dog, the noun *perro* 'dog' may be freely omitted. Not so in English, however. The noun *dog* may be omitted but the pro-noun *one* (also used as numeral and as an impersonal pronoun) must be used. The case is similar with nominal expressions involving relative clauses. Consider the examples in (72).

(72) Spanish (constructed)

a. [*El* *perro* *que* *me* *mordió*] *era* *pardo*.
ART.SG.M dog REL 1SG.ACC bite.3SG.PST be3SG.IPFV brownish.SG.M
'The dog that bit me was brownish.'

b. [*El* *que* *me* *mordió*] *era* *pardo*.
ART.SG.M REL 1SG.ACC bite.3SG.PST be.3SG.IPFV brownish.SG.M
'The one that bit me was brownish.'

As in the case illustrated (71)(b), the nominal expression in (72)(b) is perfectly acceptable in Spanish without a noun providing the sortal information about the referent. Again, English requires a pro-noun in this case. In some languages, the equivalent to the sentences in (71)(b) and (72)(b) may require the use of a common noun. This may be the equivalent of *dog* (if that's what the referent is) or some noun with a more general denotation (perhaps with the meaning 'thing'). Substitute the verbal predicate in (either of) the sentences in (72) for a referential expression such that the relationship between the nominal expression in subject position and the post copular expression is a specificational one and you get a cleft:

(73) Spanish (constructed)

a. [*El* *perro* *que* *me* *mordió*] *fue* *el* *pardo*.
ART.SG.M dog REL 1SG.ACC bite.3SG.PST be.3SG.PST ART.SG.M brownish.SG.M.
'The dog that bit me was the brownish one.'

b. [*El* *que* *me* *mordió*] *fue* *el* *pardo*.
ART.SG.M REL 1SG.ACC bite.3SG.PST be.3SG.PST ART.SG.M brownish.SG.M
'The one that bit me was the brownish one.'

A noun-anchored nominalization may be used, as in (73)(a), if one needs (for some reason) to specify that one is talking about dogs. One may use a nominalization without noun support (73)(b) if it is obvious what kind of entity is involved (or if for some reason one does not wish to constrain the sort of possible referents).

The conditions under which a common noun may be required in in relative clause constructions may vary from language to language. In Hausa, for example, relative pronouns may only be used if the referent is human (its role/function is not immediately relevant). Otherwise, a common noun is required:

(74) Hausa (Green 2007: 69, 124)

a. [*waddà* *nakè* *sô*] *Kánde* *cè*
REL.PRO.F.SG 1SG.FOC.IPFV love.VN Kande FOC.F
'The one I love is Kande.'

b. [*àbîn* *dà* *takè* *bùkātà*] *shìnkāfā* *cè*
thing.DEF.M REL 3SG.F.FOC.IPFV need.VN rice.F FOC.F
'What she needs is rice.'

In principle, the construction in (74)(a) is fully parallel to that in (74)(b). The relative pronoun is a contraction of an interrogative pronoun and the particle *dá*[52]. Note that in English, as suggested by the respective translations of the sentences in (74), the situation is in a certain sense the inverse: Only nominalizations oriented towards non-human referents are freely allowed with a relative pronoun. A sentence such as *Whom I love is Kande* is reported not to be readily acceptable to native speakers of English (see e.g. Collins 1991).

The requirement of common nouns in nominalizations may vary also between (more or less) closely related languages. As was mentioned in the previous section, in Tahitian a nominalization may be formed by introducing a clause with an article, like any other nominal expression in the language. The pattern was illustrated in example (16) in Section 3.2.5, repeated here as (75).

(75) Tahitian (Lazard & Peltzer 2000: 85)
'o *te* *taote* [*te-i* *haere* *mai*]
ID ART doctor ART-ASP go DIR
'Celui qui est venu est le docteur.' / C'est le docteur qui est venu.'
'The one who came is the doctor' / 'It's the doctor who came.'

Note that a clause introduced by the article is a full-fledged nominal expression in the language. Now, a nominal expression introduced by an article and consisting of a noun (and, possibly, other modifiers) and a clause following it is also possible in the language as illustrated in example (76), where the noun *fare* 'house' is directly followed by a subordinate clause introduced by the aspect marker *i*.

[52] Note, incidentally, that the relative clause differs from a main clause to some extent. In the perfective and imperfective aspect, particular forms of the auxiliary-like element marking subject person and aspect are used. These forms, however, do not occur exclusively in relative clauses but also in "out-of-focus" clauses in focus constructions, frame-setting adverbial clauses (also introduced by the particle *dà*), as well as in some contexts in narrative sequences (see Abdoulaye 2007; Abdoulaye et al. 2020).

(76) Tahitian (Lazard & Peltzer 2000: 78)

nō	*Teri'i*	[*te*	*fare*	*i*	*pa'apa'a*]
PREP	Terii	ART	house	ASP	burn

'La maison qui a brûlé est à Terii.'
'The house that burnt down is Terii's.'

In fact, Lazard & Peltzer (2000: 77) note that restrictive relative clause constructions in the language are formed following the pattern in (76). Non-restrictive relative clause constructions are formed by nominal expressions in apposition to relative clauses introduced by the article:

(77) Tahitian (Lazard & Peltzer 2000: 83)

teie	*te*	*hōho'a*	*o*	*Teri'i*	[(*'o*)	*tē*	*haere*	*mai*	*'araua'e*]
DEM	ART	image	PREP	Terii	ID	ART.ASP	go	DIR	soon

'Voici la photo de Terii, qui viendra tout à l'heure.'
'Here's the picture of Terii, who will come soon.'

Note, interestingly, that the second term in apposition is (optionally) introduced by the morpheme *'o* used to mark (referential) nominal expressions in predicate-like function (cf. (75) above). The pattern involving the use of a cognate of Tahitian *'o* in such appositive constructions seems to be widespread in Polynesian (see e.g. Custis (2004) on Tongan). A sentence comparable to the one in the Tahitian example in (77) is shown in the Rapa Nui example in (78), where the second of two nominal expressions in apposition is introduced by the morpheme *ko.*

(78) Rapa Nui (Kieviet 2017: 263)

He	*tu'u*	*mai*	*te*	*'avione*	*ra'e*	[*ko*	*te*	*'avione*	*ena*
NTR	arrive	DIR	ART	airplane	first	PROM	ART	airplane	MED
e	*kī*	*ena*]	*he*	*DC10*].					
IPFV	say	MED	PRED	DC10					

'The first airplane, the airplane called DC10, arrived.'

However, in Rapa Nui (unlike in Tahitian), the article plus noun anchoring pattern is the only pattern available for relative clause constructions. Consider the following cleft examples:

(79) Rapa Nui (Kieviet 2017: 461)

Ko	*te*	*nūna'a*	*era*	*'a*	*'Ōrare*	[*te*	*nūna'a*	*i*	*rē*].
PROM	ART	group	DIST	of.AL	Orare	ART	group	PFV	win

'(in a report about a music contest:) Orare's group was the group that won.'

(80) Rapa Nui (Kieviet 2017: 461)

Ko	*Timo*	[*te*	*me're*	*'ori*	*tako'a*	*roro*	*nei*].
PROM	Timo	ART	thing	dance	also	inside	PROX

'Timo is the one who is dancing inside (= in this picture).'

Kieviet (2017:14) explicitly notes that relative clause constructions without a noun are not possible in the language. As illustrated in example (80), a noun with a very general denotation may be used. It seems that the common noun *me're* 'thing' may be used as a pro-noun in a similar way as *one* in English. This strategy is apparently not used in Tahitian, where a clause may be treated in principle in the same way as any nominal

expression, introducing it with an article (the use which is virtually obligatory for nominal expressions in the language).[53]

The use of a noun in a relative clause construction may be motivated by the need to clarify the reference (i.e. orientation) of a nominalized clause. In Mandarin Chinese an oriented nominalization is marked by means of the particle *de* placed on the right edge of the subordinate clause. The default orientation, according to Li & Thompson (1989[1981]: 577) is towards an object (undergoer) argument of the clause.[54] Orientation towards an agent argument, however, is also readily available. As a general rule, the argument towards which the nominalized clause is oriented may not be expressed within the nominalized clause (i.e. it is "gapped"). If it is not a direct argument (i.e. a subject or direct object), however, a pronoun may be used (Li & Thompson 1989[1981]: 585). Note that in Chinese, all argument expressions (if recoverable or irrelevant) may be omitted. In the absence of any argument expression (and, in appropriate context), the nominalized clause may be interpreted as oriented towards any of the direct arguments:

(81) Mandarin (Li & Thompson 1989[1981]: 581)

[*zuótian*	*pīping*	*de*	*rén*]	*dōu*	*bu*	*zài*	*zhèlī*
yesterday	criticize	NMLZ	person	all	not	at	here

'The people who criticized (others) yesterday are not all here.'
'The people whom (others) criticized yesterday are not all here.'

In principle, in a sentence such as that in (81), the noun following the nominalized clause (*rén* 'person' in the case at hand) may be omitted. It is reported, however, that orientation towards some role/functions such as instrument, or local and temporal circumstances is only possible if a noun with the corresponding sortal denotation follows the clause marked with the particle *de*. Consider the following examples from Tsai (2008).

(82) Mandarin (Tsai 2008: 969)

a.

[*zoutian*	*zai*	*jiaoshi*	*yong*	*saoba*	*henhendi*	*da*	*Xiaodi*	*de*	(*ren*)]
yesterday	at	classroom	with	broom	fiercely	beat	Xiaodi	*DE*	person

shi *Akiu*
be Akiu
'The person/one who beat Xiaodi fiercely in the classroom with a broom yesterday was Akiu.'

b.

[*zoutian*	*Akiu*	*zai*	*jiaoshi*	*yong*	*saoba*	*henhendi*	*da*	*de*	(*ren*)]
yesterday	Akiu	at	classroom	with	broom	fiercely	beat	*DE*	person

shi *Xiaodi*
be Xiaodi
'The person/one whom Akiu beat fiercely in the classroom with a broom yesterday was Xiaodi.'

[53] Apparently, in Rapa Nui, a clause directly introduced by an article obtains a non-oriented interpretation. But the verb usually carries a nominalizing suffix and aspect markers are absent. On non-oriented nominalization patterns in the language, see Kieviet (2017: 88ff).

[54] "If both the subject and direct object participants are unspecified in a nominalization [i.e. not overtly realized within the nominalized clause], then that nominalization will generally be understood to have the same referent as the unspecified direct object participant of that verb" (Li & Thompson 1989[1981]: 577).

c. [*zoutian* *Akiu* *zai* *jiaoshi* *henhendi* *da* *Xiaodi* *de* *(*fangshi*)]
yesterday Akiu at classroom fiercely beat Xiaodi *DE* way
shi (*yong*) *saoba*
be with broom
'The way how Akiu beat Xiaodi fiercely in the classroom with a broom yesterday was with a broom.'

d. [*zuotian* *Akiu* *yong* *saoba* *henhendi* *da* *Xiaodi* *de* *(*didian*)]
yesterday Akiu with broom fiercely beat Xiaodi *DE* location
shi (*zai*) *jiaoshi*
be at classroom
'The location where Akiu beat Xiaodi fiercely with a broom yesterday was in the classroom.'

e. [*Akiu* *zai* *jiaoshi* *yong* *saoba* *henhendi* *da* *Xiaodi* *de* *(*shijian*)]
Akiu at classroom with broom fiercely beat Xiaodi *DE* time
shi *zuotian*
be yesterday
'The time when Akiu beat Xiaodi fiercely in the classroom with a broom was yesterday.'

The examples above would show, then, not that Mandarin Chinese does not permit the clefting of instrument, location, or time expressions but simply that, in this language, the noun anchoring/support strategy is required in these cases.

3.5 Summary

The main purpose of this chapter was to clarify the understanding of the notion of oriented clausal nominalization. In some languages, the internal structure of a nominalized clause may not differ considerably from that of an independent one apart from the (often) obligatory omission (often described as "gapping") or obligatory pronominalization (in some cases involving the use of specialized forms, i.e., relative pronouns) of an expression corresponding to the entity denoted by the nominalized clause as a whole. Minor differences such as (more) rigid word order patterns or restricted TAM marking may also distinguish nominalized clauses from independent ones. In some languages, clausal nominalization involves the use of specialized forms of the predicate within the subordinate clause which may themselves be described as involving word-level nominalization. In some cases, this involves a distinctive pattern in the coding of (some) arguments in the subordinate clause (e.g. direct arguments may be marked as possessors). Frequently, clausal nominalization involves the use of specialized markers signaling the nominal (or subordinate) status of the clause. I identified eight distinct strategies involved in oriented clausal nominalization:

i. Zero/minimal marking
ii. Noun anchoring/support
iii. Relative pronoun
iv. Nominal classifier
v. Article
vi. Article/ general subordinator
vii. Relative marker
viii. Nominal verbal form

These strategies are not (necessarily) mutually exclusive and the distinction between one strategy and another may be tenuous. Strategies i through vii were respectively illustrated in sub-sections 3.2.1 to 3.2.7 with data from diverse languages where each of the nominalization strategies may be observed in a pure form.

The use of nominal verbal forms (Strategy viii) was discussed in more detail in Section 3.3. Using data from four different languages (Turkish, Yaqui, Ingush, and Tagalog), I discussed cross-linguistic variation concerning inherent vs. contextual orientation patterns (in the sense of Shagal 2019) as well as the formal overlap (and differences) between oriented and non-oriented clausal nominalizations involving specialized (nominal) verb forms. In Section 3.3.4, I discussed clausal nominalization and clefting in Tagalog. The patterns observed in Tagalog are especially interesting in the context of the present discussion. Clausal nominalization in this language may be described as involving an article-marking strategy. On the other hand, the general sentence constitution pattern in the language has been argued to correspond quite closely to a nominal one: The idea is that verbs in Tagalog (and in Philippine-type languages generally) show a strong affinity to participial forms in other languages. Under this view, the rich voice system in Tagalog may be regarded as an elaborate paradigm of inherently-oriented participles.

Among the nominalization strategies identified, the noun anchoring/support strategy (presented in 3.2.2) is of special importance in the context of a discussion of clefting patterns and was further addressed in section 3.4. Continuing the discussion begun in Section 1.1.3 (Chapter 1), I argued for an approach which views sentences featuring VARIABLE expressions involving what I describe as noun anchored/supported clausal nominalizations (i.e.“headed” relative clause constructions) as clefts.

4 Word order

This chapter is dedicated to the discussion of word order in cleft sentences. The chapter is divided into two main sections. In 4.1 I will present five case studies illustrating the alternation between VARIABLE-VALUE and VALUE-VARIABLE configurations in clefts (and specificational sentences in general) in typologically diverse languages. In 4.2 I will focus on data from languages in which cleft sentences involve the dislocation of the VARIABLE expression and its resumption by means of a pronominal form. The discussion in the latter section is of direct relevance to the problems to be examined in Chapter 5.

4.1 VARIABLE-VALUE / VALUE-VARIABLE configurations

Specificational sentences are defined by a relationship between two nominal expressions. The reference of one (the VARIABLE) is specified by means of another (the VALUE). I have described the relation between them as subject-predicate-like, and mentioned that in many languages nominal-predicate sentences are formally identical to specificational sentences (see discussion in 1.2.2 in Chapter 1, and 2.1.3 and 2.2.1 and in Chapter 2). Though, as argued earlier, it is best to keep the notions of subject and predicate on the one hand, and VARIABLE and VALUE on the other, there is an affinity between the VARIABLE expression and the subject, and the VALUE and the predicate. In English, for instance, both specificational and nominal-predicate sentences exhibit a basic word order structure such that two expressions occur flanking the copular verb *be* (in declarative sentences). The pre-copular position is the canonical subject position and the post-copular position is generally occupied by the nominal predicate. The basic pattern in specificational sentences is, arguably, VARIABLE-copula-VALUE. The order of the terms involved in specificational sentences is reversible, but a VALUE-copula-VARIABLE is generally assumed to be a pragmatically marked variant, signaling contrast or emphasis.[55] I will not systematically address the question concerning the discourse contexts favoring VARIABLE-VALUE or VALUE-VARIABLE orders in clefts (and specificational sentences in general). Note, however, that in most of the accounts examined in which the issue is discussed, VALUE-VARIABLE order is associated with pragmatic effects variously described as involving emphasis, contrast, or counter-expectation, independently of the means by which the alternations in the relative position of the VARIABLE and VALUE expressions is achieved.[56] As the case studies included in the present section illustrate, the patterns involved in word order alternations in clefts (and specificational sentences generally) vary considerably cross-linguistically.

4.1.1 Mandarin

A cleft sentence in Mandarin consists of a participant-oriented nominalization in sentence initial position and a nominal expression following the copula *shì*. The sentence in (1)(a) illustrates a canonical (SVO) sentence in the language and the sentence in (1)(b) a (truth-conditionally equivalent) cleft sentence with a

[55] This does not apply however, when the VALUE expression is a pronominal form. In this case, the unmarked order seems to be VALUE-VARIABLE (see e.g. Van Praet 2019; Van Praet & O'Grady 2018). In some languages, there seems to be a hard constraint such that, whatever the relationship between two (referential) nominal expressions, pronominal forms (and proper names) necessarily precede other nominal expressions (see e.g. Kaufman 2018 for an account on the phenomenon in Tagalog, and Lyon 2014 for similar restrictions in Upper Nicola Okanagan Salish).

[56] For an account on the pragmatic contrast between VALUE-VARIABLE and VARIABLE-VALUE orders in a language with canonical predicate-initial word order see Nuhn (2019) on Tagalog. In this language, VARIABLE-VALUE order obtains by means of the topicalization of the VARIABLE expression.

proper noun (*Wangwu*) in post-copular position corresponding to the (obligatorily) "gapped" subject in the cleft clause (*jian guo Lisi de*).

(1) Mandarin (Wu 2017:159)

a. *Wangwu jian guo Lisi.*
Wangwu see EXP Lisi
'Wangwu has met Lisi.'

b. [*jian guo Lisi de*] *shi* *Wangwu.*
see EXP Lisi DE COP Wangwu
'(The one) who has met Lisi is Wangwu.'

Constituent order does not seem to be reversible. Sentences involving a participant-oriented nominalization in post-copular position are possible. It is not clear, however, to what extent such constructions allow specificational readings.

Li & Thompson (1989[1981]) describe the use of constructions comparable to the one in (1)(b) with the order of constituents reversed as one of the strategies available in the language to promote an object to subject position, yielding the equivalent of a passive sentence. Note the English translation of the Mandarin sentence in example (2).

(2) Mandarin (Li & Thompson 1989[1981]: 500)
zhèi ge fángzi shì [*Zhāngsān shèjì de*]
this CLF house COP Zhangsan design NMLZ
'This house was designed by Zhangsan.'

It must be pointed out that a morpheme with the shape *de* is found in many different constructions and its status as a nominalizer (as opposed to e.g. a particle marking modality, stance, or evidentiality) in many of these is controversial.[57]

It seems that clefts in Mandarin follow quite closely the word-order patterns observed in predicational copular sentence in the language in the sense that a (true) nominal predicate (in the sense discussed in 2.2.1 in Chapter 2) and the VALUE expression in a specificational sentence must occur in post-copular position. Consider the sentence in (3), which, according to the account in Tham (2008), allows a specificational but does not allow a predicational reading. That is, a reading under which the pre-copular expression is interpreted as a property predicated of the post-copular one.

(3) Mandarin (Tham 2008: 73)
laoshi shi Sanmao.
teacher COP Sanmao
'The/ *A teacher is Sanmao.'

[57] It not clear whether a common noun (perhaps an hyperonym of 'house' or a noun with a general meaning such as 'thing' could occur after de in the sentence in (2), forming a noun-supported oriented nominalization (in the sense discussed in Section 3.2.2 in Chapter 2). What I have in mind is something like *zhei ge fangzi shi Zhangsan sheji de fangzi* 'This house is the/a house Zhangsan designed'.

To clarify: Tham (2008) argues that in the sentence in (3), the pre-copular nominal expression *laoshi* 'teacher' is necessarily interpreted as referential.[58] The expression *laoshi* in (3) cannot function as a nominal predicate. That is, it cannot be used to ascribe a property to the referent of the post-copular expression. To obtain a predicational reading the order SUBJECT + COPULA + NOMINAL PREDICATE is required, as in (4).[59]

(4) Mandarin (Tham 2008: 73)
Sanmao shi laoshi.
Sanmao COP teacher
'Sanmao is a/the teacher.'

As suggested in the English translation, the nominal expression *laoshi* 'teacher' can be interpreted as a specific-definite nominal expression (i.e. denoting a unique individual) or as a (non-referential) nominal predicate. In the latter case the sentence simply attributes the property of being a teacher (and the copula *shì* may be omitted). It is possible that the sentence in (4) may allow a specificational reading in a context where an individual's name is known but not its identity and the individual in question can be identified using the expression *laoshi* 'the teacher'. That is, a reading where the proper name is the VARIABLE expression and the postcopular nominal expression the VALUE. It seems that a specificational reading of (4) in which *laoshi* 'a/the teacher' is the VARIABLE and *Sanmao* is the VALUE is not possible (but this would require clarification).

In sum, word order in Mandarin seems to be rigid both for predicational and specificational sentences. The nominal predicate (unless left-dislocated) must occur in post-copular position and this is also (apparently) necessarily the case for the VALUE expression in a specificational sentence.

4.1.2 Akan/Twi

In Chapter 2 (Section 2.2.2.2), I discussed the formal distinction in Akan/Twi (Kwa, Atlantic-Congo) between equational sentences (involving the use of the copula *ne*, which in some accounts is described as a "focus marker") and sentences involving nominal predicates (involving the use of the copula *yɛ*). In predicational (*yɛ*) copular sentences word order is fixed and a referential nominal expression must precede a nominal or adjectival predicate. Word order in equational/specificational (*ne*) copular sentences is flexible and only nominal expressions may be involved.

Clefts in Akan are formed with a participant-oriented nominalization, which can be a relative clause construction headed by a common noun, or a headless relative introduced by the elements *deɛ* or *nea*. The

[58] Though not necessarily definite. Consider the following cleft sentence from Tham (2008):

(i) Mandarin (Tham 2008: 74)

[*Yi*	*ge*	*houlai*	*bijiao*	*xingfu*	*de*	*gu_er*]	*shi*	*Sanmao*.
one	CLF	later	relatively	happy	ASSOC	orphan	be	Sanmao

'An orphan that was happier later on is Sanmao.'

[59] Though left dislocation of a nominal predicate is apparently possible:

(ii) Mandarin (Tham 2008: 74)
Laoshi, Sanmao shi.
teacher Sanmao be
'A teacher, Sanmao is.'

If I understand Tham's (2008) argumentation, the left-dslocated expression in the sentence above is non-referential. (The point the author is making by showing the sentence is to contrast it with necessarily referential pre-copular expressions in specificational sentences.)

former is mainly used in the Asante and the latter in the Akuapem dialect, according to Campbell (2020: 18).[60] As in simple equational/specificational sentences, in Akan clefts, the equational copula *ne* always occurs between the two nominal expressions involved. The order in which these expressions occur is flexible. Consider the sentences in (5):

(5) Akan (Ofori 2011: 249, 243)

a. [*Onipa* *a* *ɔ-ba-a* *ha*] *ne* *Kofi*
person REL 3SG-come-PST here COP Kofi
'The person/one who came here is Kofi.'

b. *Kofi* *ne* [*onipa* *a* *ɔ-ba-a* *ha*]
Kofi COP person REL 3SG-come-PST here
'Kofi is the person/one who came here.'

As the English translations in (5) suggest, both sentences convey a specificational relation where a proper name is used to identify the referent denoted by the participant-oriented nominalization (in the case at hand, a relative clause construction headed by the common noun *onipa* 'person').

Note that it is not always necessarily the case that the participant-oriented nominalization in a *ne* copular sentence is the VARIABLE expression. Consider, for instance, example (6) below, discussed in Ofori (2011). According to the author, in this sentence the relative clause construction is used to provide a VALUE, clarifying the reference of the proper name (Note that the relative clause construction is marked as definite by morpheme *no*.). The proper name is the VARIABLE expression.

(6) Akan (Ofori 2011: 258)
Kofi *ne* [*abfora* *a* *ɔ-ba-a* *ha* *no*]
Kofi COP child REL 3SG-come-PST here DEF
'Kofi is the child who came here.'

Ofori (2011) suggests that the sentence does not clarify who 'the child who came' was, but rather who the name Kofi refers to. In Ofori's terms:

> the focused NP (Kofi) is a specific name and gives the impression that the speaker and addressee together have encountered the name (not the person) before. The focus sentence therefore is about linking (or equating) the name Kofi […] to 'the person (i.e. child) who came here' […]. (Ofori 2011: 258)

[60] Consider the following sentences, in which the participant-oriented nominalizations introduced by *deɛ*/*nea* occur as arguments in verbal sentences (a-b), and as a term in an equational-specificational (*ne*) copular sentence (c).

(i) Akan (Campbell 2020: 40)

a. [deɛ/nea wó-dé má-a mé nó] *sua*
NMLZ 2SG-take give-PST 1SG DEF be.small
'What you gave me is small.'

b. *Mè-m̀-pé* [*dèè/nèà* *wó-ré-yɛ́* *nó*]
1SG-NEG-like NMLZ 2SG-PROG-do DEF
'I don't like what you are doing.'

c. [*Dèè/nèà* *mè-pè* *n'ă!sém̀* *pǎǎ́*] *né* *Kòjó*
NMLZ 1SG-like 3SG.matter very.much be Kojo
'The one I really like is Kojo.'

Following Ofori's (2001) account, then, the sentence in (6) is specificational and the reference of an individual named Kofi is specified. Note that such specificational relations are coded in the language in a different way from sentences associating the referent of a nominal expression with a proper name in what could be described as a "naming" relation (e.g. *The child who came here is called Kofi*). Sentences in which a fully referential expression is associated with a name involve the use of the copula *de* 'be called, be known as' (also reported in Ellis & Boadi (1969: 30) to be used as "a strong form of *yɛ*" with the meaning 'to be undeniably, to have the undeniable characteristics of.'). The "name-assigning" use of *de* is illustrated in (7).

(7) Akan (Campbell 2020: 40)

[*Dèɛ̀/Nèà*	*Kòfí*	*bísá-à*	*nò*	*sìká*	*nó*]	*dè*	*Kòjó*
NMLZ	Kofi	ask-PST	him	money	DEF	be.called	Kojo

'The one Kofi asked for money is called Kojo.'

The sentence in (7) is not a specificational sentence. The participant-oriented nominalization (in square brackets) is, as in (6) fully referential but it is not in an equational relation with the proper name. In this kind of constructions (as in predicational (*yɛ*) copular clauses) word order is reported not to be reversible (Ellis & Boadi 1969: 30).

Note that the marking of the participant-oriented nominalization as definite (as in (6) above) does not necessarily make it a fully referential expression. That is, a configuration in which a definite-marked oriented-nominalization is the VARIABLE is not precluded (cf. e.g. examples in (10) below). The presence of a definite marker in the participant-oriented nominalization presumably only signals that the referent in question is (explicitly) marked as being unique or familiar (for an account of definiteness-marking in Akan, see Bombi 2018, Bombi et al. 2019). In other words, it seems that the VALUE-VARIABLE relation holding between a proper name and definite-marked participant-oriented nominalization (as in (6)) is a possible interpretation (though not necessarily the only one). It is possible that word order may play a role, but the accounts consulted do not allow an assessment of the conditions favoring one interpretation or another.

As mentioned above, word-order reversibility reported for clefts in the language is also reported for specificational sentences in general. The following minimal pair illustrates two possible word order configurations in a simple equational/specificational sentences:

(8) Akan (Boadi 1974: 8; cited in Ofori 2011: 259)

a. *Me* *ne* *ɔsɔfoɔ*.
1SG COP priest
'I am the one who is a pastor/priest.'

b. *Ɔsɔfoɔ* *ne* *me.*
priest COP 1SG
'The one who is a pastor/priest is me.'

Note that neither of the sentences in (8) are clefts. The English translations only suggest the specificational relation holding between two nominal expressions.[61] A predicational relation between *me* 'first person singular personal pronoun' and *ɔsɔfoɔ* 'priest' would be coded as *me yɛ ɔsɔfoɔ* 'I am a priest' using the

[61] The translations suggested respectively for the sentence in (a) are "I am a pastor/priest" (Ofori 2011: 259) and ''I am priest. (It is I who am a priest, I am the one who is a priest)" (Boadi 1974: 8). Boadi's translation is retained in example (9)(a) below. The translation for sentence in (b) is taken from Ofori (2011: 259). The problem with the English translation "I am a pastor/priest" for (a)is that it very strongly suggests a predicational interpretation.

copula *yɛ* instead of the copula *ne*. Recall that in this case word order would not be reversible (and the personal pronoun would necessarily occur preceding the copula *yɛ*).

It is difficult with the available information to assess what discourse conditions would allow (or favor) the use of a VARIABLE-VALUE vs. a VALUE-VARIABLE word order in specificational sentences in Akan. As mentioned in Section 2.2.2.2 in Chapter 2, Ellis & Boadi (1969: 59) explicitly note that "emphasis" can be shifted between the expressions in pre- and post-copular positions. It may be the case that pre-copular VALUE order is associated with the expression of insistence, contrast, or counter-expectation, to judge from the reported pragmatic equivalence between specificational sentences exhibiting this word order and "focus" constructions in the language. I will now briefly discuss this kind of constructions.

Clefts constructions co-exist with so-called "focus" constructions in the language. The latter are sentences in which a (not necessarily nominal) expression "in focus" (necessarily) precedes an "out-of-focus" clause. In these constructions, the morpheme *na* occurs between the expression "in focus" and the "out-of-focus" clause. This element is also used as a conjunction in the language (see Fiedler & Schwarz 2007, Schwarz & Fiedler 2007, van Dommelen & Beerman 2019). The morpheme *na* is often described as a "focus marker" but its use in "focus" constructions may possibly be better described as a "background" marker (in the sense of Güldemann 2016) introducing a backgrounded clause rather than marking as "focal" the expression preceding it. (As a conjunction, *na* always precedes a clause, and it is not necessary for anything to precede it.) Grubic et al. (2019) compare "focus" constructions and "in situ focus sentences" (i.e. canonical sentences) in Akan and argue that the use of "focus" constructions seems to signal contrastiveness and exhaustivity (which is not necessarily the case in "in situ focus" sentences. Note, however, that the authors do not discuss (or even mention) the use of clefts in the language.[62] Boadi (1974) suggests that clefts with the specifying expression in sentence-initial position and "focus" constructions are "synonymous for all practical purposes; they share identical presuppositions and answer the same questions" (p. 14). This seems to be the case when the VALUE expression precedes the VARIABLE expression in specificational sentences in general (i.e. whether the VARIABLE expression is a relative clause construction or a simple nominal expression). Consider the examples in (9). (Note that Boadi 1974 glosses *ne* as 'focus marker', not as 'copula'.)

(9) Akan (Boadi 1974: 8)

a. *Me* *ne* *ɔsɔfoɔ.*
   1SG FOC priest
   'It is I who am a priest.' / 'I am the one who is a priest.'

b. *Me* *na* *me* *yɛ* *ɔsɔfoɔ.*
   1SG EXCL.FOC 1SG be priest
   (same)

Boadi (1974) notes that "the sentence [(9)(a)] is synonymous with the sentence [(9)(b)]"and adds that "as far as [the author] is aware, there is no context in which they are not mutually substitutable" (p. 8). Note, however, that the expression following the morpheme *na* in (9)(b) is not a nominal expression but a

[62] The authors do not discuss the use of clefts in the sense I am using the term. That is, they do not mention the use of *ne* copular sentences with a participant-oriented nominalization as a VARIABLE term. With respect to *na* "focus" constructions, however, Grubic et al. (2019) remark that "[s]ince exhaustivity and existence inferences are typical for clefts, cross-linguistically, and these particles [Akan *na*] induce a structural bi-partition into a focused constituent to their left and a backgrounded part to their right, [Grubic et al.] propose to analyze the marked focus constructions in Akan […] as clefts" (p. 152). Clearly, this understanding of clefts is different from the one I adopt in this dissertation.

backgrounded clause. The same pragmatic equivalence argued for sentences such as (9)(a) and (b) is reported to hold between (VALUE-initial) clefts and focus constructions. Boadi (1974) presents a number of cleft sentences involving two possible orders (10)(a-b) and a corresponding "focus" construction (10)(c):

(10) Akan (Boadi 1974: 14)

a. [*bere a me-ba-a ha no*] *ne* *<u>Ɔpɛpɔn.</u>*
time REL 1SG-come-PST here DEF FOC January
'The time I came here was JANUARY.'

b. *<u>Ɔpɛpɔn</u>* *ne* [*bere a me-ba-a ha no*].
January FOC time REL 1SG-come-PST here DEF
'JANUARY was the time I came here.'

c. *Ɔpɛpɔn na me-ba-ae.*
January EXCL.FOC 1SG-come-PST
'JANUARY was when I came.'

Unfortunately, it is not quite clear from Boadi's (1974) account what the difference between the conditions favoring a word-order pattern such as that found in (10)(a) and the one in (b) would be.

To summarize, word order in Akan specificational sentences (including clefts) is reported in the literature to be fully flexible (VARIABLE *ne* VALUE or VALUE *ne* VARIABLE). The word order flexibility in this kind of sentences contrasts with the rigid word order reported in the literature for predicational (*yɛ*) sentences in which a referential expression necessarily precedes a nominal (or adjectival) predicate. Specificational sentences in which the VALUE expression occurs in pre-copular position are reported to be pragmatically equivalent to "focus" constructions. In the latter, an expression (not necessarily a nominal one) occurs in sentence initial position and is followed by an "out-of-focus" clause (not a participant-oriented nominalization) introduced by the morpheme *na*. "Focus" constructions are reported, among other things, to mark contrast and trigger exhaustivity implicatures. The conditions favoring the use of one of the two possible word order configurations in clefts remains unclear. Intriguingly, some accounts on information structure in Akan do not mention clefts (as defined for the purposes of this dissertation) at all.

4.1.3 Hausa

In Hausa, copular clauses in assertive contexts (i.e. in declarative, non-subordinate clauses; see Abdoulaye et al. 2020) consist of two nominal expressions, one of which is followed by a morpheme variously described in the literature as "stabilizer", "focus marker", "copula", or "predicator". The element in question distinguishes gender and number. The relevant forms are *nē* (masculine and plural) and *cē* (feminine singular) in eastern varieties, and *na* and *ta* in western varieties of the language. These markers are argued to be of demonstrative origin (the onset consonants may be associated with gender/number distinction, the vowels with a distal vs. proximal opposition; see Schuh n.d.). Note that the forms *cē* and *nē* are not used as demonstratives in the language.[63] The markers in question typically (but not always) agree in gender and number with the subject or non-predicative expression (i.e. the expression *not* immediately followed by the morpheme). The marker occurs optionally in (non-cleft) "focus" constructions, following the focused element, which necessarily precedes an "out of focus" clause. It is important to point out that the markers

[63] Assuming the marker *nē* and *cē* are of deictic origin (see discussion in Schuh (n.d.) for a detailed account), they may be perhaps described as having developed from "demonstrative identifiers" in the sense of Diessel (1999) (see also Himmelmann 1997: 126 on the notion of 'predicative deictics' (*prädikative Deiktika*).

nē and *cē* always follow the "predicative" or "focused" expression. That is, the markers do not (necessarily) occur *between* two expressions when used in a copula-like way. Furthermore, *nē* and *cē* may occur following a nominal expression used as "predicators" or "sentence predicates in one-term deictic identification" (Abdoulaye et al. 2020).

The markers *nē and cē* may be used as copulas in both equational/specificational and predicational sentences. The following examples illustrate simple predicational (11) and specificational (12) copular sentences. Note that Green (2007) glosses c*ē* and *nē* as "focus markers".

(11) Hausa (Green 2007: 123)
wannàn màganà̀ shìrmē cè̀
this.F matter.F nonsense.M FOC.F
'This matter is nonsense.'

(12) Hausa (Green 2007: 123)
aikìn Mūsā sākà̀ nē/cē (bà kīrà̀ ba)
work.of Musa weaving.F FOC.M/FOC.F NEG smithing NEG
'Musa's work is weaving (not smithing)'

The sentence in (11) may be described as assigning the property of being nonsense to whatever matter the subject expression refers to. In (12), Musa's work is specified. In principle word-order is reversible, I will return to word-order patterns for simple specificational and predicational sentences below. Before that, I will briefly introduce clefts in the language.

Clefts in Hausa are constructed following the same pattern just described for simple specificational copular clauses. In clefts involving human referents, the cleft clause is typically introduced by the relative pronouns *wandà* (masculine), *waddà/waccè* (feminine), and *wadàndà* (plural). These are complex forms composed of an interrogative pronoun, a definite determiner, and a relativizer *dà* (Jaggar 2001: 528). In clefts involving non-human referents, a relative clause construction headed by a common noun is used. The basic word-order pattern in clefts seems to be the one where the clefted constituent (the VALUE expression) occurs in final position. This is the pattern mentioned by default in descriptions of the language (e.g. Newman 2000: 195, Jaggar 2001: 507). The order of the VARIABLE and VALUE expressions is, as in simple specificational sentences, in principle flexible. The following examples illustrate clefts in two possible word-order configurations. The cleft sentences in (13) involve a free relative clause constructions introduced by a relative pronoun, and the ones in (14) relative clause constructions involving common nouns.

(13) Hausa (Green 2007: 69)
a. [*waddà nakè̀ sô*] *Kánde cè̀*
REL.PRO.F.SG 1SG.FOC.IPFV love.VN Kande FOC.F
'The one I love is Kande.'

b. *Kánde cè̀* [waddà nakè̀ sô]
Kande FOC.F REL.PRO.F.SG 1SG.FOC.IPFV love.VN
'Kande is the one I love.'

(14) Hausa (Green 2007: 124)
a. [*àbîn dà takè bùkātà̀*] *shìnkāfā cè̀*
thing.DEF REL 3SG.F.FOC.IPFV need.VN rice.F FOC.F
'What she needs is rice.'

b. *shìnkāfā cè* [*àbîn dà takè bùkātà*]
rice.F FOC.F thing.DEF REL 3SG.F.FOC.IPFV need.VN
'Rice is what shee needs.'

Note that in either of the word-order configurations in (13) and (14) the VALUE expression in clearly marked as such. Recall that the copula (*nē/cē*) does not link two nominal expressions (as e.g. in the case of the Akan copulas discussed earlier) but marks one of the expressions involved as the VALUE. As in the case of simple copular sentences, predicational readings in sentences involving participant-oriented nominalizations seem to be available. Consider the sentence in (15).

(15) Hausa (Jaggar 2001: 458)
[*wàndà ya ràsu*] *bàbban mùtûm nē*
REL.PRO.M 3M.FOC.PFV die important.of man COP.M
'The one who died was an important man' ('The one who died was a man of importance.')

The English translation of the sentence in (15) could be interpreted as predicational or specificational. In the former case, the nominal predicate is used to ascribe a property (that of being important) to the referent of the participant-oriented nominalization. In the latter it specifies who died. Presumably, similar readings are available in Hausa.

I will return now to word order restrictions in Hausa copular clauses generally. Following the account in Green (2007), any expression marked as "focal" (i.e. followed by the marker *nē*/*cē*) can occur sentence-initially. The sentences in (16) involve proper nouns (presumably used referentially) and non-definite nominal and adjectival expressions (in principle, property-denoting nominals; cf. Newman 2000: 22, Jaggar 2001: 342 for an account of the status of adjectives in the language). In (16)(a) and (b) the "focused" (or predicative) expression is non-referential (denoting respectively an occupation and a physical property). In (16)(b) and (c), a (referential) proper name is "focused". The latter sentences may be described "quasi-specificational" (in the sense discussed in Section 2.1.5 in Chapter 2).

(16) Hausa (Green 2007: 131)
a. *dālìbī nè Audù* (*bà likità ba* (*nè*))
student.M FOC.M Audu NEG doctor NEG FOC.M
'Audu is a student (not a doctor).'

b. *dōgō nè Audù* (*bà gàjērē ba* (*nè*))
tall.M FOC.M Audu NEG short.M NEG FOC.M
'Audu is tall (not short).'

c. *Audù nè dālìbī* (*bà Bàlā ba*)
Audu FOC.M student.M NEG Bala NEG
'Audu is a student (not Bala).'

d. *Audù nè dōgō* (*bà Bàlā ba*)
Audu FOC.M tall NEG Bala.M NEG
'Audu is tall (not Bala).'

It seems that in those the cases where the non-referential expression occurs initially (a-b), a contrastive or emphatic (but not necessarily specificational) reading is triggered. In the converse case, a quasi-specificational interpretation (close to that involved in a cleft such as *The one who is tall is Andu*) is

suggested. The default pattern, however, seems to be one in which the non-referential expression marked by the copula occurs in final position (i.e. subject-predicate):

(17) Hausa (Jaggar 2001: 458)

a. *Audu* *ɗ̃àlìbī/dōgō* *nɛ̀*
Andu sutdent.M/tall.M COP.M
'Andu is a student/tall.'

b. *Kànde* *ɗ̃àlìbā/dōguwā* *nɛ̀*
Kande sutdent.F/tall.F COP.F
'Kande is a student/tall.'

In the case that the "focus"-marked expression occurs in sentence-final position, Green (2007) points out a restriction such that a property denoting expression may not occur sentence-initially:

(18) Hausa (Green 2007: 132)

a. **ɗ̃àlìbī* *Audù* *nɛ̀*
student.M Audu FOC.M
'A student is Audu.'

b. **dōgō* *Audù* *nɛ̀*
tall.M Audu FOC.M
'Tall is Audu.'

Following Green's (2007) account, then, a "quasi-specificational" interpretation (in the sense explained in Section 2.1.5 of Chapter 2) in this configuration is not possible in the language. As mentioned with respect to the examples in (16) above, this seems to be possible in the converse case, where a property-denoting expression is placed after a "focused" referential nominal expression. The restriction illustrated in the examples in (18) against "non-focused" non-referential expressions in sentence-initial position is reported to hold also in cases of topicalization. Green (2007) points out that topicalization of a non-referential expression is not possible. Consider the sentences in (19) and compare the sentences in (a) to those in (b) and (c).[64]

(19) Hausa (Green 2007: 132)

a. *Audù* (*dai*), *ɗ̃àlìbī* *nɛ̀*
Andu PTCL student.M FOC.M
'(As for) Audu, (he) is a student.'

b. **ɗ̃àlìbī* (*dai*), *Audù* *nɛ̀*
student.MPTCL Andu FOC.M
'(As for a student), (he) is Audu.'

c. **dōgō,* *Audù* *nɛ̀*
tall.M Andu FOC.M
'Tall, Audu is.'

[64] Compare the Mandarin example in f.n. 59. It seems that in Mandarin (but not in Hausa?) any expression can be topicalized/ left-dislocated.

The difference between the (non-left-dislocated) examples in (16)(c) and (d) and the one in (19)(a) lies in the presence of a distinct prosodic pattern and the (optional) presence of a modal particle[65] following the sentence-initial expression.

It is difficult to assess the pragmatic/discourse difference between the (apparently default) VARIABLE-VALUE order, in which the expression followed by the markers *nē* and *cē* occurs in second position, and the VALUE-VARIABLE, in which the order is the reverse. Perhaps the difference could be described in terms of "insistence" vs. "suspense", to use the terms of Lehmann (1984: 359). It is reported in the literature that clefts in Hausa "can be and often are focused" (Newman 2000: 195) or "can be made even more emphatic" Jaggar (2001) by "linking the subject and the complement […] with an independent pronoun + copula phrase" (p. 508). Consider the sentences in (20), where (a) is a cleft (with VARIABLE-VALUE order) and (b-c) are "emphatic" or "focused" cleft alternatives exhibiting two alternative word orders.

(20) Hausa (Jaggar 2001: 508)

a. [*àbîn dà a-kè̀ bùkātà̀*] *ruwan famfò̀ nè̀*
thing REL one-FOC.IPFV. need water pipe COP.M
'What is needed is piped water.'

b. [*àbîn dà a-kè:̀ bùkātà̀*] *shī nè̀ ruwan famfò̀*
thing REL one-FOC.IPFV need 3SG.M COP.M water pipe
'What is needed is PIPED WATER.'

c. *ruwan famfò̀ shī nè̀* [*àbîn dà a-kè̀:* *bùkātà̀*]
water pipe 3SG.M COP.M thing REL one-FOC.IPFV need
'PIPED WATER is what is needed.'

There is then a contrast between (20)(a) and (b-c) in terms of "emphasis". Again, it is not clear what the difference between (20)(b) and (c) would be. Presumably the default pattern is that in (20)(b). Newman (2000) does not mention the word order pattern illustrated in (20)(c) and Jaggar (2001) merely notes that it is also possible.

4.1.4 Korean

Clefts (and simple specificational sentences) in Korean allow both a VARIABLE-VALUE and a VALUE-VARIABLE order. In the latter configuration, however, the marking of the terms is subject to restrictions which are not present in the former. I will begin the discussion with a review of the general patterns involved in the construal of specificational sentences (including clefts) in the language (see also Section 2.3.1.3 in Chapter 2; note that much of what will be discussed in the following applies to Japanese as well).

In Korean, the copula is verbal and occurs clause-finally. The immediately pre-copular expression is (in most cases) bare (i.e. not case-marked) and the "subject" expression is generally either marked with a topic marker (*-i*/*-ka*) or a nominative marker (*-un*/*-nun*).[66] The basic pattern is similar to that of a verbal sentence (word order in the language is canonically SOV). There seems to be in principle a basic pattern for predicational and specificational sentences in terms of word order. In the case of predicational sentences the

[65] "The MP dai 'just, only' serves as a limiter/restricter [sic] or as a contrastive or corrective marker. It can also be used to soften the abruptness of a statement" (Newman 2000: 327).

[66] The choice of the allomorphs of the markers *-i*/*-ka* and *-un*/*-nun* is conditioned by the shape of the word which these attach to.

order is subject-predicate, and in the case of specificational sentences VARIABLE-VALUE. Consider the predicational sentence in (21)(a) and the specificational one in (b):

(21) Korean (Chung 2016: 19)
a. *John-un* *uysa-i-ta.*
John-TOP doctor-COP-DECL
'John is a doctor.'

b. *pemin-un* *John-i-ta.*
culprit-TOP John-COP-DECL
'The culprit is John.'

The sentences in (21) involve a sentence-initial nominal expression marked with the topic marker *-un*/*-nun* and an expression in predicate (i.e. immediately pre-copular) position. In the predicational sentence (21)(a), the topic-marked expression is fully referential and the nominal expression in predicate position denotes a property that is assigned to the referent of the topic expression. In the sentence in (21)(b), the topic-marked nominal expression denotes an individual (the culprit of a specific crime) whose reference is specified by the proper name in predicate position. Note that the nominal expressions are not (obligatorily) formally marked for referentiality or definiteness in the language. That is, a bare noun may be interpreted as referential (definite or indefinite, specific or non-specific), or non-referential. The English translations provided by the authors are meant to suggest the intended interpretation of the expressions in Korean.

Apparently, to obtain a predicational reading, a property-like nominal expression must occur in predicate position. Word-order inversion is not possible. Consider the sentence in (22). The assumption is that *uysa* 'doctor' is interpreted as non-referential.

(22) Korean (Chung 2016: 21)
**uysa-nun/ka* *John-i-ta.*
doctor-TOP/NOM John-COP-DECL
'A doctor is John.'

The sentence in (22) is unacceptable under the intended predicational reading. Under a referential interpretation of the expression *uysa-nun* 'the doctor' the sentence would be a specificational sentence (in principle like (21)(b) 'The culprit is John') To clarify: It is apparently not possible to attribute a property to a referent with the fully referential expression in immediately pre-copular position using an expression with a property-like interpretation in sentence-initial position. A (non-referential) nominal predicate must occur in (immediately) pre-copular position.

In specificational sentences, word order is reversible but the marking of the expression in initial position is subject to restrictions (this includes both the choice between a topic and a nominative marker, and prosodic realization). Note that (L) in the examples below indicates a neutral prosodic realization. This contrasts with (H) which indicates a prosodic prominence, realized as a relatively high pitch accent (Chung 2016). For the purposes of the present discussion, the prosodic contrast is relevant for nominative-marked expressions only. Consider the sentences in (23). Both are specificational. The one in (23)(a) illustrates the (arguably basic) VARIABLE-VALUE order and the one in (23)(b) the (pragmatically marked) VALUE-VARIABLE order.

(23) Korean (Chung 2016: 22)

a. *pemin-un*(L) *John-i-ta*.
culprit-TOP John-COP-DECL
'The culprit is John.'

b. *John-i*(H) *pemin-i-ta*.
John-NOM culprit-COP-DECL
'John is the culprit.'

According to Chung (2016), for a sentence such as (23)(b) to allow a specificational reading, the sentence-initial expression must be nominative-marked (i.e. not topic-marked), and it must be made prosodically prominent (H). If the sentence-initial expression is topic-marked or if it is not made prosodically prominent, the sentence is interpreted as predicational.

Chung (2016: 23) argues, however, that high pitch nominative marking does not automatically yield a specificational sentence. Consider the sentences in (24), which are both described by the author as predicational. (But with different pragmatic flavors; the one (24)(b) would qualify as "quasi-specificational", in the sense discussed in Section 2.1.5 in Chapter 2.)

(24) Korean (Chung 2016: 23)

a. *John-i*(L) *uysa-i-ta*
John-NOM doctor-COP-DECL
'John is a doctor.'

b. *John-i*(H) *uysa-i-ta*
John-NOM doctor-COP-DECL
'Only John is a doctor.'

In both cases, the bare noun in immediately pre-copular position is assumed to be interpreted as non-referential. Assuming a non-referential interpretation of *uysa* 'doctor', then, both sentences would, in Chung's terms, qualify as predicational. The difference between the interpretations with (H) and (L) prosodic realizations is described by the author in terms of exhaustivity. Note that this is a pragmatic effect; there is no Korean counterpart to English *only* in the sentence. To clarify: Chung considers a sentence such as (24)(b) above to be in principle comparable with a sentence with a (stative) verbal predicate such as (25).

(25) Korean (Chung 2016: 23)
John-i(H) *paykophu-ta*
John-NOM hungry-DECL
'It is John who is hungry.' (exhaustive)
Inference: Others are not hungry.

The difference between a sentence such as (24)(b) and one such as that in (25) would mainly be that in the former case the sentence requires a copula because the predicate is a nominal one. In the latter case, the predicate does not require a copula. The inference that nobody but the referent of the subject expression is being ascribed a certain property follows from emphasizing (i.e. "highlighting", "making prominent", or "focusing") the subject expression. This effect is apparently only possible with the nominative marker.

With respect to the alternation between the marking of the sentence-initial expression with a topic particle and high pitch nominative case suffix, clefts pattern in principle like simple specificational sentences.

Consider the example in (26) illustrating the alternation in cleft sentences. The sentence in (26)(a) illustrates the VARIABLE-VALUE pattern and the one in (26)(b) the VALUE-VARIABLE one.

(26) Korean (Chung 2016: 21)

a. [*John-i* *mek-un* *kes-un*] *panana-i-ta*
John-NOM eat-REL thing-TOP banana-COP-DECL

b. *panana-ka*(H) [*John-i* *mek-un* *kes*]-*i-ta*
banana-NOM John-NOM eat-REL thing-COP-DECL

'Bananas are what John ate.'

The VALUE expression may occur either in immediately pre-copular position, as in (26)(a), or in sentence-initial position, and high pitch nominative-marked as in (26)(b).

Recapitulating: There is an asymmetry between predicational and specificational sentences. In predicational sentences, the subject expression may take either a topic (*-un/-n*) or a nominative marker (*-ka/-i*), and the nominal predicate necessarily occurs in immediately pre-copular position (hence the unacceptability of (22)). In specificational sentences there are two possible patterns. The VARIABLE expression may occur sentence initially and take either a topic or a (low pitch) nominative marker and the VALUE expression occurs in immediately pre-copular position. (This pattern is parallel to that observed in predicational sentences.) Alternatively, the VARIABLE expression may occur in immediately pre-copular position (i.e. in predicate position) but then the VALUE expression (in subject position) must take a high pitch nominative marker. The pattern is illustrated again in the following examples embedded in a mini discourse. Consider first the question-answer pair in (27), where the expected answer would be a predicational sentence. Jo (2007) does not provide an English translation for the answers, I will come back to this.

(27) Korean (Jo 2007: 114)
John-i *cikep-i* *mwues-i-ni?*
John-NOM job-NOM what-COP-Q
'What is John's occupation?

a. *John-i/-un* *mokswu-i-a*
John-NOM/TOP carpenter-COP-DECL

b. **mokswu-ka* *John-i-a*
carpenter-NOM John-COP-DECL

In the appropriate answer (27)(a), the (referential) subject expression occurs in subject position and may take either a nominative or topic marker. The (non-referential) nominal predicate may only occur in immediately pre-copular position. The sentence in (27)(b) is, according to Jo (2007), unacceptable. Now consider the question-answer pair in (28).

(28) Korean (Jo 2007: 114)
nwu-ka *mokswu-i-ni?*
who-nom carpenter-COP-Q
'Who is a carpenter?'

a. *John-i* *mokswu-i-a*
John-NOM carpenter-COP-DECL

b. *mokswu-ka/nun* *John-i-a*
carpenter-NOM/-TOP John-COP-DECL

c. ?[*mokswu-i-n* *ken*] *John-i-a*
carpenter-COP-ADN *KES*.TOP John-COP-DECL

In this case, word order reversibility is possible. In (28)(a) the VARIABLE expression *mokswu* 'carpenter' occurs in immediately pre-copular position and the VALUE expression takes a (presumably high pitch) nominative case marker. Alternatively, as in (28)(b), the VARIABLE expression occurs in sentence-initial position and takes either a (presumably low-pitch) nominative or topic marker and the VALUE expression occurs in immediately pre-copular position. Jo (2007) notes that the alternative in (28)(c), a cleft, is "also acceptable though slightly marginal" (p. 114). Note that, interestingly, Jo (2007) does not translate the possible answers to the questions in the examples (27) and (28) into English. Elsewhere, the author translates a sentence similar to the topic-marked variant of (28)(b) as in (29)(b) below:

(29) Korean (Jo 2007: 213)
a. *Chelswu-nun* *pwuca-i-a*
Chelswu-TOP rich.person-COP-DECL
'As for Chelswu, he is rich.'

b. *pwuca-nun* *Chelswu-i-a*
rich.person-TOP Chelswu-COP-DECL
'As for being rich, it is Chelswu (who is rich).'

As suggested by the English translations, a topic-marked expression may be only loosely integrated with the rest of the sentence. In principle, it seems not to be part of the copular clause at all. This may be the case, but note that the author also suggests a variant in (28)(b) (literally: 'carpenter is Chelswu') in which the sentence-initial expression takes a (presumably low-pitch) nominative case marker. In this case, a "loosely integrated" topic interpretation would presumably not be necessary. Now, the author does not consider the expression *mokswu-ka* 'carpenter-NOM' in (28)(b) to be a referential expression but a nominative-marked nominal predicate. It may be better to describe the expression as referential (though weakly so) and not restrict the status of referential expression to fully referential expressions (in the sense discussed in Section 2.3.1 in Chapter 2). Then, perhaps, a generalization could be made such that only referential expressions take nominative marking (though this would require clarification). Word order flexibility could be said then to hold only for sentences involving two referential expressions. If this generalization turned out to be correct, the status of true (non-referential) nominal predicates may be restricted to expressions occurring in immediately pre-copular position (as suggested e.g. in Kim 2015: 13).

4.1.5 Amharic

A further account of word order-variability in specificational sentences (VARIABLE-VALUE vs. VALUE-VARIABLE) is reported in Kramer & Eilam (2012) for Amharic. The basic pattern seems to be the former. In Amharic, the canonical word order is SOV. The pattern is illustrated in the verbal sentence in (30).

(30) Amharic (Kramer & Eilam 2012: 76)
astämari-w *doro* *wät'-u-n* *bälla*
teacher-DEF chicken stew-DEF-ACC ate.3SG.M
'The teacher ate the chicken stew.'

In principle, one constituent may occur post verbally.[67] The following verbal sentences illustrate this pattern. In (31)(a), a subject expression occurs post-verbally, in (b) an object.

(31) Amharic (Kramer & Eilam 2012: 81)

a. *doro* *wät'-u-n* *bälla* *astämari-w*
chicken stew-DEF-ACC ate.3SG.M teacher-DEF
'The teacher ate the chicken stew.' (response to the question: 'What did the chicken eat')

b. *astämari-w* *bälla* *doro* *wät'-u-n*
teacher-DEF ate.3SG.M chicken stew-DEF-ACC
'The teacher ate the chicken stew.' (response to the question: 'Who ate the chicken stew.')

Kramer & Eilam (2012) note that intuitions vary among native speakers concerning the status of expressions following the (canonically clause-final) verb. Apparently, to some speakers, these have the feel of after-thoughts. The authors note that the sentences in (31) may be appropriate answers to different questions. The postposed constituent corresponds to background information. The constituent in immediately pre-verbal position is emphasized.

The authors note that

> verb-medial word order is non-canonical in Amharic. It is never required (to the best of [their] knowledge), and it is much less common than verb-final order in both written texts and spoken elicitation. It is most frequently found in clefts, where the copula comes between the focused constituent and the relative clause […] Verb-medial order is also attested in wh-questions fairly often […] with the more typical verb-final order first (note that Amharic is a wh-in-situ language generally). (Kramer & Eilam 2012: 76)

In a constituent question, the interrogative pronoun may occur in the position corresponding to its role/function in the sentence. It cannot, however, occur in post-verbal position. The following word-order patterns are reported to be felicitous in Amharic:

(32) Amharic (Kramer & Eilam 2012: 77)

a. *man* *doro* *wät'-u-n* *bälla*
who chicken stew-DEF-ACC ate.3SG.M
'Who ate the chicken stew?'

b. *man* *bälla* *doro* *wät'-u-n*
who ate.3SG.M chicken stew-DEF-ACC
'Who ate the chicken stew?'

An interrogative pronoun in post-verbal is not:

(33) Amharic (Kramer & Eilam 2012: 81)

a. *mïndïn* *anäbbäb-äcc* *Tïgïst*
what read-3SG.F Tigist
'What did Tigist read.'

[67] According to Kramer & Eilam (2012), only one.

b. **Tïgïst anäbbäb-äcc mïndïn*
Tigist read-3SG.F what
'What did Tigist read?

The examples in (34) illustrate cleft sentences in Amharic. The verb in the cleft clause in (34) is distinguished from a main clause verb by the prefix *yä-* and a definite marker (identical to that suffixed to a common noun). The same form could be used as and adnominal modifier. The structure of the clause is in principle identical to a free clause (the only difference being the form of the verb). Kramer & Eilam (2012) point out that the sentence in (34)(a) seems to have "the flavor of a pseudocleft", and that the one in (34)(b) "is similar to an *it*-cleft." (p. 77).

(34) Amharic (Kramer & Eilam 2012: 81)

a. [*doro wät'-u-n yä-bälla-w*] *astämari-w näw*
chicken stew-DEF-ACC C-ate.3SG.M-DEF teacher-DEF is
'The one who ate the chicken stew is the teacher.'

b. *astämari-w näw* [*doro wät'-u-n yä-bälla-w*]
teacher-DEF is chicken stew-DEF-ACC C-ate.3SG.M-DEF
'The one who ate the chicken stew is the teacher.

Now note that the clefted constituent cannot occur post-verbally:

(35) Amharic (Kramer & Eilam 2012: 81)
*[*doro wät'-u-n yä-bälla-w*] *näw astämari-w*
chicken stew-DEF-ACC C-ate-3SG.M-DEF is teacher-DEF
Attempted: 'It is the teacher who ate the chicken stew.'

The information structure is fixed. A specificational reading only obtains if the VALUE expression occurs in pre-copular position. The same applies in simple specificational sentences:

(36) Amharic (Kramer & Eilam 2012: 81)

a. *dʒon näw gäday-u*
John is killer-DEF
'The killer is John.'

b. **gäday-u näw dʒon*
killer-DEF is John
Attempted: 'The killer is John.'

Presumably the reading 'John is the Killer' (e.g. answering the question who is John?) could obtain in (36)(b), though this would require clarification. Note that a (quasi-?) specificational reading obtains if an indefinite is post verbal. Consider the sentence in (37)(b).

(37) Amharic (Kramer & Eilam 2012: 81)

a. *Käbbädä astämari näw*
Kebbede teacher is
'Kebbede is a teacher.'

b. *Käbbädä näw astämari*
Kebbede is teacher
'It's Kebbede who is a teacher.'

Kramer & Eilam (2012) also show the information-structural interaction of clefting with that of left-dislocation, where the left-dilocated expression is contrasted. Consider the following sentence:

(38) Amharic (Kramer & Eilam 2012: 89)
lä-Tïgïst, <u>Haylu</u> näw [kärämella-w-ïn y-asayy-ä-w]
DAT-Tigist Haylu is candy-DEF-ACC C-show-3SG.M-DEF
'To Tigist, it's Haylu who showed the candy.'

The authors note that it is "unclear whether these [(38)] structures are grammatical for non-clefts" (p. 14).

4.2 Dislocated, correlative, and correlative-like constructions

In this section, I will discuss data from languages in which cleft sentences regularly involve the dislocation of the VARIABLE expression and its resumption by means of a pronominal element. I have already discussed (arguably) related constructions elsewhere (see discussion on modern Hebrew in 2.3.1.2 in Chapter 2, and Hausa in 4.1.3 in this chapter). Given the relevance of the pattern in question in the context of cross-linguistically oriented an investigation on clefts (and especially for the discussion which will take place in Chapter 5), I will examine the pattern more closely. Consider the example in (39) (a French *ce qu-…c'est…* cleft or "pseudo-cleft").

(39) French (www)
[*Ce que je veux*], {*c'est un <u>vrai travail</u>*}.
DEM REL 1SG want.1SG.PRS DEM_be.3SG.PRS INDF.SG.M true.SG.M work
'What I want is a real job.'

The clause to the right (in curly brackets) in (39) may be regarded in principle an independent clause. Within this clause, a demonstrative in pre-copular (subject) position resumes the left-dislocated VARIABLE expression. The VALUE expression occurs in post-copular (predicate) position. Note that the same basic pattern can be observed in a specificational sentence in which the variable expression does not involve a clausal nominalization, as illustrated in (40).

(40) French (www)
Les pires, {*ce sont <u>les saumons roses et les harengs</u>*}.
DEF.PL.M worst.PL DEM be.PL.PRS DEF.PL salmon.PL pink.PL and DEF.PL hering.PL
'The worst are the pink salmon and the hering.'

I will return in 5.4.2 in Chapter 5 to a discussion of French constructions such as the one illustrated in (39). In the following sections, I will discuss data from other languages in which cleft constructions (and specificational sentences generally) are reported to involve similar patterns. I will begin in 4.2.1 with a discussion of Hindi data and discuss the problem concerning the nominal status of the dislocated constituent in correlative constructions.[68] In 4.2.2 and 4.2.3, I will discuss accounts of similar patterns in Lithuanian

[68] Note that in French the left-dislocated expression in a *ce qu-…c'est…* cleft has the potential distribution of a nominal expression. Consider the following example:

and Hungarian. In 4.2.4, I will discuss accounts of related patterns reported for Classical Nahuatl and Ixquihuacan Nahuatl. The resumptive element in Classical Nahuatl does not occupy the position associated with the subject, but rather the position associated with the predicate in the language. In Ixquihuacan Nahuatl, the resumptive pronoun is reported to develop into a copula.

4.2.1 Hindi

An oriented clausal nominalization in Hindi may be formed by means of a participle in association with lexical nouns, as illustrated in (41). Such constructions seem to have in principle the potential distribution of nominal expressions.[69]

(41) Hindi-Urdu (Bhatt 2015: 724)

[[*Atif-ko*	*lag-ii*]	*coṭ*]]	*bahut*	*gahrii*	*hai*
Atif-DAT	feel-PFV.F	hurt.F	very	deep.F	be.PRS.SG

'The injury that Atif has is quite deep.'

Constructions following the pattern illustrated in (41) are often referred to in the literature to as participial or non-finite relative clauses. A pattern more commonly discussed in accounts of relative clause constructions in Indo-Aryan languages involves the combination of two clauses usually described as correlative (or relative-correlative) constructions. In these constructions, the relative clause involves finite predicates (i.e. predicate forms identical to those used in independent clauses[70]) This pattern is also common in older Indo-European languages (see e.g. Lehmann 1984; 1988). An example from Hindi is given in (42):

(42) Hindi (Junghare 1996: 223)

{*jo*	*aam*	*pulaa*	*thaa*}	*vah*	*maine*	*khaayaa*
which	mango	yellow	was	that	I	ate

'I ate [the (that) mango which was yellow].'

The sentence-initial relative clause (marked here in curly brackets) is introduced by a specialized form. The relative pronoun *jo* may in principle inflect distinguishing nominative, accusative/dative, ergative and oblique forms, and in the non-nominative forms singular and plural. Consider the following sentence, involving the ergative singular forms *jis* (relative) and *us* (correlative/non-proximal demonstrative).

(43) Hindi (Junghare 1996: 229)

{*jis aadmii ne*	*ghar*	*jalaa diyaa*},	*us vyakh ne*	*paaThshaalaa*	*bhii*	*Jalaa dii*
who man	house	burned	that person	school	too	burned

'The man who burned the house (that person) burned the school, too.'

Unlike the forms *which* and *who* in the English glosses of the relative pronouns in the examples (42) and (43), Hindi relative pronouns are not identical to interrogative pronouns and are used exclusively in relative clauses. Correlative pronouns (such as nominative *vah* in example (42) and ergative *us* in (43)) are in principle non-proximal demonstratives, which may be used pronominally (42) or adnominally (43).

(i) French (www)

Il	*a*	*acheté*	[*ce*	*que*	*nous*	*voulions*].
3SG	have.3SG.PRS	buy.PTCP	DEM	REL	1PL	want.1PL.IPFV

'He bought what we wanted.'

As will be discussed in 4.2.1 below, it seems this does not necessarily hold for (all) Hindi correlative constructions.

[69] Bhatt (2015) analyzes the expression in square brackets in the sentence in (41) as a determiner phrase.

[70] Verbal sentences in Hindi often involve auxiliary and light verb constructions in which the lexical verb is either a participle or an action noun.

Correlative clauses in relative-correlative constructions have in principle the same shape as independent clauses with a non-proximal demonstrative as one of its arguments. The relative clause, in contrast, is unambiguously marked as a dependent clause. It cannot, however, be regarded as a constituent of a superordinate clause in a straightforward way. At least according to the characterization of nominal expressions adopted in this dissertation (based on distributional potential), the status of the relative clause in a correlative sentence as a nominal expression is problematic, because an expression of this shape does not have the distributional potential of a simple nominal expression.

A correlative cleft (or, perhaps, cleft-like construction) involves a structure similar to the one illustrated in examples (42) and (43), with the particularity that the correlative part involves a copular clause. Within the correlative clause, the correlative pronoun may be regarded as the VARIABLE expression and the immediately pre-copular nominal expression as the VALUE. Examples of this type of constructions are given in (44) and (45). Note that the order of the relative and correlative parts is reversible.

(44) Hindi (Junghare 1996: 224)

a. {*jo aadmii vahaa khaRaa hai*} *vah aadmii raam kaa mitra hai*
who man there standing is that man Ram of friend is
'The man who is standing there is Ram's friend.'
('A man is standing there, that man is Ram's friend.')

b. {*jo aadmii vahaa khaRaa hai*} *vah raam kaa mitra hai*
who man there standing is that Ram of friend is
'The man who is standing there is Ram's friend.'

c. {*jo vahaa khaRaa hai*} *vah raam kaa mitra hai*
who there standmg is that Ram of friend is
'The one who is standing there is Ram's friend.'

(45) Hindi (Junghare 1996: 225)

a. *vah aadmii raam kaa mitra hai* {*jo aadmii vahaa khaRaa hai*}
that man Ram of friend is who man there standing is
'The (that) man is Ram's friend who is standing there.'

b. *vah aadmii raam kaa mitra hai* {*jo vahaa khaRaa hai*}
that man Ram of friend is who there standing is
'The (that) man is Ram's friend the one who is standing there.'

c. *vah raam kaa mitra hai* {*jo vahaa khaRaa hai*}
that Ram of friend is who there standing is
'He is Ram's friend who is standing there.'

Again, note that nouns following the relative (and correlative) pronouns may be involved in these constructions. The relative and correlative pronouns may thus be used adnominally.[71]

Beside the patterns illustrated in (44) and (45), a pattern is possible in which the relative clause follows a nominal expression, as illustrated in the examples in (46).

[71] Junghare (1996) suggests, however, that the noun may not be present in the second part if it is absent in the first. This, independently of whether the order is relative-correlative as in (44) or correlative-relative as in (45). (Note that in both (44) and (45) the noun *aadmii* 'man', if not present in both parts, is present only in the first one.)

(46) Hindi (Junghare 1996: 226)

a. *vah aadmii, {jo vahaa khaRaa hai} vah meraa bhaaii hai*
that man who there standing is that my brother is
'The man, who is standing there, is my brother.'

b. *vah, {jo vahaa khaRaa hai}, vah meraa bhaaii hai*
that who there standmg is that my brother is
'He, who is standing there, is my brother.'

The English translations provided by Junghare to the sentences in (46) allow both nominal predicate and specificational readings. The author notes, however, that the sentences in (46) (a) and (b) may be used to answer the questions in (47) (a) and (b) respectively.

(47) Hindi (Junghare 1996: 227)

vah aadmii, {jo vahaa khaRaa hai}, vah kaun hai?
that man who there standmg is he who is
'Who is that man who is standing there?'

vah, {jo vahaa khaRaa hai}, vah kaun hai?
that who there standing is he who is
'Who is he, the one who is standing there?'

This being the case, the specificational status (or at least the availability of a specificational reading) of the sentences in (46) seems to be confirmed. Now, it seems that the relative part (including the (pro-)nominal expression preceding the relative clause in a narrow sense) in the sentences presented in (47) does have the potential distribution of a nominal expression. Consider the example in (48) in which the relative clause construction occurs embedded in a superordinate clause.

(48) Hindi-Urdu (Bhatt 2015: 711)

mujhe [vo aadmii [jo Sita-ko pasand hai]] accha: nahĩ: lag-ta:
1.DAT that man REL Sita-DAT like be.PRS.SG like NEG seem-IPFV.SG
'I don't like the man who Sita likes.'

Puri (2011) suggests that the pattern in which the relative clause follows a nominal expression, as in (48), is due to language contact. The author argues that such constructions are not found in Hindi texts prior to intensive contact with the British, and that most of the attested occurrences prior to the 20th century are found on English-Hindi translations by non-native speakers of Hindi.

4.2.2 Lithuanian

In Lithuanian, clefts involving free relative clauses (but apparently only this kind) differ from regular specificational sentences and involve a correlative construction. Unlike regular specificational sentences and clefts involving noun-supported oriented nominalizations (i.e. headed relative clauses), this kind of constructions follows an archaic pattern not involving a verbal copula. Regular specificational sentences are construed with a verbal copula and allow word-order variation, though a VARIABLE-VALUE configuration seems to be the default word order pattern, paralleling the default subject-predicate pattern in predicational copular sentences in the language.

A simple specificational sentence is described as consisting of two nominal expressions connected by the verbal copula *būti* 'be'. Consider the specificational sentence in (49).

(49) Lithuanian (Mikulskas 2016: 98)

Músų	*anglų*	*kalbos*	*mokytojas*	*yra*	*Jonas*	*Petraitis*
our	English	language.GEN.SG	teacher.NOM.SG	be.3.PRS	Jonas.NOM	Petraitis.NOM

'Our English teacher is Jonas Petraitis.'

A few remarks about Mikulskas' (2016) account on regular specificational sentences in Lithuanian are in order. Note that there is some variability in the case marking of the VARIABLE expression, which can be marked for the nominative or instrumental case, as illustrated in (50).

(50) Lithuanian (Mikulskas 2016: 121)

Mūsų	*anglų*	*kalbos*	*mokytojas/mokytoju*	*yra*	*Jonas*	*Petraitis.*
our	English	language.GEN.SG	teacher.NOM.SG/INS.SG	be.3.PRS	Jonas.NOM	Petraitis.NOM

'Our English teacher is Jonas Petraitis.'

Case alternation between the nominative and the instrumental is assumed to be a property of nominal predicates in verbal copular sentences. The subject is always in the nominative. In Lithuanian specificational sentences such as (50), the VARIABLE expression corresponds to the predicate and the VALUE expression to the subject (in formal terms). In Section 2.2.1 in Chapter 2, I mentioned that, cross-linguistically, the way an expression in predicative function (in a loose sense) is coded may be different for (true) nominal predicates and nominal expressions in equational sentences. Mikulskas (2016) argues against making a distinction between predicational (i.e. nominal-predicate) and "equative" (i.e. equational) sentences in the language on the basis of case marking. (Note that the author considers specificational sentences as a type of equational sentences.) Historically, the instrumental case is not associated with equational sentences, but neither is the verbal copula. In the modern language, Mikulskas argues, the alternation between nominative and instrumental case marks aspectual distinctions and not a distinction between predicational and specificational (equational) sentences.

Word order in specificational sentences in Lithuanian is reversible. Consider the specificational answer sentences in (51).

(51) Lithuanian (Mikulskas 2016: 145)

Kas	*yra*	*fūsų*	*miesto*	*meras?*
who.NOM	be.3.PRS	your	town.GEN.SG	mayor.NOM.SG

'Who is the mayor of your town?'

a.

Mūsų	*miesto*	*meras*	*yra*	*Petras*	*Jonaitis*.
our	town.GEN.SG	mayor.NOM.SG	be.3.PRS	Petras.NOM	Jonaitis.NOM

'The mayor of our town is Petras Jonaitis.'

b.

Petras	*Jonaitis*	*yra*	*mūsų*	*miesto*	*meras.*
Petras.NOM	Jonaitis.NOM	be.3.PRS	our	town.GEN.SG	mayor.NOM.SG

'The mayor of our town is Petras Jonaitis.'

Both answers in (51) are specificational and the VALUE containing expression is the proper noun *Petras Jonaitis*. Mikulskas (2016) notes that "the second answer [(51)(b)] is its pragmatically marked version and is more likely when information unexpected to the hearer is emphasized" (p. 145). Thus, the basic order in a specificational sentence may be the one where the VARIABLE expression occurs in the pre-copular position. The canonical word order in the language seems to be subject-predicate. Consider the sentences in (52).

(52) Lithuanian (Mikulskas 2016: 194)

a. *Vaikas* *miega*.
child.NOM.SG sleep.3.PRS
'The child sleeps.'

b. *Vaikas* *yra* *mieguistas.*
child.NOM.SG be.3.PRS sleepy.NOM.SG.M
'The child is sleepy.'

c. *Vaikas* *yra* *dar* *darželinukas*.
child.NOM.SG be.3.PRS still kindergartener.NOM.SG.M
'The child is still a kindergartener.'

d. *Vaikas* *yra* *savo* *kambaryje*.
child.NOM.SG be.3.PRS his.RPO room.LOC.SG.M
'The child is in his room.'

Mikulskas (2016: 18) notes that the verb *būti* 'be' originally had an existential meaning. In the modern language it is used with all sorts of non-verbal predicates (as illustrated in examples (52)(b-d) above) and in equational sentences. Some expressions such as general (gnomic) statements, definitions, and deictic contexts regularly involve juxtaposition of nominal expressions (p. 2). [72] Consider the following sentences:

(53) Lithuanian (Mikulskas 2016: 2)
Ponas *ne* *brolis.*
squire.NOM.SG NEG brother.NOM.SG
'A squire is not a brother.'

(54) Lithuanian (Mikulskas 2016: 2)
Štai *tas* *žmogus* *—* *mūsų* *išsigelbėjimas*.
PRST that.NOM.SG.M person.NOM.SG our rescue.NOM.SG
'This man here is our rescue.'

Beside the possibility of using the verbal copula *būti* 'be', and juxtaposition of nominal expressions, there is a copula (or copula-like) element of pronominal (demonstrative) origin identical to the neuter (uninflected) demonstrative *tai*. Its use is illustrated with an adage in the sentence in (55).

(55) Lithuanian (Mikulskas 2016: 2)
O *jaunystė* *be* *svajonių* *—* *tai* *paukštis* *be* *sparnų.*
but youth.NOM.SG without dream.GEN.PL that.N bird.NOM.SG without wing.GEN.PL
'But youth without dreams is a bird without wings.'

This pattern is possible in equational sentences in general. Identity statements can also be formed with a verbal copula (or with both the demonstrative and verbal copula). Consider the alternatives in example (56).

(56) Lithuanian (Mikulskas 2016: 60)

a. *Markas* *Twainas* (*tai* *ir*) *yra* *Samuelis* *Clemensas*.
Mark.NOM Twain.NOM that.N PRTC be.3.PRS Samuel.NOM Clemens.NOM
'Mark Twain is [the same person as] Samuel Clemens.'

[72] Note that the dash in example (54) is a Lithuanian orthographic convention.

b. *Markas* *Twainas —* *tai* *Samuelis* *Clemensas*.
Mark.NOM Twain.NOM that.N Samuel.NOM Clemens.NOM
(same meaning)

Apparently, simple predicational sentences can also be formed with the pronominal copula *tai*. Consider the sentence in (57).

(57) Lithuanian (Mikulskas 2016: 61)
Jonas *tai* *geras* *krepšininkas.*
John.NOM that.N good.NOM.SG.M basketball.player.NOM.SG
'John is a good basketball player.'

To judge from Mikulskas' (2016) account, this is not the main strategy used in the modern language to construe copular sentences (copular in a general sense, i.e. including predicational and equational sentences). Instead, the main pattern seems to be the verbal copula strategy. But the verbal copula strategy, again, to judge from Mikulkas' (2016: 59) account, seems to be (strictly speaking) necessary only in TAM other than the present indicative, as the verbal copula is the bearer of TAM.

The (invariant-neuter) pronominal form *tai* can also be used in left-dislocated constructions. In Mikulska's (2016) terms, the demonstrative in these cases "performs the function of segmentation of the information structure by marking the so-called fronted topic" (p. 61). The pattern is illustrated in example (58).

(58) Lithuanian [Mikulskas 2016: 61)
Cepelinų *tai* *aš* *nemėgstu*.
potato.dumpling.GEN.PL that.N 1.SG.NOM NEG.like.1SG.PRS
'Potato dumplings I dont like.'

Note that in this case there is no copular relationship between the constituents left and right of the demonstrative. The string following the demonstrative is a finite clause. The left-dislocated expression (*cepelinų*) is inflected according to the case required by the verb. The demonstrative, which could be interpreted as a resumptive pronoun remains uninflected (*tai* does not inflect for case or number). Note that in Mikulskas' (2016) account, clefts of the free relative variety (which I will discuss below) are associated with this type of constructions.

Mikulskas (2016) describes as clefts ("pseudo-clefts", to be precise) constructions consisting of a sentence-initial free relative clause (introduced by an inflected interrogative pronoun) followed by the neuter demonstrative *tai,* itself followed by an expression corresponding to the relativized argument (or adjunct). Consider the examples in (59) and (60):[73]

(59) Lithuanian (Mikulskas 2017: 122)
[*Ko* *aš* *nemėgstu*], *tai* *cepelinų.*
what.GEN 1SG.NOM dislike.1SG.PRS that.N potato.dumpling.GEN.PL
'What I dislike is potato dumplings.'

[73] The commas preceding or following a relative seem to reflect a Lithuanian orthographic convention not (necessarily) any prosodic pattern.

(60) Lithuanian (Mikulskas 2017: 122)

[*Ką*	*aš*	*dievinu*],	*tai*	*šokoladą*.
what.ACC	1SG.NOM	adore.1SG.PRS	that.NOM	chocloate.ACC.SG

'What I adore is chocolate.'

In (59) and (60), the sentence final expressions are nominal expressions inflected for the case required by the verb in the headless relative (I will return to this point below). The sentences in (59) and (60) resemble correlative constructions (as found e.g. in Indo-Aryan languages). There is a difference, however. Whereas in a language like Hindi the second part of a correlative construction equivalent to the sentences in (59) and (60) would involve (perhaps require) a copula, it seems to be impossible to use the verbal copula (instead or in addition to the demonstrative *tai*) in the Lithuanian cleft. This is explicitly noted by Mikulskas (2016). Consider the (unacceptable) sentence in (61).

(61) Lithuanian (Mikulskas 2017: 122)

**Ko*	*nemėgstu,*	(*tai*)	*yra*	*cepelinai.*
what.GEN	NEG.like.1sg.PRS	that.N	be.3.PRS	potato.dumplings.NOM.PL

Intended: 'What I hate is potato dumplings.'

Note, however that the use of the verbal copula is possible if the VARIABLE expression is a headed relative clause construction as in (62) and (63), where it is headed by a demonstrative pronoun.

(62) Lithuanian (Mikulskas 2017: 123)

[*Tai,*	*ko*	*nemėgstu*],	*yra*	*cepelinai*.
that.N	what.GEN	NEG.like.1SG.PRS	be.3.PRS	potato.dumpling.NOM.PL

'That which I like is potato dumplings.'

(63) Lithuanian (Mikulskas 2017: 123)

[*Tai,*	*ką*	*aš*	*dievinu*],	*yra*	*šokoladą*.
that.N	what.ACC	1SG.NOM	adore.1SG.PRES	be.3.PRS	chocloate.NOM.SG

'What I adore is chocolate.'

In these sentences, the headed relative clause construction seems to be treated as the subject, and the predicative part (a complement to the verbal copula) does not show case attraction. It is not clear what status Mikulskas (2016) assigns to such constructions. The author refers to the examples in (62) and (63) as "rearrangements of pseudoclefts" (p. 123). Apparently, such "rearrangements" do pattern like regular specificational sentences. Not so, according to Mikulskas (2016), clefts with free relatives (i.e. the constructions I describe as correlative clefts). Consider for instance a Lithuanian counterpart of an English *do*-cleft in (64).

(64) Lithuanian (Mikulskas 2017: 120)

Per	*atostogas*	*Jonas*	*beveik*	*nieko*	*neveikė*.
during	holiday.ACC(PL)	Jonas.NOM	almost	nothing.GEN.SG	NEG.do.3.PST
[*Ką*	*jis*	*padarė*],			
what	3SG.NOM.M	PFX.do.3.PST			
tai	*pagaliau*	*sutvarkė*	*savo*	*kelionių*	*nuotraukas*.
that.N	finally	PFX.order.3.PST	RPO	travel.GEN.PL	photograph.ACC.PL

'John did almost nothing during his holiday. What he did do was bring some order in his travel photographs.'

Mikulskas (2016) points out that the expression following the neuter pronoun *tai* in (64) is a finite clause. The author argues that in this case "the pronoun performs the function of segmenting information structure rather than the structural function of a syntactic connector" (p. 120). Note that this is possible also in some varieties of English (e.g. *What he wants is he wants a steaming cup of coffee*). Such constructions are sometimes described as "amalgam pseudo-clefts" (e.g. O'Neill 2015). Unfortunately, it is not clear from Mikulskas' (2016) account whether the equivalent of a *do*-cleft such as that in (64) is possible with a relative clause construction headed by a demonstrative (as in (62) and (63)), and, if so, what the VALUE expression would look like.

Another issue motivating a non-equational analysis for Lithuanian clefts (of the correlative/free relative variety), identified in Mikulskas (2016), concerns the case marking of the clefted constituent. As mentioned earlier, in regular specificational copular sentences the VALUE expression is consistently marked nominative (or instrumental) while in clefts it necessarily takes the case for which the relative pronoun is marked. That is, the case corresponding to the relativized argument (or adjunct). Consider the sentences in (65)-(68) (and the examples in (59)-(60) for genitive and accusative marking on the VALUE expression respectively).

(65) Lithuanian (Mikulskas 2017: 122)

[*Kas*	*mane*	*labiausiai*	*žavi*],	*tai*	*alyvų*	žydėjimas.
what.NOM	1SG.ACC	most	delight.PRS.3	that.N	lilac.GEN.PL	blossoming.NOM

'What delights me most of all is lilac blossoming.'

(66) Lithuanian (Mikulskas 2017: 122)

[*Kam*	*aš*	*esu*	*neabejingas*],	
what.DAT	1SG.NOM	be.PRS.1SG	NEG.indifferent.NOM.SG.M	
tai	*raudunoam*	*vynui*.		
that.N	red.DAT.SG.M	wine.DAT.SG		

'What I'm partial to is red wine.'

(67) Lithuanian (Mikulskas 2017: 122)

[*Kuo*	*aš*	*dar*	*bjauriuosi*],	*tai*	*paskalomis*.
what.INS	1SG.NOM	more	loathe.PRS.1SG	that.N	slander.INS.PL

'What I loathe is slander.'

(68) Lithuanian (Mikulskas 2017: 122)

[*Kur*	*aš*	*dar*	*kartą*	*norėčiau*	*išsimaudyti*],	
where	1SG.NOM	more	once	want.IRR.1SG	sea.LOC.SG	
tai	*Libjijos*	*jūroje*.				
that.N	Libya.GEN	sea.LOC.SG				

'Where I would like to have a swim once more is in the Libyan Sea.'

It could be argued that the verbal copula *būti* 'be' is associated with particular case forms (the nominative and instrumental) and thus, as it were, simply blocks case attraction. Mikulskas (2016) prefers to argue for a "monoclausal" analysis of Lithuanian correlative clefts. In his view, a Lithuanian "pseudo-cleft" would involve an "appositive" relation between two expressions. In the author's terms: "In the Lithuanian pseudocleft […] the sequence [*tai* 'that'+ NP] will have to be analyzed as an appositive of the question word used at the beginning of the free relative clause" (p. 123). The relationship between these expressions would, he argues, not be much different to that holding between *vieno dalyko* 'one thing' and *cepelinu* 'potato dumplings' (both in the genitive, required for object arguments of the verb 'like') in the example in (69) below, which Mikuslkas (2016) describes as a "paraphrase of a pseudocleft" (p. 124):

(69) Lithuanian (Mikulskas 2016: 124)

Nemėgstu	*vieno*	*dalyko —*	*cepelinų.*
NEG.like.1SG.PRS	one.GEN.SG.M	thing.GEN.SG	potato.dumpling.GEN.PL

'I dislike one thing (and that is) potato dumplings.'

Now, the sentence in (69) would not, indeed, fit the definition of clefts adopted in the present investigation. But clefts of the correlative/ free relative variety such as those in (59)-(60) and (65)-(68) do seem to. They appear to involve a participant-oriented nominalization and exhibit a structure very similar to (some variants) of equational sentences in the language. However, the (arguably innovative) pattern involving the copular use of the verb *būti* 'be', which is available in other equational sentences, is not available in these constructions. (Unless the VARIABLE expression is construed following a pro-noun support pattern as in (62) and (63).

4.2.3 Hungarian

A correlative-like pattern for clefts is also reported in Hungarian. I will base the discussion on the account in Hartmann et al. (2013). Before discussing clefts, I will present the account of specificational sentences in Hungarian by Matić (2007).

In his account of specificational sentences in Hungarian, Matić (2007) describes a (simple) specificational sentence as one in which an individual-denoting expression is in focus position and a property-denoting expression is in topic position. The account is based on the following assumption about sentence structure in Hungarian (based on Kiss 1992, 2002; Matić 2007: 26).

(70) Hungarian sentence structure (Matić 2007: 27)

[topic position] [focus position]~[$_{VP}$[Verb][referential phrases]]$_{WIDE\ FOCUS}$

The schematic representation in (70) is to be understood as follows. There are two pre-verbal slots in the core sentence (left dislocation will be discussed below). The first slot in the core sentence is the topic position. This is the position which the VARIABLE expression in a specificational sentence occupies. An (immediately pre-verbal) narrow-focus position follows. The VALUE expression may occupy this position. Alternatively, it may occur in post-verbal position. The pragmatic effects observed in specificational sentences may be observed in verbal sentences as well. The verbal sentences in (71) and (72) illustrate the information structural effects of word order.[74]

(71) Hungarian (Matić 2007: 27)

a. [*Marika*]$_{TOPIC}$ [*ISKOLÁBA*]$_{FOCUS}$ ~ [*mente* *Jánossal*]$_{VP}$
Marika school.in went János.with
'Was Marika betrifft, so ging sie mit János in die SCHULE (nicht auf die Post).'
'As for Marika, she went with János to SCHOOL (not to the post office).'

b. [*Jánossal*]$_{TOPIC}$ [*ISKOLÁBA*]$_{FOCUS}$ ~ [*mente* *Marika*]$_{VP}$
János.with school.in went Marika
'Was János betrifft, so ging Marika mit ihm in die SCHULE (nicht auf die Post).'
'As for János, Marika went with him to SCHOOL (not to the post office).'

[74] Note that the use of square brackets corresponds to that in the scheme in (62). (Square brackets here are not used to indicate the presence of an oriented nominalization).

c. [*Iskolába*]$_{\text{TOPIC}}$ [*JÁNOSSAL*]$_{\text{FOCUS}}$ ~ [*mente* *Marika*]$_{\text{VP}}$
school.in János.with went Marika
'Was die Schule betrifft, so ging Marika dahin mit JÁNOS (nicht mit Feri).'
'As for the school, Marika went there with JÁNOS (not with Feri).'

d. [*Marika*]$_{\text{TOPIC}}$ [*MENTE* *Jánossal*]$_{\text{VP}}$
Marika went János.with
'Was Marika betrifft, so ging sie mit János (sie blieb nicht allein zu Hause).'
'As for Marika, she went with János (she didn't stay home alone).'

Note that the topic position need not be occupied by a referential expression:

(72) Hungarian (Matić 2007: 27)
[*Okosnak*]$_{\text{TOPIC}}$ [*IMRÉT*]$_{\text{FOCUS}}$ ~ [*tartom*]$_{\text{VP}}$
smart.DAT Imre.ACC clever.1SG
'Für klug halte ich IMRE (nicht János).'
'Intelligent I consider IMRE (not János).'

The position an argument (or adjunct) may occupy is in principle indifferent to its role/function, which is consistently marked by case. An important point is that the preverbal topic and focus positions may be occupied by almost any kind of expressions (i.e. not necessarily referential/nominal). Only the post-verbal position, according to Matić (2007: 27), must involve referential expressions.

I will turn now to copular sentences. Hungarian has a verbal copula identical to the existential verb *van*. In its copular use, it is obligatorily absent in the third person singular present tense. The distinction between predicational and specificational sentences in Matić's (2007) account is made as follows: A predicational sentence as described by Matić for Hungarian is one in which an expression denoting an individual occurs in topic position and a property-like expression occurs either in immediately pre-verbal focus position or in post-verbal position. As mentioned earlier, in Matić's account, a specificational sentence is that in which a property-denoting expression occurs in topic position and the individual-denoting expression in focus position or in post verbal position (Matić 2007: 37). Matić's approach is thus compatible with accounts that regard specification as reverse predication (see discussion in 2.1.4 in Chapter 2).[75] An opposition between equation vs. predication plays no role whatsoever in Matić's (2007) account. The question whether a specificational sentence may be best considered a type of equational sentence (or not) is not addressed.

The examples in (73) and (74) illustrate sentences which Matić (2016) describes as predicational. In (73), a nominal predicate occurs in immediately pre-verbal narrow-focus position. In (74), a nominal predicate occurs in post-verbal position.

[75] An important difference between some "inverse predication" accounts and the account in Matić (2007) is that Matić argues against an inversion analysis in terms of subject and predicate for Hungarian. Matić defines the subject in Hungarian as term with which the verb shows person-number agreement. This term may in principle occur in any position (i.e. in the topic, narrow focus, or post-verbal).

(73) Hungarian (Matić 2007: 34)
('Why did the "Black Scorpion" team win on the "Aki II" terrain? Because they were the only team on this terrain?)

Azért	*nyertek,*	*mert*	*ők*	*a*	*legjóbb*	*csapat*	*voltak.*
therefore	won.3PL	because	they	DEF	best	group	was.3PL

'Nein, sie haben deswegen gesiegt, weil sie die BESTE Truppe waren.'
'No, they won because they were the BEST team.'

(74) Hungarian (Matić 2007: 37)

[*Mi*]$_{\text{TOPIC}}$	[*voltunk*	*az*	*első*	*Hitler*	*által*	*megszált*	*ország.*]$_{\text{VP = WIDE FOCUS}}$
We	was.1PL	DEF	first	Hitler	of	occupied	land

'Wir waren das erste von Hitler besetzte Land.'
'We were the first country occupied by Hitler.'

In (73), the expression described by Matić as property-denoting occurs in focus position (immediately preceding the verbal copula) yielding an emphatic/contrastive reading. In (74), this emphasis is lacking because the non-topic expression is in post verbal (wide focus) rather than immediately pre-verbal (focus) position. In both cases, the topic position is occupied by individual-denoting expressions (in the cases at hand, personal pronouns).

The examples in (75) and (76) below illustrate specificational sentences. Again, the examples show the possible positional variation of the non-topical expression. In (75), the VALUE expression occurs in immediately pre-verbal narrow-focus position. In (76), it occurs in post-verbal position.

(75) Hungarian (Matić 2007: 35)
(Context: In an discussion on internet gaming, a participant claims that the "Lions" were the best group. Another participant corrects him. It was the "Hungarian Legion" which got to play against four further teams.)

A	*legjóbb*	*csapat*	*a*	*mezőnyben*	*ők*	*voltak!*
DEF	best	group	DEF	field.in	they	was.3PL

'Die beste Gruppe aud dem Feld waren (doch) SIE.'
'THEY were the best team in the field!.'

(76) Hungarian (Matić 2007: 37)
(Among those infected through a bite, one percent become seriously ill. They develop menigitis…)

[*Ilyen*	*megbetegedés*]$_{\text{TOPIC}}$	[*volt*	*az*	*1999-es*	*New_York-i*	*eset*]$_{\text{VP=WIDE FOCUS}}$
such	disease	was	DEF	1999-ADJ	New_York-ADJ	case

'Eine solche Erkrankung war der New Yorker Fall im Jahre 1999.'
'One such illness was the case in New York in the year 1999.'

In the sentence in (75), a personal pronoun (the VALUE expression) occurs in focus position while a co-denotational definite description containing the VARIABLE (in Matić's terms, a property-denoting expression) occurs in topic position. In the sentence in (76), the expression in post-verbal position provides a VALUE referring to an instance of 'such an illness' (i.e. a case of meningitis induced by a bite). The sentence in (75) involves emphasis/contrast. The one in (76) does not.

In addition to the clause-initial topic position there is the possibility of marking an expression as topical through left dislocation. Left dislocation in Hungarian is characterized by the presence of a resumptive

pronominal form at the beginning of the core clause. Matić (2016) shows that, in cases of left dislocation, the choice of resumptive pronoun with expressions denoting human referents hints at the predicational vs. specificational nature of the sentence in a way similar to that observed in the familiar tag question test in English (see discussion in 2.3.1 in Chapter 2). In a specificational sentence, the choice of the resumptive pronoun is limited to the demonstrative *az* 'that', normally used to refer to non-human referents, even though the VARIABLE-containing expression (property-denoting, in Matić's terms) may refer to a human. The contrast is illustrated in the examples in (77).

(77) Hungarian (Matić 2007: 39)

a. *János, ő/az a legjóbb barátom volt.*
János he/that DEF best friend.POSS was
'Janos, der war mein bester Freund.'
'János, he was my best friend.'

b. *A legjóbb barátom, az János volt.*
DEF best friend.POSS that János was.
'Mein bester freund, das war János.'
'My best friend, that was János.'

c. **A legjóbb barátom, ő János volt.*
DEF best friend.POSS he János was.

In (77)(a), where the proper name *János* is left-dislocated, both a third-person singular personal pronoun and the demonstrative are in principle acceptable. The choice of resumptive pronoun is this case seems not to be very reliable in distinguishing predicational and equational sentences with a left-dislocated subject in Hungarian, as both *ő* and *az* may refer anaphorically to a fully referential expression.[76] As illustrated in (77)(b) and (c), however, in a specificational sentence the personal pronoun is not acceptable, although, as Matić points out, it is perfectly clear that the expression described as property-denoting (*a legjóbb barátom* 'my friend') implies a single, male, human referent (which would normally be compatible with the pronoun *ő* '(s)he').

The same pattern can be observed in constituent questions:

(78) Hungarian (Matić 2007: 41)

a. *Mi volt János?*
what was János?
Ő/?*Az volt a legjóbb Budapesti forgovos.*
he/?that was DEF best Budapestian dentist
'Was war János? (gefragt z.B nach seiner Rückkehr aus Budapest nach Szeged)' — 'Er war der beste Zahnarzt in Budapest.'
'What was János (enquired e.g. after his return from Budapest to Szeged)' — 'He was the best dentist in Budapest.'

[76] Note that Matić does not, as far as I can see, consider in his account the possibility that a proper name may be used a VARIABLE expression. As discussed in 2.3.1 in Chapter 2, such sentences are not considered to be specificational in many accounts but are rather described with terms such as "descriptively identifying" (Declerck 1988). Assuming it may be better to subsume the relations between a VARIABLE expression and a VALUE expression (in the sense I am using the terms) under the label "specificational" regardless of the kind of expression containing a VARIABLE (i.e. including proper names), the sentence in (77)(a) could be both specificational or predicational. (The use of the personal pronoun *ő* 'he' would presumably, however, allow a predicational reading only.)

b. *Ki volt a legjóbb Budapest forgovos?*
who was DEF best Budapestian dentist
*Az/*Ő János volt.*
that/he János was
'Wer war der beste Zahnarzt in Budapest? — 'Das war János.'
'Who was the best dentist in Budapest?' — 'That was János.'

In the example in (78)(a), the resumptive pronoun may be said to refer to the best dentist (a human) but the use of the personal pronoun *ő* is not appropriate. This in Matić's view allows the conclusion that „the denotatum […] is not an individual with the feature [+ human] but some extensionally distinct category, namely the property of being the best friend" (p. 40).[77] I will turn now to the account of clefts in Hungarian by Hartmann et al. (2013).

Hartmann et al. (2013) describe clefts in Hungarian as consisting of a (left dislocated) free relative clause followed by a demonstrative pronoun, a nominal expression, and a copula. Consider the following examples.

(79) Hungarian (Hartmann et al. 2013: 79)
[*Aki a gitáron játszik*], *az* *ÉN* *vagyok*
who the guitar.on plays that(DEM) I am
'Who is playing the guitar is me.'

(80) Hungarian (Hartmann et al. 2013: 79)
[*Ami-t János elolvasott*] *az* *A HAMLET(*-ET)* *volt*
what-ACC JOHN PTCL.read that(DEM) the Hamlet(-ACC) was
'What John read was Hamlet.'

The sentences in (79) and (80) illustrate clefts formed with a free relative in left-dislocated position. The free relatives are introduced by a relative pronoun (with the shape of an interrogative pronoun preceded by the vowel *a*) which is inflected for the case corresponding to the role/function of the relativized argument. The left-dislocated expression is followed by a resumptive pronoun, itself followed by the clefted constituent (which is necessarily in the nominative). The copula verb would not present in third person present declarative sentences.

Note that the clefted constituent can co-occur with another expression in immediately pre-verbal (pre-copular) focus position, as illustrated in (81).

(81) Hungarian (Hartmann et al. 2013: 77)
[*Aki-t a diákok legjobban szerettek*], *az* *TAVALY* *volt* *JÁNOS*
who-ACC the students best liked.3PL that(DEM) last.year was John
(*és* *TAVALYELŐTT* *volt* *PÉTER*)
and the.year.before was Peter
'It was last year that the one who the students liked best was John (and it was in the year before that it was Peter).'

[77] „[…] was den Schluss zulässt, dass das Denotatum der Phrase kein Individuum mit dem Merkmal [+menschlich] ist, sondern eine extensional unterschiedliche Kategorie, nämlich die Eigenschaft, der beste Freund zu sein" (Matić 2016: 40).

A specificational relation holds between the free relative and the (post-copular) VALUE expression *János*. But the pre-copular (narrow focus) position is occupied by a temporal adverbial.

Note that the pattern involving a left-dislocated free relative illustrated for clefts in the examples (79)-(81) is in principle available for both (specificational) clefts and (non-equational) predicational sentences with a free relative as a subject. The sentence in (82)(a) is a predicational sentence, the one in (b) a cleft. Hartmann et al. (2013) call the former "predicational pseudo-cleft" and the latter "specificational pseudo-cleft".

(82) Hungarian (Hartmann et al. 2013: 68)

a. [*Ami-t* *János* *nézett*], *az* *szép* *volt*.
what-ACC John looked.3SG that(DEM) beautiful was
'What John was looking at was beautiful.'

b. [*Ami-t* *János* *nézett*], *az* *A MONA_LISA* *volt*.
what-ACC John looked.3SG that(DEM) the Mona_Lisa was
'What John was looking at was Mona Lisa.'

In both examples (82)(a) and (b), the free relative denotes an object that was being looked at but in the predicational sentence the identity of the referent (whatever it was) is not in question. A property is simply attributed to it. In the specificational sentence in (82)(b), the point is to specify what the thing looked at was (here, a particular work of art).

The difference between (specificational) clefts and predicational sentences with a left-dislocated free relative is reflected in the use of resumptive pronouns in a similar way as the difference between simple specificational and predicational sentences discussed earlier and illustrated in examples (77) and (78). (As long as human referents are involved; apparently the personal pronoun *ő* '(s)he' is used to refer to human referents only.) The contrast between a cleft (83)(a) and a predicational sentence with a left dislocated free relative (83)(b) is illustrated below.

(83) Hungarian (Hartmann et al. 2013: 82)

a. [*Aki* *jövőre* *megnyeri* *az* *X-faktort*], ??ő/az szerintem
who next.year PARTICLE.wins the X-factor.ACC he/that in.my.opinion
a szomszéd nagyobbik fia *lesz*
the neighbor older son.GEN will.be
'Who will win X Factor next year will be the older son of our neighbor in my opinion.'

b. [*Aki* *tavaly* *megnyerte* *az* *X-faktort*], *ő/az* *szerintem*
who last.year PARTICLE.won the X-factor.ACC he/that in.my.opinion
valóban *tehetséges* *volt*.
indeed talented was
'Who won X Factor last year was really talented in my opinion.'

Hartmann et al. (2013) point out that there seem to be speakers who prefer the demonstrative in any context (i.e. whether the sentence is predicational or specificational). However, speakers who find the use of the personal pronoun in predicational sentences fully appropriate reject it in specificational sentences (p 82). Interestingly, the authors note that in predicational sentences (with a left-dislocated free relative) the presence of the demonstrative *az* (which would indicate left dislocation) is "more or less optional, which means that the *wh*-clause can either be left-dislocated or occupy the topic position" (Hartmann et al. 2013: 82). In (specificational) clefts, however, left dislocation is reported to be very strongly preferred. Consider the sentences in (84).

(84) Hungarian (Hartmann et al. 2013: 83)

a. *Az-t* *hittem* *hogy* [*amelyik* *színészt* *legjobban* *kedveled*],
that-ACC believed.1SG that which actor.ACC best liked.2SG[78]
?(*az*) *amerikai*.
that(DEM) American
'I thought that the actor you liked best is American.

b. *Az-t* *hittem* *hogy* [*amelyik* *színészt* *a* *legjobban* *kedveled*],
that-ACC believed.1SG that which actor.ACC the best like.2SG
?*(?*az*)[79] *Schwarzenegger*.
that(DEM) Schwarzenegger
'I thought that the actor you like best is Schwarzenegger.

Hartmann et al. (2013) point out that the assessment of the left-dislocated status of a both predicational and specificational sentences with free relatives such as (84)(a) and (b) is not quite straightforward. On the basis of an acceptability rating task involving 17 participants, the authors conclude that there does seems to be a clear difference between specificational and predicational sentences with respect to the acceptability of the presence or absence of the demonstrative *az.* The authors note, however, that the issue is "complicated by the fact that the presence or absence of the copula [not used in the present tense] seems to play a role in the acceptability of the presence vs. absence of the pronominal *az"* (p. 83).

Though not discussed in Hartmann et al. (2013), it seems that Hungarian cleft constructions permit, beside a VARIABLE-VALUE order such as that illustrated in the examples discussed so far, a VALUE-VARIABLE configuration as well. Consider the example in (85).

(85) Hungarian (Coyaud 1975:45)
SZABO-UR *az* [*aki* *hust* *evett* *tegnap*]
Szabo-Mr that who meat ate yesterday
'It was Mr Szabo who ate meat yesterday.'

It is not clear from Coyaud's (1975) account, unfortunately, how the sentence should be analyzed in terms of the sentence schemas discussed earlier. Assuming the demonstrative *az* is used in (85) as a resumptive element, the VALUE expression *Szabu-ur* 'Mr Szabo' may be analyzed as a left-dislocated nominal expression and the VARIABLE expression as occurring in the sentence-final slot reserved, following Matić's (2007) account, for referential expressions. (If this assumption is correct, in a variant of the sentence in (85) featuring an overt verbal copula, the copula would precede the VARIABLE expression.)

4.2.4 Nahuatl

Canonical word order in Classical Nahuatl is described as predicate initial (Launey 1986). The predicate word (which may be any content word[80]) may be preceded by particles forming together with it a predicate complex. In canonical sentences, the predicate complex is followed by argument and adjunct expressions, which may (but need not) be introduced by the article *in* or a demonstrative. Argument and adjunct expressions introduced by an article or by a demonstrative may also precede the predicate complex. Launey

[78] Note that the glossing (and the corresponding translation) of the form *kedveled* 'like, present definite tense' differs in examples (a) and (b) in the source.
[79] It is not quite clear to me what exactly the question mark within the parentheses indicates.
[80] Launey (1986) proposes three major word classes in the language: verbs, nouns, and locatives. All of these may function as predicates.

(1986) describes these constructions as topicalizations. Topicalization plays an important role in clefting in the language. I will begin the discussion by illustrating first the canonical sentence pattern. Consider the sentence in (86).

(86) Classical Nahuatl (Sahagún vol. I: 71; cited in Launey 1986: 1492)

Ō-quim-īxpopoyōtilî	*in*	*diablo*	*in*	*huēhuetquê.*
ANT-3PL.OBJ-render.blind.PRF	ART	devil	ART	old.PL

'Le diable a aveuglé les enciens.'
'The devil blinded the old ones.'

In (86), the verb occurs in sentence-initial position and is followed by two nominal expressions corresponding to the subject and the object. The presence of these is facultative and the order is variable, though Launey (1986: 1492) reports a tendency for the subject to precede the object and arguments to precede adjuncts. As mentioned in Section 3.2.6 in Chapter 3, clausal nominalizations in the language are introduced by the article *in* and exhibit in principle the same structure as independent clauses. Consider the sentence in (87), where the second argument expression *in quittaquê* 'what they had seen' is an oriented clausal nominalization.

(87) Classical Nahuatl (Sahagún vol XII: 17; cited in Launey 1996: 1492)

Qu-ilhuīcô	*in*	*Motēuczōma*	[*in*	*qu-ittaquê*]
3SG.OBJ-tell.AND.COMPL.PL	ART	Moctezuma	ART	3SG.OBJ-see.PRF.PL

'Ils allèrent dire à Moctezuma ce qu'ils avaient vu.'
'They went to tell Moctezuma what they had seen.'

The verbal form *quilhuīcô* 'they went to tell him/her' is in principle a full clause and could by itself form a complete independent sentence. The form *quittaquê* 'they have seen it/him/her' is marked as a subordinate clause by means of the article preceding it. Note that the expression *in quittaquê* is not inherently oriented. In the case at hand, it is interpreted as oriented toward the object ('what they had seen') but it could also denote the subject ('the ones who had seen it/him/her'). In the appropriate context, it could be interpreted as a complement clause or as an adverbial clause ('that they have seen it/him/her' or 'when they saw it/him/her'). A clause (or a predicate word) introduced by the article *in* is necessarily subordinate, however. It cannot be interpreted as the main predicate of a sentence.

As mentioned at the beginning of this section, the main predicate in a sentence may be a noun, possibly preceded by a particle and (canonically) followed by a subject expression introduced by the article *in.* Expressions introduced by the article, however, may not occur in predicate position. That is, the juxtaposition of two expressions introduced by an article or a demonstrative does not yield a clause in the language. To construe an equational sentence involving two article-marked expressions, one of the expressions is topicalized and a free pronoun (not introduced by the article) occurs in predicate position. Consider the specificational sentence in (88).

(88) Classical Nahuatl (Sahagún vol XI: 9; cited in Launey 1986: 632)

auh	*in*	*ī-tlacual*	*yèhuātl*	*in*	*quimichin*
and	ART	3SG.POSS-food	3SG	ART	mouse

'Et sa nourriture, ce sont (uniquement) les ("la") souris.
'And the mouse is its (only) food.' ('And its food, it's the mouse.')

In (88), a nominal expression (in this case, involving a possessed noun preceded by the article *in*) is topicalized. The third person singular pronoun *yèhuātl* (short form *yê*) is followed by a nominal expression

introduced by the particle *in*. The topicalized expression may be described as the VARIABLE and the sentence-final one as the VALUE. Neither of these, however, may be regarded as the predicate (in formal terms). Consider now the sentences in (89).

(89) Classical Nahuatl (Sahagún vol XII: 34; cited in Launey 1986: 631)
ca *amo* *zan* *aca*
ASRT NEG only somebody
'It was not just anybody (who accosted us).'

ca *yèhuātl* *in* *tēlpōchtli* *Tezcatlipōca*
ASRT 3SG ART youngster Tezcatlipoca
'It was the youth Tezcatlipoca.'

In the second of the two consecutive sentences in (89) the third person pronoun *yèhuātl* may be said to take up a referent from the preceding context. Note that the free pronoun is preceded by the assertive particle *ca* and followed by a (subject) expression introduced by the article *in.* In both (88) and (89), then, the free pronoun functions as the main predicate in the sentence. Or, to put it in other words, the pronoun occupies the main-predicate slot in the sentence.

Now, expressions not introduced by an article (or a demonstrative) may freely occur in predicate position, as in (90) (see also examples in 3.2.6 in Chapter 3).

(90) Classical Nahuatl (Sahagún vol. VI: 57; cited in Launey 1986: 634)
cuix *nanacatl* [*in* *ti-qu-itta*] [*in* *ti-c-mati*]
DUB mushroom ART 1PL.SBJ-3SG.OBJ-see ART 1PL.SBJ-3SG.OBJ-feel
'Do we perhaps see, [do we] perhaps perceive as if crazed […]?'
('Perhaps what we see and what we feel are mushroom-induced visions?')

In (90), the coordinated expressions in square brackets following the noun *nanacatl* 'mushroom' in predicate position (preceded by the dubitative particle *cuix*) are participant-oriented nominalizations. It is not clear whether sentences such as the one illustrated in (90) must necessarily be regarded as involving a predicational (as opposed to specificational) relation between the nominal expression in predicate position and the nominal expressions that follow (see discussion in 2.3.2 in Chapter 2 on a similar problem in other languages). If so, such sentences are not to be considered clefts, at least not following the definition adopted in this investigation. In the case of (90), this is possibly so. In many cases, it is difficult to decide for the predicational or predicational status of a sentence. As argued in Chapter 2 (Section 2.3.2), it is difficult to draw a clear distinction between equational sentences and nominal-predicate sentences on the basis of formal criteria alone. Consider the sentences in (20) and (21) (discussed previously in Section 3.2.6 in Chapter 3 and repeated here for convenience)

(91) Classical Nahuatl (Sahagún vol. II: 122; cited in Launey 1986: 1315)
zan *nō* *cihuâ* [*in* *tla-napaloâ*]
only also woman.PL ART NSP.OBJ-carry.in.arms.PL
'Also only the women carried (litters) on their shoulders.'
'The ones who carried (litters) on their shoulders were also only the women.'

(92) Clasical Nahuatl (Sahgún vol. I: 39; cited in Launey 1986: 1316)
Tzapotēcâ [*in huel īn-teōuh catca*]
Zapotec.PL ART very 3PL.POSS-god be.LOC.PRF
'He was the proper god of the Zapotec'
'The ones whose proper god he was were the Zapotec.'

It is not clear whether the nominal expressions in predicate position are (necessarily) to be interpreted as non-referential (note that referential expressions in the language do not necessarily require a determiner). In any case, in sentences where the VALUE expression is introduced by a determiner (as in (88) and (89) above), the use of a free pronoun in the predicate slot is required. The examples in (93) and (94) illustrate clefts with article-marked VALUE expressions and a personal pronoun in predicate position.

(93) Classical Nahuatl (Sahagún vol. IX: 7; cited in Launey 1986: 1536)
auh in yèhuāntin tlàtòquê, [*in qui-mo-tlāuhtiāya*], *yèhuātl in tlàtocātilmàtli*
and ART 3PL lord.PL ART 3SG.OBJ-RFL-offer.PST.PL 3SG ART lord.cape
'And these rulers assumed as favours the rulers' capes.'
('And these rulers, what they received was the ruler's capes.'

(94) Classical Nahuatl (Olmos (1875[1547]: 236; cited in Launey 1986: 632)[81]
aocmo yèhuātl in òtli in xopechtli [*in qui-toca*]
NEG 3SG ART path ART foundation ART 3SG.OBJ-follow
'It's no longer the (good) way, the (good) example that he follows.'
('What he follows is no longer the path and the foundation.')

As shown in the examples in (93) and (94), word order in this kind of constructions is flexible. Note that the participant-oriented nominalization may occur preceding the free pronoun in predicate position (93) or follow the nominal expressions in canonical argument position as in (94). In the first case, the VARIABLE expression can be described as topicalized.[82]

---

[81] Note that in the source text the sentence cited by Launey (1986: 1536) is embedded in an oriented nominalization:
(i) Classical Nahuatl (Olmos 1875[1547]: 236)
in aucmo yehuatl yn utli, yn xopechtli in qui-toca
ART NEG 3SG ART way ART foundation ART 3SG.OBJ-folow
'the one who what he follows is no longer the path and the foundation'

[82] It should be pointed out that, beside constructions such as the ones presented in (93) and (94), which can be described as clefts, other constructions pragmatically related to clefts involving free pronouns in sentence-initial position occur in the language. Consider the following sentence:

(i) Classical Nahuatl (Carochi 1892[1645]: 483; cited in Launey 1986: 640)
ca tèhuātl ti-yāz
ASRT 2SG 2SG.SBJ-go.FUT
'It is you who will go.' ('YOU will go')

In the sentence in (i), a free pronoun occurs immediately following a (necessarily clause-initial) assertive particle but the verbal expression following it is not marked as a nominalization (i.e. it is not preceded by an article). It was mentioned earlier that argument (and adjunct) expressions are not always preceded by a determiner. Participant-oriented nominalizations (and clausal nominalizations generally) are consistently introduced by the article *in*. The sentence in (i) is not, arguably, a cleft. It may be perhaps better analyzed as a verbal sentence with an argument in pre-verbal position following a (non-canonical) pattern in which arguments and adjuncts (not introduced by a determiner) may precede a verbal predicate. This pattern is described in Launey (1986: 1500) as a "compact construction". The use of the free pronoun in a context such as (i) may trigger pragmatic effects similar to those often associated with clefts (e.g. an exhaustivity implicature or contrast) but this may arguably be explained as an effect triggered by the

A very similar pattern to that discussed for Classical Nahuatl is reported in Sasaki (2014) to occur in Ixquihuacan Nahuatl. As in Classical Nahuatl, nouns in Ixquihuacan Nahuatl can freely function as predicates. In the present tense, no copula is used. The examples in (95) illustrate predicational sentences involving a proper noun (introduced by the article *n*) and a bare noun denoting an occupation.

(95) Ixquihuacan Nahuatl (Sasaki 2014: 2)

a. *n* *Elias* *tlapahtāni*
ART Elias doctor
'Elias is a doctor.'

b. *tlamachtāni* *n* *Elias*
teacher ART Elias
'Elias is a teacher.

Note that word order is flexible. Sasaki (2014) does not provide information regarding the conditions under which either of the possible word-order patterns illustrated in (95) may be preferred. There is no indication in his account, however, that the pre-predicate occurrence of the subject expression (in the cases at hand, the expression introduced by the article) may be a pragmatically marked option (as it may be, according to Launey (1986), in classical Nahuatl). The examples are intended simply to illustrate that a nominal-predicate sentence does not require (in the present tense) the use of a copula. Juxtaposition of nominals, however, is only possible when the nominal expression in predicative function is not marked by the article *n*, which Sasaki (2014: 3) describes as a definite article. According to the author, a sentence such as (96) is unacceptable.

(96) Ixquihuacan Nahuatl (Sasaki 2014: 3)
**n* *Elias* *n* *tlapahtāni*
ART Elias ART doctor
Intended: 'Elias is the doctor.'

In sentences such as the one in (96), a personal pronoun is used following a pattern reminiscent of that observed in Classical Nahuatl is used. In the first person singular, the third person pronoun *yeh* coalesces with the article *n* resulting in the form *yēn*. Consider the examples: in (97), illustrating to possible answers to a question inquiring the identity of the mayor of a town.

(97) Ixquihuacan Nahuatl (Sasaki 2014: 2)
(Q: Who is the mayor?)

a. *n* *to-tiāchkah* *deh* *nikã* *to-puēbloh* *yēn* (<yeh n) *Victor*
ART 1PL.POSS-chief of here 1PL.POSS-village COP 3SG ART Victor
'The chief of our village is Victor.'

b. *Yēn* (<yeh n) *Victor* *n* *to-tiāchkah* *deh* *nikã* *to-puēbloh*
COP 3SG ART Victor ART 1PL.POSS-chief of here 1PL.POSS-village
'Victor is the chief of our village.'

The sentences in (97)(a) and (b) are presented as alternatives which can be felicitously answer the question. Sasaki (2014) shows another, alternative, word order possibility, illustrated in (98).

---

very occurrence of a free pronoun. It may well be the case that the type of constructions illustrated in (i) is more frequent than clefts in the language.

(98) Ixquihaucan Nahuatl (Sasaki 2014: 6)

n	*Victor*	*yēn* (<yeh	n)	*to-tiāchkah*	*deh*	*nikã*	*to-puēbloh*
ART	Victor	COP	3SG	ART 1PL.POSS-chief	of	here	1PL.POSS-village

'Victor is the mayor of this village' (used to introduce Victor to others)

In this case, where the form *yēn* does not immediately precede the proper name Victor but the definite description 'the mayor of this village', the VARIABLE-VALUE relations are inverted. Now the personal name is the VARIABLE expression and the definite description the VALUE.

Note that the third person free pronoun is sensitive to number, as illustrated in (99).

(99) Ixquihuacan Nahuatl (Sasaki 2014: 2)
(Q: Who made the side dish?)

yehwā	*n*	*Juan*	*īwā*	*Maria*	*n*	*ō-ki-chih-chih-keh*
3PL	ART	Juan	and	Maria	ART	PST-3SG.OBJ-RED-make-PL

'It was Juan and Maria who made it'

This is a cleft comparable to the Classical Nahuatl example presented earlier in (94). The verbal form introduced by the article in sentence final position is a participant-oriented clausal nominalization, which is formed by a verbal predicate which could itself form a complete sentence (*ōkichihkeh* 'they made it') and is marked as a nominal expression by being introduced by the article *n*.

Sasaki (2014) points out that the third-person pronoun is used in a copula-like manner with definite (article marked) nominal expressions (including proper names), as well as demonstratives and pronouns. The sentences in (100) illustrate the pattern with demonstratives:

(100) Ixquihuacan Nahuatl (Sasaki 2014: 2)

a.

[*n*	*tlēn*	*ō-ni-k-tēmowā-ya*]	*yeh*	*nikānkah*
ART	REL	PST-1SG.SBJ-3SG.OBJ-look.for-IPFV	3SG	this

'This is what I was looking for.'

b.

*[*n*	*tlēn*	*ō-ni-k-tēmowā-ya*]	*nikānkah*
ART	REL	PST-1SG.SBJ-3SG.OBJ-look.for-IPFV	this

'This is what I was looking for.'

c.

ō-ni-k-tēmowā-ya	*nikānkah*
PST-1SG.SBJ-3SG.OBJ-look.for-IPFV	this

'I was looking for this.'

The sentence in (100)(a) illustrates a cleft sentence where the VALUE expression is the proximal demonstrative *nikānkah*. Juxtaposition of a participant-oriented nominalization and a demonstrative, as illustrated in (100)(b) is not acceptable. A corresponding canonical verbal sentence is shown in (100)(c).

As mentioned above, Sasaki (2014) reports that a third person pronoun is also used in a copula-like manner where the VALUE expression is a personal pronoun:

(101) Ixquihuacan Nahuatl (Sasaki 2014: 2)

wan	*yeh*	*nehwātl*	[*n*	*ō-ni-k-chih-chih*]
and	3SG	1SG	ART	PST-3SG.SBJ-3SG.OBJ-RED-create.PST

'I made it.' ('I was the one who made it')

This use of the third person pronoun is reminiscent of the copular use of pronominal forms in languages such as Arabic and Modern Hebrew (see discussion in 2.3.1.2 in Chapter 2). It is surprising if we assume that the use of pronominal forms in Nahuatl may be best understood as involving the substitution of a determiner-marked (or, inherently non-predicative) expression by a co-referent personal pronoun used, as it were, as a pro-predicate. Apparently, in Ixquihacan Nahuatl only a third person pronoun (singular or plural) may assume this function. Other personal pronouns may, it seems, only occur as subjects in nominal-predicate sentences such as (102).

(102) Ixquihuacan Nahuatl (Sasaki 2014: 2)
tehwātl *ti-no-siwa-h*
2SG 2SG.SBJ-1SG.POSS-woman-SG.POSS
'You are my wife'

Sasaki (2014) points out that the construction involving the use of a third person pronoun in a copula-like manner seems not to be available if the VALUE is not a definite (marked) expression. Consider example (103).

(103) Ixquihuacan Nahuatl (Sasaki 2014: 8)

a. **yēh* (*yēyi*) *tōtolti-meh* *n* *ō-sē-kin-koh* *n* *āxā*
3SG three egg-PL ART PST-1PL.SBJ-3PL.OBJ-buy.PST ART today
Intended: 'What we bought today was (three) eggs.'

b. *ō-sē-kin-koh* (*yēyi*) *tōtolti-meh* *n* *āxā*
PST-1PL.SBJ-3PL.OBJ-buy.PST three egg-PL ART today
'We bought (three) eggs today.'

A sentence such as (103)(a) is reported not to be acceptable. Interestingly, Sasaki (2014) points out that in cases such as illustrated in (103), canonical sentences must be used "regardless of information structure" (p. 8). It is not clear from Sasaki's (2014) account whether a sentence involving a participant-oriented nominalization and a bare nominal (without the third person pronoun used in a copula-like manner) is possible in the language at all. Perhaps it is, but the only possible interpretation would be a predicational one, as is apparently the case in sentences involving (bare) nominal predicate such as those illustrated in (95).

4.3 Summary

This Chapter discussed word-order in clefts and specificational sentences generally. In Section 4.1, I presented five case studies illustrating cross-linguistic variation in the patterns involved in the alternation between VARIABLE-VALUE and VALUE-VARIABLE configurations. In 4.1.1, I discussed word order in specificational sentences and clefts in Mandarin. The structure of both specificational and nominal-predicate sentences in this language involves, as in the case of English, two expressions flanking a copula (though the use of the copula is much less regular than it is in English). It seems that in Mandarin only subject-copula-predicate and VARIABLE-copula-VALUE configurations are allowed (excluding, perhaps, constructions involving dislocation). In Akan/Twi, discussed in Section 4.1.2, both specificational and nominal-predicate sentences involve a copula occurring between two expressions. As had been mentioned in 2.2.2.2 in Chapter 2, in this language specificational sentences and nominal-predicate sentences are formally distinct, featuring two different copulas. Word-order order flexibility is possible in specificational, but not in nominal-predicate sentences. Note that in Mandarin there is a strong correspondence between the subject in a

nominal-predicate sentence and the VARIABLE expression in a specificational sentence (and between the nominal predicate and the VALUE expression). Akan/Twi, by contrast, exhibits a clear dissociation between subject/object and VARIABLE/VALUE relations. Unlike in Mandarin and Akan/Twi, in Hausa (discussed in Section 4.1.3) specificational and many (but not all) nominal-predicate sentences do not (in the case of main declarative clauses at least) follow a pattern involving a copula flanked by subject and nominal predicate (or VARIABLE and VALUE expressions). Rather, a morpheme of presumably deictic origin immediately follows the predicate (or VALUE) expression forming a tight syntactic unit. Word order is reversible. In all of the languages discussed in Sections 4.1.1 to 4.1.3, the basic word order in verbal sentences is reported to follow an SVO pattern. Sections 4.1.4 and 4.1.5 dealt with word order in clefts (and specificational sentences generally) in two genetically and areally unrelated languages exhibiting a basic SOV word order in verbal sentences: Korean and Amharic. In both languages, specificational and nominal-predicate sentences involve verbal copulas in clause final position. The basic pattern in specificational sentences in both languages seems to involve a VARIABLE-VALUE configuration, with the VALUE expression immediately preceding the copula. In these constructions, the VARIABLE expression occupies the canonical subject position and the VALUE expression occupies the position corresponding to that of a nominal predicate. In both of these languages, VALUE-VARIABLE order is possible as well, but the patterns involved in these alternative constructions differ. In the case of Korean, the VALUE-VARIABLE configuration involves the placement of the VALUE expression in canonical (sentence initial) subject position but entails morphological (and prosodic) restrictions with respect to the coding possibilities available for subject expressions in nominal-predicate sentences. In Amharic, by contrast, the VALUE expression in specificational sentences with VALUE-VARIABLE word order remains in immediately pre-copular (predicate) position, and the VARIABLE is placed in (possibly extra-clausal) post-verbal position.[83]

In the second part of this chapter (Section 4.2), I examined data from languages in which clefts (and, to some extent specificational sentences generally) require the dislocation of the VARIABLE expression and its resumption in a core clause by means of a pronominal form. In 4.2.1, I examined relative-correlative sentences in Hindi and discussed the problem concerning the nominal status of the relative clause in this type of constructions. In relative-correlative sentences, the relative part is clearly marked as a dependent clause but its status as a clausal nominalization is problematic as it does not exhibit the distributional potential of a nominal expression. This renders the status of the constructions in question problematic with respect to the definition of clefts adopted in this dissertation. A (pro-)noun-anchored variant of the relative-correlative construction exists in the language, in which the relative part may perhaps be more easily characterized as a nominalization. It seems to be a relatively recent, possibly contact-induced, innovation in the language. Lithuanian (discussed in 4.2.2) exhibits a pattern reminiscent of the Hindi relative-correlative construction in cleft sentences. The strategy seems to be obligatory except in cases in which the VARIABLE expression involves (pro-)noun anchoring (i.e. "headed" or "light-headed" relative clause constructions). In these cases, a further strategy is available. The alternative strategies correspond to two patterns available in the language for the formation of nominal (i.e. nominal-predicate and equational) sentences. The first strategy involves the dislocation of a nominal expression and its resumption by a demonstrative in a copula-less nominal clause. The second pattern involves the use of a (originally locative/stative) verbal copula, the use of which is reported to be incompatible with the dislocated, correlative-like cleft construction. In 4.2.3, I discussed clefts (and specificational sentences generally) in Hungarian, another language in which the

[83] Note that the strategy placing the VARIABLE expression in post-copular position described for Amharic is available, and not uncommon in Japanese (Yuto Yamazaki p.c.). It is presumably available in Korean as well. Unfortunately, I am not aware of accounts in the literature on cleft constructions discussing such constructions in either language.

dislocation of the VARIABLE expression is reported to occur in clefts (but not necessarily in specificational sentences generally). In Section 4.2.4, I examined data from Classical Nahuatl and Ixquihuacan Nahuatl. In both languages, nominal-predicate sentences involve the juxtaposition of a nominal predicate and a referential subject expression (canonically, in that order, at least in the case of Classical Nahuatl). Specific-definite referential expressions are generally introduced either by an article (or article-like element) or a demonstrative. There is a restriction in these languages such that two expressions introduced by either of these elements in juxtaposition cannot form a sentence. In equational sentences involving two definite-marked nominal expressions, a free pronoun occurs in predicate position. In specificational sentences, the free pronoun anaphorically resumes the dislocated VARIABLE expression (in clefts, an oriented clausal nominalization). Formally, the VALUE expression aligns with the subject of a nominal-predicate sentence and the resumptive pronoun behaves, as it were, as a pro-predicate. In Ixquihuacan Nahuatl, this kind of constructions seems to have developed further and the resumptive pronoun is reported to behave in many ways as an equational copula.

5 (Non-pseudo-) clefts

The present chapter is dedicated to the discussion of constructions of the type for which the term 'cleft' was originally coined (in Jespersen 1969[1937]). Sentences of this type (such as the English *it*-cleft or the French *c'est…qu…* cleft) pose a problem to the approach to clefting adopted in this dissertation. Under the cross-linguistically oriented approach to clefting I am following, a cleft sentence is simply a specificational sentence in which the VARIABLE expression involves an oriented clausal nominalization. As discussed in 1.1.2 (Chapter 1), however, constructions in different languages often described as "clefts" (in contrast to "pseudo-clefts") present a number of idiosyncrasies such that an analysis as specificational sentences is difficult (or at least not straightforward). The chapter is organized as follows: In Section 5.1, I will (using English data) recapitulate the problems constructions such as the English *it*-cleft (and similar constructions in other languages) pose to an analysis as clefts with respect to the definition adopted in this dissertation. In Section 5.2 I will discuss two main patterns occurring in (major) modern Germanic and Romance languages (in the latter group, with the exception of Romanian; see e.g. Gheorge 2017). The patterns to be discussed share the trait of involving a cleft clause with the form of a modifying (as opposed to free) relative clause, often (but not necessarily) formally indistinguishable from a complement clause. The two patterns are distinguished by the presence (or absence) of a cleft pronoun. Sections 5.3 and 5.4 are dedicated to the discussion of the English *it*-cleft and the French *c'est…qu-…* cleft respectively. Here, I will examine two different approaches which seek to account for the idiosyncratic properties of these constructions. The first, which I will refer to as the "discontinuous-constituent" analysis, views the cleft pronoun and the cleft clause as a discontinuous constituent (i.e. a "light-headed" relative clause construction with the cleft pronoun as a "head" and the cleft clause as an extraposed restrictive relative clause). Under the second analysis, which I will refer to as the "extraposition" analysis) the cleft clause is viewed as a (formally anomalous) free relative clause in sentence-final position cataphorically linked with the cleft pronoun (in languages in which such an element occurs). In Section 5.5 I will discuss data from languages outside of Germanic and Romance in which constructions reminiscent of the patterns discussed in 5.2 are reported to occur. The chapter concludes with a summary in 5.6.

5.1 Problems

In this section I will discuss in some detail what it means for *it*-clefts (and analogous constructions in other languages) to be problematic as clefts. I will discuss the problems using English data but most of what I will have to say about English *it*-clefts applies to other constructions in Germanic and Romance as well. In the examples I will continue to enclose clausal nominalizations in square brackets, and, in clefts, I will underline the VALUE expression (i.e. the clefted constituent). Note, however, that I also use square brackets to mark cleft clauses in *it*-clefts although these generally do not have the potential distribution of nominal expressions. In other words, square brackets may be used to mark either a clausal nominalization or a cleft clause. Note that I may also use square brackets to mark clausal nominalizations in non-cleft sentences. The following example is meant only to remind the reader of the terminology used.

(1) English (constructed)

It	*is*	*clefts*	[(*that*) *this example is about*].
CLEFT PRONOUN	COPULA	CLEFTED CONSTITUENT	CLEFT CLAUSE

I will now discuss the problems the *it*-cleft may pose with respect to my definition of clefts, i.e. the problems concerning the analysis of an *it*-cleft as a specificational sentence in which the VARIABLE expression is an oriented clausal nominalization.

5.1.1 The form of the cleft clause

The most problematic characteristic of *it*-clefts is that the cleft clause does not have the form of a nominal expression. It is typically identical to a (restrictive) adnominal relative clause (which, without a "head" noun, may not function as a nominal expression). In Chapter 3, I mentioned that, in many languages, the distinction is not necessarily relevant. In some languages, however, it is. In Germanic and in Romance, some oriented subordinate clauses cannot be used as nominal expressions by themselves. Consider the following examples. Note that a subordinate clause with the same shape as the cleft clause in (2) cannot function as an argument in a verbal sentence (3).

(2) English (constructed)
It was <u>my cat</u> [*that died*].

(3) English (constructed)
**I buried* [*that died*].

Compare now the sentences in (2) and (3) with those in (4). Assume that the reading of (4)(a) and (b) is in both cases specificational and that the expression in square brackets is the VARIABLE expression. (Imagine the sentences are used e.g. to answer the question *Which one* (*of the cats*) *died?*)

(4) English (constructed)
a. [*The one that died*] *was <u>my cat</u>*.

b. *<u>My cat</u> was* [*the one that died*].

Notice now that an expression with the same shape as the expression in brackets in the (pseudo-)cleft sentence in (4), unlike the one featured in (2), may be used as an argument in a verbal sentence. This is illustrated in (5) (to be compared with example (3)).

(5) English (constructed)
I buried [*the one that died*].

As I will further discuss below, cleft sentences introduced by *that* and those without any introducing element (usually only possible with relativized non-subjects in English) are particularly problematic because even their oriented (i.e. relative-clause) status becomes problematic. In other cases, the oriented nature of the cleft clause is not in question (i.e. it can be argued to be clearly oriented). Still, its status as a nominal expression may be problematic. Consider the sentences in (6) and (7):

(6) English (www)
It was <u>John</u> [*who died*].

(7) English (constructed)
?*I buried* [*who died*].

The cleft clause in (6) is arguably oriented but still does not have the shape of a regular nominal expression. Native speakers of English are reported to consider the sentence in (7) unusual if not unacceptable (see e.g. Collins 1991: 29). Though clearly oriented, a clause introduced by the relative pronoun *who* cannot normally

function as a nominal expression unless associated with a noun (8), a pro-noun (9) or some pronominal form (10)-(11).

(8) English (www)
Son, [*the man who left*] *died so long ago.*

(9) English (www)
[*The one who left*] *died in battle*.

(10) English (www)
[*Those who left*] *died of starvation in strange places.*

(11) English (www)
[*He who returns from a journey*] *is not the same as* [*he who left*].

Note, however, that the restriction is not absolute and relative clauses introduced by the relative pronoun *who* used as nominal expressions do occur:

(12) English (www)
I missed [*who you used to be*].

(13) English (www)
so that's [*who took those sexy cat photos*]

There may be some overlap between forms that may introduce oriented nominalizations and forms that may not (freely) introduce oriented nominalizations on their own (i.e. without an associated noun, pro-noun or pronominal). In general, however, cleft clauses in *it*-clefts (and similar constructions in other languages) pattern with (restrictive) adnominal relative clauses and not with oriented clausal nominalizations (i.e. "free" relative clauses).

There is, however, the intuition that cleft clauses are not (restrictive) relative clauses. Or, if they are, they do not form a relative clause construction together with the words they occur adjacent to (i.e. with the clefted constituent). Rather, if the cleft clause in the English *it*-cleft forms a constituent with another expression in the sentence, it is with the cleft pronoun. I will come back to this later. In the following sub-section, I will address the problem of distinguishing *it*-clefts from superficially similar constructions sometimes described as "presentational" or "presentative" sentences. A comparison between *it*-clefts and these constructions will serve to explain in what sense the clefted constituent and the cleft clause do not form a constituent.

5.1.2 Contrast with so-called presentative/presentational constructions

It is often mentioned in the literature that sentences such as the following (disregarding prosody and context) may be ambiguous:

(14) English (constructed)
It was a kitten that was trying to get in.

The sentence in (14) allows two bracketing possibilities. Consider the sentences in (15) and (16). Only the latter one is a cleft.

(15) English (constructed)
(*What was making that noise?*)
It was [*a kitten that was trying to get into the kitchen*].

(16) English (constructed)
(*Did you say a puppy was trying to get into the kitchen?*)
No, it was <u>a kitten</u> [*that was trying to get into the kitchen*].

Both can be described as specificational sentences. In (15), the whole expression in square brackets is the VALUE expression (the VARIABLE expression is the pronoun *it*). In the cleft in (16), *a kitten* is a VALUE expression, and the cleft clause contains the VARIABLE. The cleft pronoun is presumably part of the VARIABLE expression. The sentences can be paraphrased as follows:

(17) English (constructed)
(*What was making that noise?*)

a. *It was* [*a kitten that was trying to get into the kitchen*].

b. [*The thing that/ what was making that noise*] *was* [*<u>a kitten that was trying to get into the kitchen</u>*].

(18) English (constructed)
(*Did you say a puppy was trying to get into the kitchen?*)

a. *No, it was <u>a kitten</u>* [*that was trying to get into the kitchen*].

b. *No,* [*the thing that/ what was trying to get into the kitchen*] *was <u>a kitten.</u>*

The pronoun *it* in the "presentational" sentence in (17)(a) is equivalent to the (*th-* or *wh-*) cleft clause in (17)(b). Both are specificational sentences but only (17)(b) is a cleft. (The reason is that in (17)(a) it is not an oriented clausal nominalization but a pronoun.) Let us turn to the sentences in (18). In principle, the (cleft) pronoun *it* in (18)(a) may be said to be equivalent to the cleft clause in (18)(b) as well, but in (18)(a) it co-occurs with a cleft clause. So instead of involving two terms (corresponding to the VARIABLE and the VALUE), the *it*-cleft in (18)(a) involves three. I will address this issue in the following sub-section.

5.1.3 The status of the cleft pronoun

To begin, let's assume that we want to analyze *it*-clefts as specificational sentences. We need to assign the VARIABLE status to one expression and the VALUE to another. The problem is, we have three expressions. There are basically two solutions to this problem. I will refer to these as the extraposition analysis and the discontinuous-constituent analysis. Under the extraposition analysis both the cleft pronoun and the cleft clause are two exponents of the VARIABLE expression. The cleft pronoun may be described as an anticipatory pronoun (in the sense of Kaltenböck 1999). The pronoun *it* is cataphorically linked with the cleft clause. A parallel can be drawn to other construction types. Compare the clausal subject extraposition sentence in (19) and the *it*-cleft in (20). Only the clausal nominalization in (20) is oriented.

(19) English; clausal subject extraposition (www)
It was a disgrace [(*that Africa and Europe should meet up and not discuss Darfur*)].

(20) English; *it*-cleft (www)
It was <u>John</u> [(*who made this mess*)].

Note that the nominalizations in sentence-final positions can be omitted. In that case, the pronoun *it* has nothing to be cataphorically linked with. Perhaps the content expressed is retrievable and need not be uttered. Note that if word order is reversed, anticipatory *it* disappears:

(21) English; clausal subject without extraposition (cf. (19))
[*That Africa and Europe should meet up and not discuss Darfur*] *was a disgrace*.

(22) English; wh-cleft (cf. (20))
?[*Who made this mess*] *was John*.

Now, the sentence in (22) is, to judge from some accounts in the literature, presumably problematic for many speakers of English. The acceptability of such a reversal may be more severely affected if the cleft clause is not introduced by a relative pronoun (i.e. a *wh*-form) but by the subordination marker *that* (cf. ???*that made this mess was John*), or if the cleft clause is bare. Note, however, that (non-oriented) subject clauses in canonical subject position (i.e. without extraposition and *it*-insertion) may also be considerably degraded if not introduced by *that.* Consider the following examples:

(23) English (www)
a. *It's a pity* [*they didn't meet*].

b. *[*They didn't meet*] *is a pity*.

(24) English (www)
a. *It was John* [*I called*]

b. *[*I called*] *was John*.

Perhaps the (b) versions in (23) and (24) are downright unacceptable rather than merely degraded (at least in written form). The (a) versions seem to be fine. But what is the status of the expressions in square brackets? In terms of their distributional potential their status of nominal expressions is problematic. Now, if the expression in square brackets in (24)(a) is not a nominal expression, the status of the sentence as a specificational sentence is problematic. One way to save its status as a specificational sentence is to argue that the cleft clause in (24)(a) is not a nominal expression but rather part of a nominal expression. This would be the approach taken under the discontinuous-constituent analysis. The idea is that an oriented subordinate clause may not be acceptable as a nominal expression by itself but only in combination with a noun or a pronoun. The oriented subordinate clause and the noun (or pronoun) need not, however, occur immediately adjacent to each other. The principle is illustrated in the oft-cited sentences in (25).

(25) English (Francis 2010: 35)
a. *<u>New sets</u> soon appeared* [*that were able to receive all the TV channels*].

b. [*<u>New sets</u> that were able to receive all the TV channels*] *soon appeared*.

The expression is square brackets in (25)(a) does not, by itself, have the distributional potential of a nominal expression. The one in (25)(b) clearly does. The discontinuous-constituent approach to *it*-clefts may be regarded as an attempt to solve at once the problem concerning the status of the cleft pronoun as part of the VARIABLE expression in a specificational sentence, and to account for the formal similarity between cleft clauses and restrictive adnominal relative clauses. In 1.1.3 in Chapter 1, I mentioned that this analysis of *it*-clefts corresponds to the original analysis proposed by Jespersen (1949[1927]). I will discuss two further accounts following the discontinuous-constituent approach in 5.3.2 and 5.3.3. For the moment, I will point out two problems with this approach. Firstly, the pronoun *it* seems not to occur (at least in standard present-day English varieties) adjacent to (i.e. forming a constituent with) restrictive relative clauses. Compare the sentences in (25) with the sentences in (26).

(26) English (www)

a. (…*when I needed help in Calgary*,) *it was* <u>*John*</u> [*I called*].

b. *[*It I called*] *was* <u>*John*</u>.

Another problem with the discontinuous-constituent account is that, as will be discussed in 5.2.2, in a number of languages, the cleft pronoun is systematically absent and yet cleft sentences occur in which the cleft clause does not have the form of a nominal expression but that of a restrictive adnominal relative clause. These languages also systematically lack anticipatory pronouns cataphorically linked with (non-oriented) extraposed clausal subjects, which argues for considering the cleft pronoun an anticipatory pronoun rather than part of a discontinuous construction.

5.2 Two patterns: Germanic and Romance

There are two basic patterns I will discuss in this chapter. The first pattern consists of a matrix copular clause with a pronoun or demonstrative (cleft pronoun) in subject position and a nominal predicate as its two main constituents. A subordinate clause with the form of an adnominal relative clause follows the nominal predicate but the relationship between these is not that of head/antecedent and a modifying relative. The second pattern does not involve a cleft pronoun. The first pattern occurs in all major modern Germanic languages (and in French). The second pattern occurs (albeit not to the same extent) in all major modern Romance languages other than French (except for Romanian; see Georghe 2017). The patterns are illustrated in parallel with an English and a Portuguese sentence for the Germanic (27) and the Romance (28) pattern respectively.

(27) English (www)

It	*was*	<u>*she/her*</u>	[*that ran away*].
CLEFT PRONOUN	COPULA	CLEFTED CONSTITUENT	CLEFT CLAUSE

(28) Portuguese (www)

	Foi	<u>*ela*</u>	[*que fugiu*].
	COPULA	CLEFTED CONSTITUENT	CLEFT CLAUSE

5.2.1 The Germanic pattern

The following examples from the survey on relativization and clefting in major Germanic and Romance languages by Smits (1989) illustrate the Germanic pattern.

(29) German (Smits 1989: 246)
Es war <u>*der Komplize*</u>, [*den er verurteilte*].
'It was <u>the accomplice</u> [whom he condemned].'

(30) Dutch (Smits 1989: 346)
Het waren <u>*zijn eigen broers*</u> [*die hij aangaf*].
'It was <u>his own brothers</u> [whom he denounced].'

(31) Icelandic (Smits 1989: 311)
Það voru hestarnir/hestana [*sem María sá*].
'It was <u>the horses</u> (NOM/ACC) [that Maria saw].'

(32) Danish (Smits 1989: 231)
*Det var mig/*jeg* [*som købte den jakke*].
'It was me/*I [that bought this jacket].'

(33) Swedish (Smits 1989: 398)
*Det var jag/*mig* [*som köpte denna rock*].
'It was I/*me [that bought this coat].'

(34) Norwegian (Smits 1989: 361)
Det var jeg/meg [*som kjøpte denne frakken*].
'It was I/me [that bought this coat].'

(35) French (Smits 1989: 298)
Ce sont ces vestes [*que nous avons achetées*].
'It's these coats that we bought.'

There is some variation within the pattern. For instance, the copula may show agreement with the clefted constituent in some languages (e.g. German, Dutch, and Icelandic, and French) but not in others (see Harmann & Heycock 2020). Case-marking on the clefted constituent may also vary. For instance, in the Icelandic example in (31) the clefted constituent corresponds to the direct object of the verb in the cleft clause and the clefted constituent may inflect for the nominative or the accusative. In the Danish example in (32), the personal clefted pronoun cannot take the subject form (*jeg* 'I') but must take the non-subject form appropriate in the language for pronouns in post-copular position. The opposite is reported for Swedish (33). In Norwegian (34) there is a choice between the corresponding forms. But note that Smits (1989) reports that the use of the subject pronoun form (*jeg*) is not possible if the role/function (in the cleft clause) is not the subject. In this case, *mig* 'me' is required.

Note that all of the languages considered in examples (29)-(35) allow in principle oblique nominal expressions as clefted constituents. Whether the marking of the role/function of the constituent in question occurs on the clefted constituent itself (rather than in the clefted clause) varies from language to language but there seems to be a preference for the marking on the clefted constituent in some languages. Smits (1989: 361) suggests that function/role marking on the clefted constituent rather than on the cleft clause is preferred in Swedish (36)(a) but not necessarily in Norwegian (37). (The strategy illustrated in (36)(b) and (37) involves preposition stranding in the cleft clause.)

(36) Swedish (Smits 1989: 398)
a. *Det var om Etiopien* [*som Pelle skrev en bok*].
'It was about Ethiopia [that Pelle wrote a book].'

b. ?*Det var om Etiopien* [*som Pelle skrev en bok om*].
'It was Ethiopia [that Pelle wrote a book about].'

(37) Norwegian (Smits 1989: 361)
Det er en viktig sak [(*som*) *Knut skriver om*].
'It is an important affair [that Knut writes about].'

Note, on the other hand, that preposition stranding (signaling the role/function of the clefted constituent within the cleft clause) seems to be unproblematic in Swedish clefts where the cleft clause is a free relative. (The null sign *Ø* indicates that the relative particle *som* could occur in this position; I will return to this further on.)

(38) Swedish (Smits 1989: 399)
[*Vad Ø vi bekymrar oss om*] *är* *ventilerna*.
'[What we worry about] are the valves.'

Languages differ as to the kinds of expressions that may be clefted. Swedish, for instance allows *do*-cleft-like constructions, where the clefted constituent is a verb (plus non-subject arguments) as in (39), but apparently only marginally predicative adjectives and secondary predicates (40).

(39) Swedish (Smits 1989: 398)
Det var måla grinden [*Ø Knud gjorde*].
'It was paint the gate [Knud did]'

(40) Swedish (Smits 1989: 398)
%Det var blek [*Ø Kirsten blev*].
'It was pale [Kirsten became].'

In all of the Germanic languages represented in Smits' (1989) survey, clefts involving a cleft clause with the form of an adnominal relative co-exist with clefts in which the cleft clause has the form of a free relative (i.e. "pseudo-clefts"). Consider the following examples illustrating free relative patterns in Swedish. (Note that the order of constituents can be inverted.)

(41) Swedish (Smits 1989: 399)
a. [*Den Ø jag såg där i går*] *var* *han.*
'[The one I saw there yesterday] was him.'

b. [*Den som vet mest om det*] *är* *han.*
'[The one that knows most about this] is him.'

c. *Blomkål är* [*vad Ø jag har köpt*].
'Cauliflower is [what I've bought].'

d. [*Vad som mest oroade mig*], *var* *mina fienders tystnad*.
'What bothered me most was my enemies' silence.'

A determiner alone (e.g. *den* common/animate) as in (41)(a) or *vad* 'what' alone as in (41)(c) can introduce free relative. The relative particle *som* can appear in the construction as in (41)(b) and (d). But note that the relative particle on its own cannot introduce a free relative clause construction.

5.2.2 The Romance pattern

The second pattern to be considered occurs in some Romance languages. There is no cleft pronoun but the cleft clause, as in the Germanic pattern, has the form of an adnominal relative. Again, in all languages reported in Smits (1989) to exhibit this pattern, clefts with a free relative as a VARIABLE expression (i.e. "pseudo-clefts") also occur. The contrast between both patterns is illustrated in the Italian examples below.

(42) Italian (Smits 1989: 328)
Sono io [*che ho comprato questa casa*].
'It is I [who bought this house].'
('It is I [that bought this house]' KM)

(43) Itaian (Smits 1989: 328)
[*Chi l'ha pagato*] *sono* <u>*io*</u>.
'[Who paid for it] was <u>I</u>'

The relative particle *che* used in (42) cannot introduce a free relative in (standard) Italian. It is used to introduce restrictive relatives and also complement clauses (like English *that*). The relative pronoun *chi* used in (43) cannot introduce adnominal relatives. The sentence in (43) is a basic cleft in the language ("pseudo-cleft" in some accounts) and the one in (42) a ("non-pseudo-") cleft following the Romance pattern as defined above.

Beside *che*, other elements can introduce the ("non-pseudo-") cleft clause in Italian. For instance, forms combining a preposition and *quale* 'which', as illustrated in (44) may be used. Note that these forms can also be used to introduce restrictive adnominal relative clauses, as illustrated in (45).

(44) Italian (Smits 1989: 314)
E' <u>*Luigi*</u> [*al quale abbiamo prestato l'accendino*].
'It is <u>Luigi</u> [to whom we have lent the lighter].'
('It is <u>Luigi</u> [to which we have lent the lighter] KM).

(45) Italian (Smits 1989: 314)
[*L'uomo* al quale *parlavi*] *è cieco.*
'[The man you were talking to] is blind.'
('[The man to which you were talking] is blind.' KM)

The use of the combination of a preposition and *quale* 'which' is restricted to relative clause constructions involving the relativization of obliques. Note, however, that it is always possible (at least in all major Romance languages) to mark the role/function on the clefted constituent itself, allowing then the use of a general subordinator (e.g. Italian *che*). Consider the example in (46).

(46) Italian (Smits 1989: 314)
E' <u>*ai figli*</u> [*che sta sempre pensando].*
'It is (of) the kids that he is always thinking of.'

As mentioned earlier, the subordinator *che* is also used to introduce complement clauses in the language as in the sentence with a clausal subject illustrated illustrated in (47):

(47) Italian (www)
E' una fortuna [*che i due documenti non siano stati approvati*].
It's a fortune [that the two documents have not been approved].

There is, as in English, also a formal overlap with non-cleft constructions. Note, however, that clausal subject constructions such as the one illustrated in (47) very often involve verbs in the subjunctive mood (as is the case here) while clefts do not.

Note that word order is not definitory of the type of clefts under discussion (i.e. the Romance type "non-pseudo-cleft"). The pattern described in (28) (i.e. COPULA + CLEFTED CONSTITUENT + CLEFT CLAUSE) may vary. Consider the following Portuguese examples:

(48) Portuguese (Guesser & Trianon 2020: 10)
a. *Foi* <u>*o João*</u> [*que comeu o bolo*].
'It was <u>John</u> [that ate the cake].'

b. *O João foi* [*que comeu o bolo*].
'It was John [that ate the cake].'

Both sentences in (48) have in common that the cleft clause has the shape of an adnominal relative clause (and also of a complement clause, e.g. *é verdade* [*que comeu o bolo*] 'It's true [that (s/he) ate the cake']). Perhaps the pattern illustrated in (48)(b), where the clefted constituent precedes the copula rather than following it, may be dispreferred in some varieties of Portuguese (Guesser & Trianon 2020) and it may not be acceptable in other languages. Still, both (48)(a) and (b) have in common that the cleft clause does not have the shape of a relative clause construction which could occur by itself as an argument expression.

Note that cleft sentences involving cleft clauses which could function as regular nominal expressions (albeit restrictedly) may exhibit the same word order pattern as "non-pseudo-clefts". Compare the sentences in (48)(a) and (b) with those in (49)(b) and (c). The canonical pattern of a *quem*-cleft is presumably the one illustrated in (49)(a).

(49) Portuguese (Guesser & Trianon 2020: 10)
a. [*Quem comeu o bolo*] *foi o João.*
'[(The one) who ate the cake] was John.'

b. *Foi o João* [*quem comeu o bolo*].
'John was [(the one) who ate the cake].'

c. *O João foi* [*quem comeu o bolo*].
'John was [(the one) who ate the cake].'

The relative pronoun *quem* 'who' cannot normally (by itself) be used to introduce adnominal relatives in European Portuguese, but it can in other varieties (Guesser & Trianon 2020). Thus, in some varieties of Portuguese, the cleft clause in (49) may not resemble an adnominal relative clause while in others it may. The distinction between "clefts" and "pseudo-clefts" is more straightforward in clefts involving a VARIABLE expression introduced by an article followed by the relative particle *que*. In these cases, the cleft clause has the form of an oriented clausal nominalization and not that of an adnominal relative clause. However, there may also be an overlap with bare *que*-clefts with respect to word order. (This is a formal overlap which does not occur, for instance, between English *it*-clefts and *th*-clefts.) Consider the sentences in (50).

(50) Portuguese (Guesser & Trianon 2020: 15)
a. *Foi a telefonia* [*que o Pedro estragou*].
'It was the telephone signal that Pedro messed up.'

b. *Foi a telefonia* [*o que o Pedro estragou*].
'The telephone signal was the (thing) that Pedro messed up.'

To clarify: The sentence in (50)(a) is a bare *que*-cleft. The cleft clause (in square brackets) does not have the shape of an oriented nominalization. It does not have the distribution of a nominal expression (as defined in Chapter 3) unless, as mentioned earlier, it is used as a non-oriented nominalization (i.e. complement clause). The cleft clause in (50)(b) has exactly the distributional potential of a regular nominal expression. What makes (50)(a) anomalous as a cleft (with respect to the definition adopted in the present investigation) is (exclusively) the form of the cleft clause. The sentence in (50)(b) has in principle the same structure as a (simple) specificational sentence in the language (minus the presence of an oriented subordinate clause). The bare *que*-cleft in (50)(a) does not because the cleft clause does not look like an oriented clausal nominalization.

Note that, as mentioned at the beginning of the present section (5.2), the acceptability of clefts with the cleft clause introduced by a bare general subordinator in Romance varies across languages. Sentences such as the one illustrated in (50)(a) are fully acceptable in Portuguese—and perhaps even preferred to those such as in (50)(b). The situation in Spanish is different. In principle, bare *que*-cefts are possible in the language. Consider the example in (51).

(51) Spanish (Guasch Leguizamón 1951: 325; cited in Dufter 2010: 258)
No fué Dios [*que puso los primeros nombres a las criaturas*], *sino Adán*.
'It was not God [that gave the creatures their first names], but Adam.'

Note, however, that the example is taken from a prescriptively-oriented Argentinian publication and the sentence is originally presented as an example of a pattern to be avoided. (Note that Argentinian Spanish is reported to be especially permissive with respect to the constructions in question.) Dufter (2010: 258) cites Bentivoglio et al. (1999), who argue that bare *que*-clefts such as the one illustrated in (51), where the clefted constituent is a nominal expression not introduced by a preposition, are practically absent even in informal conversation in Venezuelan Spanish. As mentioned in 2.3.3 in Chapter 2, bare *que*-clefts in Spanish are much less problematic in sentences where the clefted constituent corresponds to an oblique and is thus introduced by a preposition, as in (52).

(52) Spanish (Di Tullio & Kailuweit 2012: 146)
a. *Es con Marcos* [*que quiero conversar*].
'It is with Marcos [that I want to converse}.'

b. *Con Marcos es* [*que quiero conversar*].
'With Marcos it is [that I want to converse].'

The bare *que*-cleft is in many cases the preferred (and in some cases the only fully acceptable) pattern in sentences where the clefted constituent is an adjunct, as in (53) and (54)

(53) Spanish (Di Tullio & Kailuweit 2012: 146)
Es por tu bien [*que te lo pido*].
'It is for your good [that I ask (this) from you].'

(54) Spanish (Di Tullio & Kailuweit 2012: 146)
Fue entonces [*que me enamoré de ella*].
'It was then [that I fell in love with her].'

5.2.3 Cleft(-like) constructions involving complement clauses

It is sometimes difficult to draw the line between clefts and similar constructions which, however, do not have an oriented nominalization and thus clearly do not fit my definition of clefts. Consider the following Norwegian examples in (55) and (56), which are described as clefts in Smits (1989):

(55) Norwegian (Smits 1989: 371)
Det var om Etiopia [*at han skrev en bok*].
'It was about Ethiopia [that he wrote a book].'

(56) Norwegian (Smits 1989: 371)
Det var på kjøkkenet [*at Tarald bygde et skip*].
'It was in the kitchen that Tarald built a ship.'

In his investigation of the history of relative clauses in Nordic languages, Wagener (2017) does not report *at* (or its cognates) as an element used to introduce relative clauses. Rather, it is used to introduce non-oriented clausal nominalizations. Thus, the Norwegian examples in (55) and (56) are problematic with respect to the definition of clefts I propose because the cleft clauses do not resemble oriented clausal nominalizations (though they resemble non-oriented ones).Consider also the German examples with very similar sentences in the examples in (57) and (58). Again, the element *daß* (spelled *dass* after the orthographic reform of 1996) is exclusively used to introduce non-oriented clausal nominalizations in the language.

(57) German (Smits 1989: 247)
Es war in der Küche, [*daß Hans ein Schiff zu bauen versuchte*].
It was in the kitchen [that Hans tried to build a ship].

(58) German (Smits 1989: 247)
Es war über Amerika, [*daß Lothar ein Buch geschrieben hat*].
It was on America [that Lothar wrote a book].

Note that there are (in principle possible) alternatives of the sentences in (57) and (58) which may more straightforwardly be described as involving an oriented nominalization. However, they are reported not to be readily acceptable. Compare the sentence in (59) with the one in (57).

(59) German (Smits 1989: 247)
??*Es war in der Küche,* [*wo Hans ein Schiff zu bauen versuchte*].
It was in the kitchen [where Hans tried to build a ship].

The cleft clause in (59) would be less problematic with respect to the definition of clefts I propose but apparently it is not readily accepted by German speakers, at least according to Smits (1989). It must be pointed out, however, that sentences comparable to that in (59) do occur. Consider the sentence in (60).

(60) German (www)
Es war in Amerika, [*wo zuerst der Werbeslogan auftauchte*].
'It was in America [where the advertising slogan first appeared].'

Note, however, that neither of the alternatives in (57)-(58) or (60) seems to be particularly favored in German. It is doubtful that such sentences are frequent in the spoken language (or even in written German).

5.3 English

As announced at the beginning of this chapter, I will discuss the English *it*-cleft to consider an approach to clefts of the Germanic type under which these constructions are regarded as specificational copular sentences and their structure is explained by analyzing the cleft pronoun and the cleft clause as a discontinuous constituent. This approach is the one taken by Jespersen (1949[1927]) in his first analysis of the English *it*-cleft (see discussion in 1.1.3 in Chapter 1). Under the discontinuous-constituent analysis, the cleft clause is viewed as a restrictive relative clause—and not as a subordinate clause resembling a restrictive adnominal relative clause. It is assumed, however, that its antecedent (or "head") is not the clefted constituent, to which it is immediately adjacent, but rather the cleft pronoun. This approach has been adopted subsequently in one form or another in the literature (see e.g. the discussion in Patten 2012a: 5ff for a recent review). In 5.3.2 I will discuss the account by Bolinger (1972), who argues on synchronic terms and provides a number of explanations concerning the difficulty involved in recognizing the cleft clause and the cleft

pronoun as a discontinuous constituent. In 5.3.3, I will discuss the accounts by Patten (2012a, 2012b) and Ball (1991). Both of these accounts approach the *it*-cleft from a diachronic perspective. Patten's (2012a) account is a monograph-length investigation of the history of the English *it*-cleft and an account of its status from a Construction Grammar perspective. She argues that the status of the English *it*-cleft is best understood taking into consideration the stage of the language in which the *it*-cleft developed. She claims that the *it*-cleft developed in Old English as a distinct construction type. Patten's investigation has an important antecedent in Ball (1991), another major investigation of the history of clefts in English. Unlike Patten (2012a), Ball (1991) does not adopt a discontinuous-constituent approach. She situates the origin of the *it*-cleft in Old English and analyzes cleft pronouns as expletive subjects, the use of which, she argues, develops in Middle English. I will comment on some aspects of Ball's account together with Patten's. Before I discuss the discontinuous constituent accounts of the English *it*-cleft, I will briefly provide in 5.3.1 a brief description of the construction and its main parts to give an account of the variation occurring in the construction.

5.3.1 Characteristics of the English *it*-cleft

The English *it*-cleft consists of a clause which consists of the invariable (3rd person neuter) pronoun *it* (the cleft pronoun) in subject position, the verbal copula *be* and a nominal expression (the clefted constituent). This clause is followed by the cleft clause, which has the form of a restrictive relative clause.

There is a construction which is considered in some accounts to be a cleft variant, sometimes referred to as demonstrative cleft[84] (see e.g. Huddleston & Pullum 2012: 1420). In this case, a demonstrative is used instead of the pronoun *it*. Consider the following examples:

(61) English (Ball 1991: 11)
This is not <u>language teaching problems</u> [*that we are talking about*].

(62) English (Patten 2012a: 75)
This is <u>Oliver London</u> [*we're talking about*], *isn't it?*

Patten (2012: 75) notes that in these cases the use of demonstratives deictically or contextually restricts the reference of the entities considered. Note that very similar constructions are not specificational:

(63) English (Ball 1991: 11)
That was quite a big favor [*you did for me*].

In these cases, the clefted constituent does not specify the reference of an entity but provides a qualification of a referent whose reference is already established. The difference is often difficult to assess.

The copula in *it*-clefts (though not necessarily demonstrative clefts) is consistently inflected in the third person singular and thus the cleft pronoun together with the copula tend to be realized as an invariable unit (*it's* in standard contemporary Enbglish varieties). In this way, the invariant *it's* is comparable to a predicator or focus-marking particle, motivating the view that the *it*-cleft may better not be best analysed as a specificational copular sentence but as a construction sui generis. With respect of its TAM inflectional

[84] Note that in some accounts (e.g. Calude 2008), the term demonstrative cleft is used to describe constructions in which the clefted constituent is a demonstrative (e.g. *That's what I was thinking*). In some accounts, demonstrative clefts (in the sense presently intended) are referred to as *th*-clefts (because the demonstratives *this/these* and *that/those* are used as cleft pronouns). But in other accounts, the term *th*-cleft is used to describe cleft constructions where the VARIABLE expression involves a noun, or a pro-noun preceded by an article (e.g. *The man/one who did this was John*).

potential, the copula may remain in the present tense regardless of the tense of the predicate in the cleft clause. In principle, however, it may inflect for tense and take auxiliaries:

(64) English (www)
It couldn't have been <u>them</u> [*who took the necklace*].

In interrogative clauses the order of the cleft pronoun and the copula may invert like a regular interrogative copular sentence:

(65) English (www)
Is it <u>me</u> [*who did this*]*?*

In this respect, the part of a cleft sentence consisting of the cleft pronoun, the copula, and the clefted constituent behaves exactly like a regular copular clause in the language.

The kind of expressions that may occur as clefted constituents is almost unrestricted (see e.g. Collins 1991). Consider the following examples form Delahunty (1984).

(66) English (Delahunty 1984)
a. *It was <u>stupidity and greed</u>* [*that caused the Boston T shutdown*]. (p. 74)

b. *It was <u>the mousse</u>* [*that I decided to eat for dessert*]. (p. 74)

c. *It was <u>Fred</u>* [*that we went to the movies with*]. (p. 75)

d. *It was <u>with Fred</u>* [*that we went to the movies*]. (p. 75)

e. *It was <u>in the hallway</u>* [*that we waited*]. (p. 75)

f. *It was <u>to buy shaving cream</u>* [*that Fred set out for the store*]. (p. 84)

g. *It was <u>that he was supposed to get there early</u>* [*that Fred forgot*]. (p. 82)

h. *It was <u>stealing</u>* [*that Fred regretted*]. (p. 75)

i. *It is <u>nude</u>* [*that Sandy most likes to swim*]. (p. 77)

j. *It was <u>green</u>* [*that he painted this boat*]. (p. 77)

k. *It wasn't <u>on</u>* [*that he pulled his boots*], *it was off.* (p. 76)

Thus, beside direct (66)(a-b) and oblique arguments (c-d), a wide variety of expressions such as local adjuncts (e), purpose clauses (f), finite complement clauses (g) and gerundial complements (h), depictive (i) and resultative (j) secondary predicates, and even particles (k) can be clefted. The extent to which some of these may be described as specificational sentences is problematic. Note that some of these sentences are reported not to be easily paraphraseable as *wh*- or *th*-clefts.[85] On the other hand, the clefting of predicative adjectives (e.g. *It's happy he was*) is reported to be restricted to some varieties (e.g. Irish English, see Siemund & Beal 2011).Note also that *do*-clefts (e.g. *What he did was paint the house green*) are not

[85] Consider the following sentences:

(i) [*How he pulled his boots*] *was <u>on</u>.*

(ii) [*The way he pulled his boots*] *was <u>on</u>.*

I have no information regarding the extent of which such sentences may be acceptable.

commonly construed as *it*-clefts. With respect to the restrictions observed in the different cleft types (basically, *it*-, *wh*- and *th*-clefts), it is important to consider that clefting in English is highly formulaic (Hopper 2004). I will turn now to the properties of the cleft clause.

The cleft clause can virtually always be introduced by *that* (or lack a relative marker in the case of non-subject relativization). But other relative markers are common also. This is relevant with respect to the assessment of the often very problematic status of the cleft clause as an oriented clause. *That* may be also used to introduce (non-oriented) complement clauses in the language. Furthermore, in many (though not all) environments the use of *that* to introduce a non-oriented clausal nominalization is optional (complement clauses generally involving an overt subject expression).

As just pointed out, the cleft clause in an *it*-cleft sometimes resembles a complement clause. Nevertheless, it can be said that cleft clauses resemble very closely (restrictive) adnominal relatives in the language. Importantly, there is evidence suggesting that cleft clauses and (restrictive) adnominal relative clauses show similar patterns with respect to their historical development. Ball (1994b) reports that the patterns for (restrictive) adnominal relative clauses and cleft clauses in English are very similar in the course of the history of the language. She cites a number of accounts that argue for the preponderance of *that* introducing cleft clauses but finds no support in a corpus study for claims that other markers may be considered marginal in this context. She finds that the patterns of relative clause marking have changed continuously in the language throughout a span of seven centuries and that the changes observed apply to cleft clauses as well. That is, (restrictive) adnominal relatives and cleft clauses developed in parallel. The development consists in a rise of *wh*-forms (gradually displacing *that*) from the 17th century until the 19th century, followed by a slow decline of *wh*-forms in favor of *that* in the 20th century. She finds a strong correlation in the patterns of restrictive relative clauses and cleft clauses.

In a corpus study of contemporary English, Collins (1991: 35) finds the following distribution of the markers listed in Table 1.

that	483	64.2%
which	52	6.9%
who	91	12.1%
when	14	1.9%
Ø	108	14.4%
preposition + *which*	4	0.5%
Total	752	100%

Table 1 Count and proportion of cleft clause types in *it*-clefts (Collins 1991:35)

Clearly, the majority involves cleft clauses introduced by the general subordinator *that* (and null), but other alternatives do not seem to be, strictly speaking, marginal.

A relative marker can be omitted, in principle, under the same conditions that make the omission possible in adnominal relative clauses in the language. In clefts, as in restrictive relative clauses generally, the omission of a relative marker is freely permitted if a non-subject is clefted:

(67) English (www)
a. *It's <u>them</u>* [*I want to see*].

b. *It was <u>in June</u>* [*we got married*].

Omission of a relative marker (generally, not only in clefts) for subject-oriented subordinate clauses is, however also reported in older stages of the language and in non-standard varieties. The null sign (Ø) in the following examples indicates a position in which a marker of subordination (e.g. *that*) would be expected in Modern Standard English.

(68) English (ca. 1678; Aphra Behn, *Oroonoko*; cited in Ball 1994b: 189)
It is not <u>Titles</u> [*Ø make Men Brave or Good*].

(69) English (ca. 1611-1623; Shakespeare, *A Winter's Tale*; cited in Tagliamonte et al. 2005: 77)
There was [*a man Ø dwelt by a churchyard*].

(70) English (1596; Shakespeare, *The Merchant of Venice*; cited in Tagliamonte et al. 2005: 77)
I haue [*a minde Ø presages me such thrift*].

(71) English (Cumnock, East Ayrshire, Scotland; Tagliamonte et al. 2005: 98)
And it was <u>them</u> [*Ø had the institute where the pool was*].

(72) English (Cumnock, East Ayrshire, Scotland; Tagliamonte et al. 2005: 90)
There's [*no many folk Ø liked going to the pit to work*].

(73) English (Cumnock, East Ayrshire, Scotland; Tagliamonte et al. 2005: 96)
I have [*a woman Ø comes in on a Thursday morning*].

These patterns seem to appear more commonly in clefts, existential, and possessive constructions (Tagliamonte et al. 2015: 96). Non-standard varieties may allow forms that are associated with free relative clauses exclusively:

(74) English (Lancaster-Oslo/Bergen Corpus; Collins 1991: 42)
'*It's my <u>wife</u>* [*wot's died*]. *Surely I ought to know*,' *he said*.

(75) English (Cullybackey, County Antrim, Northern Ireland; Tagliamonte et al. 2005: 76)
I would have carried [*the groceries what they needed*].

It may be the case that if a variety of English allows a particular element to introduce cleft clauses, the same element is accepted introducing adnominal relative clauses. This has been not investigated systematically to my knowledge but seems to be implied in Ball (1994b).

5.3.2 Discontinuous constituent account: Bolinger (1972)

Bolinger (1972) argues that *it*-clefts in principle behave like constructions involving relative clause extraposition, but that this is difficult to recognize for a number of reasons which lead, in the author's terms, to the "impression that cleft sentences represent a class distinct from ordinary adjective-clause embedding" (p. 111). His argumentation for a discontinuous-constituent analysis runs as follows: Firstly, all elements that introduce restrictive relative clauses (e.g. *that, who, whom, whose, which, where, when, etc.*) can introduce cleft clauses (76), but elements which can only introduce headless relative clauses (*what*) (77) cannot.

(76) English (Bolinger 1972: 111)
a. *It was <u>John</u>* [*who did it*].

b. *It was up here* [*where I put it*].

c. *It will be pretty soon* [*when you have to do it*].

(77) English (Bolinger 1972: 111)
a. [*What I want*] *is this*.

b. **It's this what I want*.

It was mentioned in the previous section that *what* is reported in *it*-clefts in non-standard varieties but also that it is reported in clearly adnominal relative clauses. Again, it might be the case that those varieties that allow the use of *what* in *it*-clefts, also allow it to introduce of adnominal relative clauses. Bolinger (1972) argues that one of the reasons that prevent the cleft pronoun and the cleft clause from being recognized as a discontinuous constituent is the formal similarity between *it*-clefts and sentences with sentence-final clausal subjects, which also involve the pronoun *it* in sentence-initial position. Consider the sentences with a non-oriented clausal subject in (78) and the cleft sentences in (79).

(78) English (Bolinger 1972: 111)
a. [*That he did it*] *was nice*.

b. *It was nice* [*that he did it*].

(79) English (Bolinger 1972: 111)
a. [*Where he lives*] *is here*.

b. *It's here* [*where he lives*].

Note that in both cases, the pronoun *it* occurs only when the clausal nominalization (occurs in sentence-final position. In (78), the clause introduced by *that* has the same status whether it occurs in sentence-initial or sentence-final position. The relationship between the pronoun *it* and the clausal subject in (78)(b) is one of cataphoric co-indexation. In (79), Bollinger argues, the clause introduced by *where* is a free-relative in (a) and an extraposed adnominal relative forming a discontinuous constituent with the cleft pronoun.

Another source of potential confusion, Bolinger argues, is the similarity between constructions involving right-dislocated free relatives and *it*-clefts. Consider the following examples:

(80) English (Bolinger 1972: 112)
a. *It's this,* [*what I want*].

b. *It's here,* [*where I live*].

Bolinger (1972) notes that "comma disjunctures are characteristically ignored in rapid speech" (p. 112). As mentioned earlier, a cleft clause introduced by the relative pronoun *what* would be (in Standard English at least) unacceptable in an *it*-cleft but a sentence with a right dislocated free relative introduced by *what* is not, as illustrated in (80)(a). In a sentence such as (80)(b), the only distinction between an *it*-cleft and a sentence with a right-dislocated free relative would be prosodic (if realized).

Bolinger (1972) points out that the status of the cleft clause as an extraposed restricted relative is obscured by the impossibility of attaching a relative clause directly to the pronoun *it*.

(81) English (Bolinger 1972: 109)
a. *[*It that he stole*] *was money*.

b. *It was money* [*that he stole*].

But replacing *it* with *that* allows both the continuous and discontinuous pattern:

(82) English (Bolinger 1972: 109)
a. [*That which he stole*] *was money*.

b. *That was money* [*which/that he stole*].

Note that these are "demonstrative clefts", described at the beginning of the previous section. Bolinger (1972) argues that the "acceptance of *that which* and [the] rejection of **it that* (or **it which*) has the earmarks of a relatively unimportant fact of surface arrangement, of a kind with the rejection of *he who, they who,* etc in colloquial speech" (p. 110).

5.3.3 Two historical accounts of the development of the English *it*-cleft

Patten (2012a, 2012b) approaches the English *it*-cleft from a diachronically informed Construction-Grammar perspective. She views the *it*-cleft as a specificational copular sentence (like *th*- and *wh*-clefts). The peculiar feature of the *it*-cleft construction, according to the author, is that the subject term is a discontinuous constituent consisting of the cleft pronoun *it* and an (obligatorily) extraposed restrictive relative. Patten (2012a, 2212b) presents an account of the English *it*-cleft where the idiosyncrasies of the construction are explained as involving a fossilization of patterns that were once productive in the language. Once established as a construction in the language, the *it*-cleft developed further, partially independently from the development of the general patterns followed by relative clause constructions in the language. Patten (2012a; 2102b) contrasts a discontinuous constituent analysis to the analysis presented in the diachronic study of the English *it*-cleft in Ball (1991). Ball situates the rise of English *it*-clefts in Early Middle English in parallel with the rise of expletives. For Ball (1991), the *it*-cleft does not involve a discontinuous constituent. Rather, the author argues, the origins of the *it*-cleft are to be found in constructions involving a free relative clause in sentence-final position. Changes in word-order patterns in the language (tending to a rigid SVO order) and the ban of null subjects explain the development of the pattern observed in the *it*-cleft.

For Patten's (2012a; 2012b) discontinuous-constituent account to work, two main pieces of historical evidence must hold. Firstly, the pronoun *it* or its historical predecessor must be attested in older stages of the language as part of a relative clause construction adjacent to the relative clause—together forming a free (or "light-headed") relative clause construction. A second piece of evidence required is that the discontinuous construction should be attested at a stage before the pronoun *it* becomes (quasi-) obligatory in subject position when a clausal subject occurs sentence-finally (i.e. extraposed). Otherwise, a discontinuous-constituent explanation becomes less compelling. Both conditions are met, but the necessity of a discontinuous-constituent explanation remains problematic. I will begin the discussion by examining the patterns involved in the formation of oriented clausal nominalizations in Old English, the period in which Patten (2012a; 2012b) claims the *it*-cleft originates.

5.3.3.1 Nominalization strategies in Old English

In considering the historical development of the *it*-cleft, it must be kept in mind that the forms that introduce adnominal and free relative clause constructions overlap in the earliest attested stages of English (see e.g. Seppänen 2004, Mitchell 1985a, 1985b). This is a fact that Patten (2012a; 2012b), however, does not pay much attention to. There are two patterns involved in the formation of relative clause constructions in Old

English involving overt markers of subordination.[86] One involves the use of the invariable (i.e. non-inflecting) element *þe* (and its variant *ðe*) and the other a demonstrative *se* 'that'. The demonstrative occurs in the appropriate gender, number, and the case corresponding to the referent's role/function in the subordinate clause (e.g. the nominative forms *se, þæt, seo, þa* for the masculine, neuter, feminine, and plural respectively). Note that the neuter form is also used to introduce non-oriented nominalized clauses (i.e. complement clauses). The demonstrative *se* can also be used adnominally (i.e. much like an article).

Consider first the following sentences illustrating relative clauses introduced by the invariable relative particle *þe*.

(83) Old English (*Ohthere*; cited in Ball 1991: 27))
Ohthere sæde þæt sio scir hatte Helgoland [*þe he on bude*].
'Ohthere said that [the district that he lived in] was called Helgoland.'

(84) Old English (*Beowulf*; cited in Ball 1991: 27)
Þa wæs eaðfynde [*þe him elles hwær gerumlicor ræste sohte*].
'Thereafter it was easy to find [(the man) who sought rest from himself elsewhere, farther away…]

(85) Old English (Ælfric, *Catholic Homilies*; cited in Ball 1991: 27)
…and forði ic sprece [*ðe he me het*]
'And for this reason I speak [what he has commanded me].'

The clause introduced by *þe* in (83) is not adjacent to the nominal expression it modifies but can be in some (rather loose) sense be described as adnominal, restrictively modifying the expression *sio scir* 'that/the district'. In (84) and (85) there are no (other) nominal expressions the nominalized clauses can be associated with. Note that in (84) the nominalized clause refers to a human and the one in (85) to an inanimate referent.

The pattern involving the demonstrative *se* introducing relative clauses is illustrated in the following examples:

(86) Old English (*The Battle of Maldon*; cited in Ball 1991: 29)
Het þa hæleða hleo healdan þa bricge wigan wigheardne se wæs haten Wulfstan
'The bulwark of heroes then commanded to hold a bridge a war hardened warrior who/he was called Wulfstan.

(87) Old English (*West Saxon Gospels*, Matthew 15.11; cited in Ball 1991: 30)
ne besmyt þone man þæt on hys muð gæð.
What goes into his mouth doesn't defile a man.

It may be better to argue for the sentence in (86) to be analyzed as involving two nominal expressions in apposition. (Perhaps one may even analyze the clause involving the nominative masculine form *se* as an independent clause where the pronominal form is used as an overt third person pronoun 'he', as suggested in the Modern English translation.) In the case of (87), the neuter nominative form *þæt* can be analyzed as a relative pronoun but in the case at hand, it clearly introduces a free relative clause construction.

[86] I will ignore free-choice ('whatever') relative clause construction of the following type:

(i) Old English (*West Saxon Gosples*, Mark 6.23; cited in Ball 1991: 26)
Soðes ic þe sylle [*swa whæt swa þu me byddest*]…
'Truly, I will give you [whatever you ask of me']

Note that a demonstrative and the element *þe* may be combined. Where a demonstrative and the relative particle occur in combination, two patterns are possible. In one case, the demonstrative is inflected for the case corresponding to the role/function of the referent in the main clause. In the other, the demonstrative occurs in the case corresponding to the role/function in the subordinate clause. Ball (1991: 31) describes these patterns as "demonstrative + complementizer" and "compound relative" respectively.[87] In later stages of the language *þe* is replaced (or merges with) the neuter form of the demonstrative *þæt*.

Note that the forms corresponding to *þæt* (i.e. *þat/that*) can introduce constructions which may be analyzed as free/headless relatives (as in example (87) above) at least until the late fifteenth century. (The use of *what* as a relative pronoun introducing non-multiple-choice relatives is a relatively late development). Consider the following example from Middle English involving the form *that* introducing a free relative clause.

(88) Middle English (1485; Malory, *La Morte d'Arthur*; cited in Ball 1991: 318)
It maye not be fals <u>that alle men say</u>…
'<u>What all men say</u> cannot be false…'

A problem with a construction such as the one presented in (88) is that it is difficult to decide whether the clause introduced by *that* should be analyzed as a free relative clause co-referent with the pronoun in sentence-initial position (cf. *it cannot be false, what all men say*) or as forming a constituent with it. In other words, whether the pronoun is in a cataphoric relation with a free relative clause or whether the clause introduced by *that* is an extraposed adnominal relative clause, is not a trivial question to answer. I will turn now to data which could be taken to support the second analysis—and thus Patten's (2012a, 2012b) proposal.

5.3.3.2 On (*h*)*it*+relative clause constructions in the history of English

Free relative clause constructions headed by demonstratives and pronouns (e.g. *that which*, *he who*) are in principle possible in present-day English, though perhaps the more common pattern involves constructions introduced by the relative pronoun *what* alone, or constructions involving noun (or pro-noun) anchoring/support as in *the one that* or *the thing that*. It was mentioned earlier, however, that the combination of the pronoun *it* and a relative pronoun or relative marker (e.g. *it who, it which, it that*) is not commonly found in the present-day language. Patten (2012b: 558) and Ball (1991: 263) show that the neuter pronoun *it* may occur directly adjacent to relative clauses (as a "light head", or forming a "compound relativizer") at least until the end of the Middle English period (i.e. until ca. 1500).

(89) Middle English (early 15th c; *The Book of Privy Counselling*; Ball 1991: 59; Patten 2012b: 558)
þis is it þæt settiþ þee in silence.
'This is what sets you in silence.'

(90) Middle English (ca. 1450-1460; Pecock, *The Folewer to the Donet*; cited in Ball (1991: 263)])
To þe xx^e^ argument it is to be seid þat þe ij^e^ premysse is vntrewe. And to <u>it which is arguyd for proof of þilk ij^e^ premysse</u>, it is to be seid þat…
'To the 20th argument it is to be said that the 2nd premise is false. And to <u>that which is argued for proof of this 2nd premise</u>, it is to be said that…'

[87] The patterns can be further combined in constructions involving a demonstrative and a "compound relative". That is, a demonstrative occurring in the case according to the role/function of the referent in the main clause, followed by a demonstrative in the case corresponding to the role/function of the referent in the subordinate clause, followed by the relative particle.

Data such as that presented in (89) and (90) are crucial as evidence supporting a discontinuous-constituent analysis because they show that the pronoun *it* and a relative pronoun (or marker) can (in principle) form a constituent.

Ball (1991: 59), reports that relative clause constructions headed by *it* do occur in Middle English and even in Early Modern English but, importantly, not in Old English. The constructions she finds, however, only refer to non-human (though not necessarily inanimate) referents.[88] This is a problem for a discontinuous-constituent analysis of clefts and it could be taken to support an approach proposing an "expletive" or "dummy" status for the cleft pronoun. Another problem is that in all attestations of constructions involving the pronoun (*h*)*it* adjacent to a relative clause, the relative clause occurs in sentence-final position. Ball (1991: 264) reports that (*h*)*it* + relative clause (in clefts or otherwise) is not attested in sentence-initial position at all. Patten (2012b: 559) cites Suárez-Gómez (2006: 80) and O'Neil (1976) reporting that at least until Late Old English relative clauses tend to occur in clause-final position. In Old English, according to Suárez-Gomez (2006), "extraposed relative clauses are almost twice as frequent as clause-internal relative clauses." (p. 80). These facts suggest that the reasons for the pronoun and the relative clause to systematically occur in separation may be found in general tendencies in the language (Patten 2012b: 559). Still, the fact remains that no unambiguously non-extraposed relative clauses are attested in the stage of the language in which Patten (2012) claims the *it*-cleft to originate.

5.3.3.3 On the origin of the English *it*-cleft: Old or Middle English?

In Ball's (1991: 27) account, the Old English antecedent to the *it*-cleft is a copular clause with the structure she describes as NP/PRO + BEON + REL-CLAUSE. A nominal expression precedes a form of the verb *beon* 'be', which is followed by a (free/headless) relative clause. The pattern is illustrated in (91).

(91) Old English (*Ælfric, Catholic Homilies*; Ball 1991:27; Patten 2012b: 556)
min fæder is [*þe me wuldrað*]
'It is my father [that glorifies me].'

In the sentence in (91) the pre-copular term is the VALUE expression, and the VARIABLE expression occurs in post copular position. Considering that the relative particle *þe* may introduce free relative clause constructions, the sentence in (91)(a) fits perfectly well the definition of clefts proposed in this dissertation. It does not feature any of the of the idiosyncratic properties of the *it*-cleft.

According to Ball (1991), *it*-clefts only developed in Middle English. One of the earliest examples Ball identifies is the sentence in (92).

(92) Middle English (1280-90; *South English Legendary;* in Ball 1991: 158; Patten 2012b: 556)
'A-bidez,'quath þis holie man: 'ore loured is guod and freo, þe deuel it is [*þat bringuth þis wedur...*]'
'Stay, said this holy man, our Lord is good and free. The devil it is [that brings this weather…]'

Ball (1991: 158) discusses two possible analyses for this construction. She decides to analyze the construction as involving a "topicalized focus" (i.e. the VALUE expression *þe deuel* 'the devil' in sentence-initial position) and an expletive pronoun *it* (p. 159). In the other analysis, the construction would be

[88] Consider the following example.

(i) Middle English (ca. 1440; *Gesta Romanorum*; cited in Ball (1991: 59)
Goo ȝe out of þe hall echon, for I se [*it for whos love I suffre this torment*]
'Everyone of you leave the hall, for I see [the one for whose love I suffer this torment]' [= the serpent]

analogous to constructions in Old English following a pattern (similar to that) described by the author as PRO + HIT/ÞÆT + BEON + REL-CLAUSE. In this pattern, according to Ball, the pronoun *hit* would not be an expletive. The pattern is illustrated in the following example:

(93) Old English (*Lives of Saints, Eustace*; cited in Ball 1991: 68)
…and axodon hine hwæðer he hit wære [*þe heora cempena lareow geo wæs*].
'…and asked him whether it were he [who formerly was the teacher of their soldiers].'

In principle, this could be described as a cleft. Ball (1991: 68) points out that *hit* is the complement (i.e. not the subject) of the copula and thus does not have the structure of an *it*-cleft. It is not clear from Ball's account how she would propose to describe the relation holding between the pronoun *hit* and the relative clause introduced by *þe*. A discontinuous-constituent analysis would be in principle possible. Ball (1991: 68) argues against its status as a (specificational) cleft arguing that the relation between the expressions involved (*he* and *hit* + the relative clause) is not specificational. The author argues that

> [d]espite the translation, this [(93)] is not a specificational *hit*-cleft. What is at issue is not who used to teach the soldiers, but rather who this man is […] the correct translation is in fact: '…and asked him whether he was the one who was formerly the teacher of their soldiers' […] The question is then whether Eustace and the soldiers' master are one and the same. (Ball 1991: 68)

An examination of the text in which the sentence in (93) occurs reveals that in fact the question concerns the identity between the soldier's master—also referred to as Placidas some lines earlier—and Eustace (Placidas Eustachius), who attempts not to be recognized. Now, as discussed in Chapter 2, I propose specification to be a relationship holding between two coreferential nominal expressions within a sentence which may be said to hold as soon as there is an asymmetry concerning the referential status of the expressions involved. This could be the case in (93). I also pointed out, however, that the difference between identity statements and specificational sentences is a rather tenuous one.

Patten (2012b: 556) notes that Ball also dismisses the cleft status of Old English constructions such as the following, where the pronoun *hit* precedes the VALUE expression:

(94) Old English (Ælfric, *Catholic Homilies*; in Ball 1991:39; Patten 2012b: 556)
þa cwædon þa geleafullan, nis hit na Petrus [*þæt þær cnucað*]*, ac is his ængel.'*
'Then the faithful said: It isn't Peter [who is knocking there], but his angel.'

Patten (2012b: 557) points out that Ball (1991) justifies her dismissal of (94) as a cleft arguing that it does not correspond to "the stereotypical *it-cleft* of the Linguistics Literature" (Ball 1991:45). Elsewhere, Ball (1991: 66) states her objection with respect to the status of the sentence in (94) as a cleft in terms similar to those used with respect to the construction in (93). That is, that the sense is not specificational. Considering the (textual) context, however, the sentence in (94) arguably does involve a specificational relation and *Petrus* can be regarded as the VALUE expression. Patten (2012b: 557) argues that Ball's (19919 reason for dismissing constructions such as (94) precursors of the *it*-cleft is that they pose a challenge to an account that relies on the notion of expletive pronouns to explain the cleft pronoun in English *it*-clefts. Since Ball situates the rise of expletives (or the restriction against null subjects) in the Middle English period, *it*-clefts are not expected to occur in Old English. But from this it does not necessarily follow that the pronoun *hit* and a relative clause as in the case of the sentence in (94) need to be analyzed as involving a discontinuous relative clause construction headed by the pronoun. Given the fact that free and adnominal relative clauses formally overlap until (at least) Middle English, a discontinuous-constituent analysis is possible but not necessary. The pronoun may also be analyzed as co-referential with a sentence-final free/headless relative

clause construction. The pattern is available with sentence-final non-oriented nominalized clauses in Old English. Consider the following example:

(95) Old English (Ramhöj 2016: 64)
Hit is awritten on Paules bocum [*ðæt sio Godes luvu sie geðyld*].
'It is written in the book of Paul [that the love of God is patience].

The relation between the pronoun *hit* in cleft sentences such as (93) and (94) above could be regarded as analogous (in principle) to that holding in (95) between the pronoun *hit* and the (non-oriented) clausal nominalization introduced by *þæt*. The use of a pronoun co-referential with a sentence-final non-oriented clausal nominalization is not as firmly established in Old English as it is in the present-day language, as illustrated in (96), where a non-oriented clausal nominalization occurs in sentence final position but the pronoun *hit* is absent. (Compare the corresponding Modern English translation.)

(96) Old English (Ælfric, *Catholic Homilies*)
Micel wurðscipe is cristenra manna, [*þæt gehwilc hæbbe fram his acennednysse him betæhtne engel to hyrdrædene*]
'It is a great honour for christian men, [that every one has from his birth an angel assigned to him in fellowship].'

Note that the sentence in (96) is taken from the same text as the cleft in (94), where a cleft pronoun is used. Patten (2012b) argues that clefts in which a cleft pronoun in Old English is absent (e.g. (91) above) are problematic because they seem to be generally translations from Latin (in the case at hand, *min fæder is þe me wuldrað* from *est Pater meus, qui glorificat me*). Patten (2012b: 556) seems to suggest that if this pattern is present in Old English, it may be restricted to translations. (Note that all the examples of the NP/PRO + BEON + REL-CLAUSE pattern in Ball (1991: 41-42) are in fact translations from Latin originals.) But dismissing constructions without the pronoun *hit* as calques seems problematic. Grønvik (1991) reports clefts to be attested in the oldest recorded stages of all Germanic languages. However, according to this author the oldest attested examples of clefting do not involve cleft pronouns.

Patten's (2012a, 2012b) argumentation with respect to the status of the cleft pronoun may reflect a bias inherent in a project whose explicit goal is to prove the discontinuous-constituent status of the cleft pronoun and the cleft clause. Patten (2012) points out that

> we might speculate that, over time, *it*-clefts have come to align less with (the now integrated) relative clauses, and more with other kinds of extraposition construction. […] while *it*-clefts have a distinct structure, in that they *do not involve the extraposition of a complete subject expression*, the *superficial similarity* that exists between *it*-clefts and cases of *it*-extraposition could well provide support for the *it*-cleft's outward appearance. (Patten 2012a: 160; emphasis mine)

Though Patten acknowledges the formal affinity between *it*-clefts and (non-oriented) clausal subject extraposition constructions, she rejects Ball's (1991: 498ff) approach, under which the development of the English *it*-cleft may be seen not merely as "aligning" with the latter type of constructions but as involving a blending between this kind of constructions and specificational sentences.

5.4 French

There are two major clefting patterns in French. The first one, which we call *c'est…qu-…* construction (e.g. *c'est moi qui l'ai fait* 'it's me who did it') follows a pattern very similar to the English *it*-cleft. The second pattern involves left dislocation and resembles a correlative construction (see discussion in Section 4.2 in

Chapter 4). A discontinuous-constituent analysis (such as that discussed for English in 5.3.2) seems not to have found appealing to explain the status of the cleft clause in *c'est…qu-…* clefts. Following the scheme in Section 5.3 on the English *it*-cleft, I will begin this section by presenting a general description of the French *c'est…qu-…* cleft in 5.4.1. I will then very briefly discuss the properties of the *ce qu-…c'est…*cleft in 5.4.2. An understanding of this construction is crucial for the discussion of the extraposition account. In 5.4.3 I will discuss the extraposition account for the *c'est…qu-…* cleft. In the last part of this section (5.4.4), I will present historical data showing the evolution of the *c'est…qu-…* cleft in French.

5.4.1 General characteristics of the French *c'est…qu-…* cleft

The *c'est…qu-…* cleft, like the the English *it*-cleft, follows the Germanic pattern described in 5.2.1. It features a cleft pronoun, a copula, a clefted constituent, and a cleft clause formally identical to a restrictive relative clause following the clefted constituent. The matrix copular clause differs in some respects from its counterpart in the English *it*-cleft. The cleft pronoun is not a personal pronoun but a demonstrative. In French, the cleft pronoun is the invariable demonstrative *ce*. The copula (the verb *être* 'be') may inflect for tense, person, and number but may take the present tense, third person singular form regardless of the form of the verb in the cleft clause. Consider the following examples:

(97) French (Rouquier 2018: 4)
a. *Ce fut <u>Lulli</u>* [*qui inventa ces symphonies*].
'It was <u>Lulli</u> [who invented these symphonies].'

b. *C'est <u>Lulli</u>* [*qui inventa ces symphonies*].
'It's <u>Lulli</u> [who invented these symphonies].'

(98) French (Rouquier 2018: 4)
a. *C'etait <u>une histoire de revenants</u>* [*qu'elle racontait*].
'It was a ghost story she was telling.'

b. *C'est <u>une histoire de revenants</u>* [*qu'elle racontait*].
'It's <u>a ghost story</u> [she was telling].'

Apparently, there is no context in which sentences such as (97)(a) and (98)(a) may not be substituted for the respective sentences in the (b) examples. Tense-aspect congruence may be freely neutralized (see also discussion in 2.3.3 in Chapter 2). Unlike the copula in the English *it*-cleft, which is reported to occur invariably in the third person singular form, the copula *être* 'be' in the French *c'est…qu-…* cleft can agree in number and person with the clefted constituent (while the cleft pronoun *ce* remains in the singular and is invariably third person). Consider the following sentences:

(99) French (Rouquier 2018: 5)
a. *Ce sont <u>eux</u>* [*qui regardent*]
'It's <u>them</u> [who are watching].'

b. *Ce <u>sont eux</u>* [*que je regarde*].
'It's <u>them</u> [I'm watching].'

Rouquier (2018: 5) notes that number agreement with the clefted constituent is more or less restricted to formal registers. The author refers to Carlier (2004; 2005), who regards the demonstrative as the subject and explains plural agreement in terms of a transmission of the plural feature to an element in subject position (i.e. *ce*) underspecified for number. Rouquier, however, argues that "the existence of examples […] in which

the clefted element is a prepositional syntagma shows that […] regressive plural agreement patterns are probably cases of hypercorrection" (p. 5). Consider the following examples (all culled by the author from the internet):

(100) French (Rouquier 2018: 5)

a. *Ce sont à eux* [*que vont mes premiers pensées*].
'It's to them [that my first thoughts go].'

b. *Ce sont à eux* [*que les Romains empruterunèrent beaucoup de choses à cet égard*].
'It's from them [that the Romans borrowed a lot of things in this respect].'

c. *Ce sont pour eux* [que nous travaillons chaque jour].
'It's for them [that we work every day].'

Note that the agreement patterns in the sentences in (100) are not considered standard. Blanche-Benveniste (2002; also cited in Rouquier 2018: 5) denies the status of full verb (*verbe constructeur*) to *être* 'be' in *c'est* and assigns it the status of an auxiliary with weakly verbal status (*faible degré de verbalité*). That is, the element *c'est* is viewed as a predicator rather than a demonstrative in subject position and copula.

Many different types of expressions may occur in the clefted constituent position, including direct arguments (101)(a-b), obliques and adjuncts (c), adverbs (d), subordinate clauses (e), secondary predicates (f), and infinitives (g).

(101) French (Rouquier 2018: 6)

a. *C'est ce paquet-la* [*que j'ai ouvert en premier*].
'It's that package over there [that I opened first].'

b. *C'était elle* [*qui marchait maintenant devant*].
'It was her [who was now walking in front].'

c. *C'est dans les villes* [*que les réfugiés arrivent*].
'It's in the towns [that the refugees are arriving].'

d. *C'est exprès* [*que tu as cassé le grand miroir dans ma chamber*].
'It's on purpose [that you broke the big mirror in my room].'

e. *Mais c'est justement parce que de tels prodiges n'existent pas* [*que les dispositifs de sauvegarde ont été inventés*].
'But it's precisely because such prodigies do not exist [that safeguard devices were invented].

f. *C'est bien mûres* [*que je les préfère*].
'It's well ripe [that I prefer them].

g. *C'est partir très tôt le matin* [*que je redoutais*].
'It's leaving very early in the morning [that I dreaded].'

The cleft clause can be introduced practically in all cases by the elements *qui/que*. In standard French, the element *qui* is used for subject relativization (and for the relativization of obliques as in *à qui* 'to whom' or *avec qui* 'with whom'), and the form *que* is used for non-subject relativization. Other relative pronouns (and relative adverbs) such as *dont* 'whose' or *où* 'where' (102) are possible. The latter can generally be replaced by *que* (103).

(102) French (Sandfeld 1977[1936]: 131)
C'est là [*où je mange quand je suis tout seul*].
'It's there [where I eat when I'm all alone].'

(103) French (Sandfeld 1977[1936]: 131)
C'est ici [*qu'il faut chercher la mairie*].
'It's here [that one has to look for the town hall].'

The form *que* cannot introduce free relatives in Standard French (just like Italian *che* and Ibero-Romance *que* discussed in the 5.2.2). The form *qui* can introduce free relatives but is subject to some restrictions. A free relative marked only with *qui* in the modern standard language usually has a non-specific/multiple-choice ('whoever') reading. Generally, a relative clause construction involving a demonstrative (a "light-headed" relative clause in the sense of Citko 2004) is preferred for free relatives referring to specific entities. The cleft clause may be assumed to be oriented (at least in cases of argument clefting) but there is a tendency to use the form *que,* which (like English *that*) is identical to the element that introduces complement clauses. Consider the example in (104).

(104) French (www)
J'ai vu [*que tu as utilisé le sèche-linge*].
'I saw [that you used the dryer].'

There is a general tendency to neutralize distinctions in the cleft clause in favor of a general/polyfunctional subordinator *que*. In substandard varieties, gapless relative clauses resembling complement clauses may occur. Consider the examples in (105) and (106).

(105) French (Sandfeld 1977[1936]: 176)
C'est moi [*que je suis Desiré Lecoq*].
'It's me [that is Desiré Lecoq].'
(Lit. 'It's me [that I am Desiré Lecoq].')

(106) French (Sandfeld 1977[1936]: 176)
C'est nous [*qu'on fait le pain*].
'Its us [who make the bread].'
(Lit. It's us [that one makes the bread].'

There is a general tendency to neutralize distinctions in the marking of relative clauses in favor of a general subordinator. However, as in English, the patterns involved in cleft clauses and in (restrictive) adnominal relative clauses seem to converge formally to a certain extent.

As in the case of the English *it*-cleft, the cleft clause in a c'est…qu-… cleft does not form a constituent with the clefted constituent. In their investigation of *c'est…qu-…* clefts, Clech-Darbon et al. (1999: 87) point that a cleft clause may co-occur with a regular restrictive relative clause. Consider the sentence in (107), where the cleft clause (in square brackets) follows a clefted constituent involving a relative clause.

(107) French (Clech-Darbon et al. 1999: 87)
C'est l'article que Chomsky a écrit l'an dernier [*que j'ai lu hier*].
'It's the paper that Chomsky wrote last year [that I read yesterday].'

Clech-Darbon et al. (1999: 87) note that relative clause stacking is not allowed in French (unlike e.g. in English). Consider the sentences in (108). The co-occurrence of two relative clauses, as in (108)(a) is unacceptable without a coordinating conjunction (as in (108)(b)). In English the conjunction is optional.

(108) French (Clech-Darbon et al. 1999: 87)
a. *[*L'article que Chomsky a écrit l'an dernier que j'ai lu hier*] *m'a intéressé*.

b. [*L'article que Chomsky a écrit l'an dernier et que j'ai lu hier*] *m'a intéressé*.
'[The paper that Chomsky wrote last year (and) that I read yesterday] interested me.'

The contrast between the restriction illustrated in (108) and absence of this restriction in the cleft sentence in (107) can be taken as a nice piece of evidence for the special status of the cleft clause in the French *c'cest…qu-…* cleft. It resembles formally very closely a restrictive adnominal relative clause but clearly does not behave like one, because it does not modify the expression immediately preceding it. Tellingly, the title of the account by Clech-Darbon et al. (1999) is "Are there cleft sentences in French?". The authors suggest that the cleft clause in the *c'cest…qu-…* cleft may be best analyzed simply as relative clause contruction which, in this particular configuration, does not require an overt antecedent.

5.4.2 The ce qu-…c'est… cleft

The *ce qu-…c'est…*cleft in French consists of a left-dislocated oriented nominalization followed by a clause consisting of the pronoun *ce,* the copula *être* 'be' and a VALUE expression. The left-dislocated cleft clause may be free/headless or headed. The example in (109) illustrates a cleft with a dislocated free/headless relative clause construction as the VARIABLE expression and an infinitival clause as the VALUE:

(109) French (De Cat 2017: 93)
[*Ce qu'il veut*], *c'est* <u>*rester près de toi*</u>.
'[What he wants] is <u>to stay near you</u>.'

The demonstrative *ce* here occurs twice. Once heading the left-dislocated cleft clause and once in subject position in the following copular clause. The construction can be described as a sort of correlative construction similar to those described in Section 4.2 in Chapter 4. The example in (110) illustrates the VARIABLE expression introduced by an element other than *ce*.

(110) French (De Cat 2017: 93)
[*Une que j'aimerais bien recontrer*], *c'est* <u>*Marie Yacoub*</u>.
'[Someone I'd love to meet] is <u>Marie Yacoub</u>.'

I am calling the construction *ce qu-… c'est* cleft just to avoid the term "pseudo-cleft" but the presence of the demonstrative *ce* introducing the left-dislocated is not definitiory.

The following example illustrates a sentence which can be described as both a simple specificational sentence and a cleft as it features a simple nominal expression and a free relative clause construction in apposition:

(111) French (Apothéloz 2012: 198)
Le premier pas, [*ce qui a tout débloqué*], *c'est* <u>*la suppresion du priviège de juristiction*</u>.
'The first step, [what opened everything up], was <u>the suppression of jurisdiction privilege</u>.

Either of the two left-dislocated expressions could be omitted. Now, the pattern in *ce qu-...c'est* clefts follows a general pattern in spoken French where "non-pronominal elements are obligatorily dislocated if they are topics" (De Cat 2017:77). Consider the predicational sentence in (112).

(112) French (De Cat 2017: 77)
a. *Le malais, c'est difficile*.
b. *#Le malais est difficile*.
'Malay is difficult.'

De Cat (2017) notes the obligatory dislocation of subjects in (spoken) sentences involving individual level predicates, which, she argues, "force a topic interpretation of their subject […] irrespective of the context" (p. 77). The pattern found in *ce qu-… c'est* clefts can be regarded as following a very general pattern in the language in sentences exhibiting a topic-comment information structure.

5.4.3 The extraposition account for the French *c'est…qu-…* cleft

A discontinuous-constituent analysis such as that discussed in 5.3.2 for the English *it*-cleft has not received much attention in the literature on the French *c'est…qu-…* cleft. (In fact, I am not aware of any account that explicitly argues for a discontinuous constituent explanation.) The approach favored for French analyzes the cleft clause as an extraposed free relative co-indexed with the demonstrative in subject position.

Rouquier (2018: 11) mentions Sandfeld (1977[1936]) and Togeby (1983) as proponents of the extraposition approach. Togeby (1983: 899) argues that in cleft sentences there is a free relative as a subject and a nominal expression as a predicate. He compares the following examples:

(113) French (Togeby 1983: 899; cited in Rouquier 2018: 11)
a. [*Qui changea*], *ce fut <u>le duc</u>*.
'[The one who changed] was <u>the duke</u>.'
(lit. 'Who changed, it was the duke.')
b. *Ce fut <u>le duc</u>* [*qui changea*].
'It was <u>the duke</u> [who changed].'

The point here is that the difference between the sentences in (113)(a) and (b) can be described simply in terms of word order. In (113)(a) the VARIABLE expression is left-dislocated and in (b) it is extraposed to the right periphery. The pronominal form *ce* is present in both variants. In one case its relation to the VARIABLE expression is anaphoric and in the other case cataphoric. There is arguably not much motivation to posit a discontinuous constituent analysis. Note that in Togeby's examples the cleft clauses are in both cases bare (in the sense that they are introduced simply by the relative pronoun *qui*). This is not the regular pattern for a sentence such as (113)(a). The regular pattern would involve the demonstrative *ce* (or rather, *celui*, for a human referent) preceding the relative pronoun. The sentence in (113)(a) is presumably not completely unacceptable, however.

In his historical investigation of relative clauses in French, Kunstmann (1990: 286) accepts a cataphoric relation between the cleft clause and the cleft pronoun in older stages of the language but argues that constructions in which the cleft clause can be regarded as a relative (free or otherwise) in the present-day language are better regarded as archaisms. He argues that in Modern French "it would be better to avoid the

term ‘relative’ in relation to clefts” (Kunstmann 1990: 283).[89] To illustrate this point, he gives the following examples:

(114) French (Kunstmann 1990: 283)
a. *C’est* <u>*en forgeant*</u> [*qu’on devient forgeron*].
‘It’s <u>by smithing</u> [that one becomes a blacksmith].’

b. *C’est* <u>*contre son gré*</u> [*qu’il a donné son accord*].
‘It’s <u>against his will</u> [that he agreed].’

It is arguably the case that the subordinate clauses introduced by *qu-* in the sentences in (114) are not oriented nominalizations and the relationship between the post-copular expression and the subordinate clause is not specificational. Kunstmann argues that clauses introduced by *qu-* in *c’est…qu-…*clefts (in present-day French) are not relative clauses (free or otherwise). The author argues that this is generally the case in (present-day) French *c’est…qu-…* clefts. This not only in cases such as those illustrated in (114), but generally. As reported in 5.3.1, cleft clauses (like relative clauses generally) in French exhibit an alternation between the introductory elements *que* and *qui* (and other forms such as *dont* ‘whose’) depending on the role/function of the relativized argument. Kunstmann (1990) proposes, however, that in (present day) *c’est…qu-…* clefts, “the morpheme *que* plays merely a subordinating role; if *qui* replaces it when the subject is extracted, it is due to the MasQUErade rule” (p. 284).[90] The rule he alludes to—the term is attributed to Perlmutter (1972)—describes the form *qui* as consisting of the subordinator *que* and the vowel *i*. The latter serves as a subject index reflecting the (general) restriction in the language against non-overt subjects. Under this analysis, French *c’est…qu-…* clefts would be better not regarded as a (particular kind of) specificational sentences. In the case of the constructions in (114), I am very much inclined to agree with Kunstmann’s view. But constructions which do not involve the general subordinator *que* and for which the complement (i.e. non-oriented) clause analysis of the cleft clause is therefore not compelling continue to be used, and many cleft clauses introduced by *que/qui* can be readily analyzed as oriented nominalizations. There are in some sense three types of “cleft” constructions in French (excluding now the correlative-like *ce qu-…c’est…* type). At the one extreme, there are archaic constructions which may be considered clefts consisting of an extraposed free relative (the latter introduced e.g by unambiguously oriented pronominal forms such as *dont* ‘whose’ or *où* ‘where’, or prepositions marking the role/function of the participant the toward which the clausal nominalization is oriented). At the other extreme there are sentences like those illustrated in (114), which resemble sentences involving a (non-oriented) extraposed clausal subject (i.e. an extraposed subject complement clause). In the middle, there are argument clefts, which can (in principle) be analyzed as involving an oriented nominalization, but with the problem that the cleft sentence does not look like a normal free relative.

I will now briefly turn to Sandfeld’s (1977[1936]) account, which seems to be the main point of reference for later extraposition accounts. Sandfeld (1977[1936]: 119f) discusses *c’est…qu-…* clefts in a study of subordination in French in a section on free relatives (*propositions relatives indépendantes*). He points out that Modern French (generally) tends to disfavour free relative clause constructions introduced by simple forms such as *qui* ‘who’ or *où* ‘where’ in favour of forms involving a determiner such as *celui qui* ‘the one who’, *ce qui* ‘that which’, or *là où* ‘there where’ (p. 119). In Sandfeld’s account, clefts represent one context (among others) in which bare/simple free relative clause constructions have not been substituted by more

[89] “il viendrait mieux éviter le terme “relative” à propos de la clivée” (Kunstmann 1990: 283).
[90] “le morpheme que ne joue qu’un role de subordonnant; si *qui* le remplace quand le sujet se trouve extraposé, c’est par l’application de la règle MasQUErade” (Kunstmann 1990: 284).

complex forms. Tellingly, the section where his account is found is called "La proposition relative indépendante (sans déterminatif) en extraposition" (p. 119f). This is in stark contrast to Jespersen's (1949[1927]; 1969[1937]) accounts, where the *it*-cleft is discussed in the context of a description of adnominal relative clauses.

Sandfeld (1977[1936]) argues that sentences in which the free relative precedes the clefted constituent may be regarded as exhibiting a primitive (original) pattern, still to be found (albeit less frequently) in the language:

(115) French (Sandfeld 1977[1936]: 119)
[*Qui fut étonné de cette charité de nouvelle espèce*], *ce fut* <u>*le pasteur Laboulaye*</u>.
'[The one who was surprised by this new sort of charity] was <u>the pastor Laboulaye</u>.'

These "primitive" forms are nothing other than clefts of type *ce qu-…c'est* described above. In the example in (115) (as in the examples (113) from Togeby 1983) the free relative is bare. As mentioned earlier, the regular pattern would involve a free relative introduced by a determiner (e.g. *ce*).

Sandfeld (1965[1936]) explains the preponderance of the *c'est…qu-…* construction (instead of the *ce qu-…c'est…* type) in terms of a word order configuration allowing the insistence (emphasis) on the VALUE expression (i.e. the clefted constituent).[91] He argues that this pattern has become normal in the language.[92]

To make his point clearer, Sandfeld (1977[1936]) compares the extraposition of a free relative with the subject extraposition in simple specificational copular clauses. Consider the examples in (116), which follow (in principle) the pattern observed in *c'est…qu-…* clefts, with the difference that the expressions in sentence-final position are not clausal nominalizations.

(116) French (Sandfeld 1977[1936]: 120)
a. *Ce n'est pas elle, la coupable*.
'It wasn't her, the culprit.'

b. *Ce sont les vielles les plus gaies.*
'It's the old ones (who are) the merriest.'

c. *C'était lui l'imprudent.*
'It was him (who was) the imprudent one.'

d. *C'étaient les autres les sages.*
'It was the others (who were) the wise ones.'

e. *C'était la seconde citation la bonne*.
'It was the second citation (which was) the good one.'

[91] The notion of "insistence" associated with a VALUE-VARIABLE configuration is reminiscent of Lehmann's (1984: 359) characterization of "clefts" (VALUE-VARIABLE) and "pseudo-clefts" (VARIABLE-VALUE) in terms of "insistence" and "suspension" respectively (see Section 1.3.3 in Chapter 1).

[92] "c'est l'ordre inverse qui est devenu ordinaire, dès qu'on veut insister sur l'attribut: *ce fut le pasteur qui fut étonné*" (Sandfeld 1977[1936]: 119). Now, the VALUE-VARIABLE order is attested since the earliest stages of the language (and in fact can be traced to go back to Latin). Sandfeld's account has been interpreted as suggesting that the VALUE-VARIABLE order may be a French innovation and has been criticized in this regard (e.g. in Löfstedt 1966) but the interest of Sandfeld's account is not it's historical correctness but his intuition that the cleft clause in a *c'est…qu-* cleft is to be best regarded as a free relative (i.e., in principle, a full nominal expression).

f. *C'est vous l'assassin*.
'It's you (who are) the assassin.'

g. *C'est vous le maître*.
'It's you (who are) the master.'

h. *C'était lui le vrai mari*.
'It was him (who was) the real husband.'

Again, considering the similarity between clefts and this kind of sentences[93] a discontinuous-constituent analysis is not very compelling. The cataphoric demonstrative *ce* could not appear directly adjacent to the extraposed subjects.

Still, cleft clauses in *c'est…qu-…* cleft sentences resemble adnominal relatives. Sandfeld (1977[1936]: 120-121) argues that cleft clauses in *c'est…qu-…* clefts tend to be "confused" with adnominal relative clauses.[94] Sandfeld also mentions a tendency in the popular (substandard) language to neutralize the *qui/que* distinction in favor of *que*.[95] But this trend, as indicated in 5.4.1, does not seem to concern clefts exclusively. The trend can be better described as a general tendency involving the development of a relative pronoun into a general subordinator. This seems to be a widespread tendency in Romance languages (see Dufter 2010). In the following sub-section, I will discuss the historical processes involved in the development of the *c'est…qu-…* cleft reported in the literature.

5.4.4 Historical development of the French *c'est…qu-…* cleft

The historical processes involved in the development of the *c'est…qu-…* cleft can be resumed in the following three points:[96]

- Free and adnominal relatives become formally distinct. Free relatives introduced by bare relative pronouns are replaced by those involving complex forms. Bare relatives continue to be used but to some extent restricted for free choice ('whoever') uses.
- A cleft pronoun becomes obligatory in VALUE-VARIABLE configurations.
- There is a historical process involving the reduction/simplification of the cleft clause (parallel to the simplification/reduction of relative clauses generally). The reduction/simplification of the cleft clause renders it not only similar to an adnominal relative clause but also to a complement clause. This development can be described in terms of a shift in the function/role marking from the cleft clause to the clefted constituent.

In Old French bare relative pronouns could regularly introduce free relatives (as was also the case in Latin; see example (5) in Section 3.2.3 in Chapter 3). I will focus on the use of the relative pronouns *qui* and *que*.

[93] I do not have any information on the prosodic similarities between a right-dislocated constituent of the kind seen in the sentences in (116) and cleft clauses in *c'est…qu-* clefts. Note that only one of the examples (all from literary sources) adduced in Sandfeld (1977[1937]) involves a comma. De Cat (2017: 34-43) claims that right-dislocation in spoken French can be generally distinguished prosodically but that a pause is not necessary.

[94] "La transposition a eu pour conséquence que la nature de la proposition relative indépendante a été méconnue et qu'elle a été confondue avec les propositions relatives adjointes." (Sandfeld (1977[1937]: 120).

[95] "la langue populaire a fait de *que* un conjonctif universel qui peut s'employer dans tous les cas possibles" (Sandfeld 1977[1937]: 175).

[96] A point I will not discuss concerns the rise and establishment of the *c'est…qu-…* cleft as a high-frequency construction in the language. For an account of this development see Dufter (2008) and references there.

(The development of use of *que* is especially relevant, as it is in cases involving this form that the distinction between adnominal and free relatives is most clear, and where formal overlap with non-oriented clausal nominalizations occurs.) In Old French, the distribution of the forms *qui* and *que* introducing relative clauses was somewhat different than that found in the modern language. In Modern French, *qui* is normally used for relativized subjects and *que* for non-subjects. In Old French, *qui* was used for animates and *que* for inanimate referents (mostly objects but also inanimate subjects, which, however, occur less frequently). Crucially, both could introduce free relatives and the use of bare *qui* was less restricted than in the modern language. Kunstmann (1990: 327-328) notes, however, that *qu*(*e*) instead of *qui* (and instead of *cui* for obliques) in cases involving human referents in free relatives does occasionally occur in Old French. This pattern is observed especially (but not exclusively) in Anglo-Norman texts. Later, the distribution of *qui* and *que* will shift from an animacy-determined pattern to one determined by the role/function of the referent (in the subordinate clause).

The use of complex forms (demonstrative + relative pronoun) gradually substitutes bare *que/qui*. Pierrard (1995: 114) notes that the use of bare *que* is practically ruled out in Middle French for free relatives referring to subjects (i.e. subjects of the subordinate clause) and that it is generally substituted by *ce que* also for objects in this period. Pierrard (1995: 115) cites Wunderli (1978: 221), who argues that in *ce que* sequences in Old French, *ce* is always an anaphoric demonstrative modified by a relative. It is only in Middle French, according to Pierrard (1995: 115), that the use of *ce* becomes obligatory in free relatives involving the relative pronoun *que*. This development is accompanied by the rise of reinforced forms (e.g. *ceci*, *cela*) for the autonomous use of the demonstrative (gradually replacing bare *ce* as a demonstrative). The point here is that from Middle French onwards the complex form *ce* + *que* is necessary to introduce (non-subject-oriented) free relatives. This development, however, will not apply to the cleft clause in the *c'est…qu-…* cleft.

Note that before the complex form *ce* + *que* becomes obligatory as an introducer of free relatives, there is still the possibility of non-adjacency of the demonstrative *ce* and a clause introduced by *que*. This is not normally possible in present-day French. Compare the Middle French sentence (from a play) in (117) and its present-day language version in (118).

(117) Middle French (14th century; *Miracles de Nostre Dame*; Jokinen 1986: 174; in Pierrard 1995: 115)
Tu as tout ce bien entendu
[*Que Sevestre m'a respondu*]
'You have well understood all that Sevestre answered me.'
('You have all this well understood, what Sevestre answered me.')

(118) Modern French (Konrad 2019: 92)
Tu as bien entendu [*tout ce que Sevestre m'a répondu*].
You have well understood all that Sevestre answered me.

The relation between *tout ce* 'all this' and the relative introduced by *que* in (117), however, is not discussed by Pierrard (1995) in terms of discontinuous constituency but of cataphoric relations. This is the same analysis that is posited for French *c'est…qu-…* clefts in accounts regarding the cleft clause as an anomalous (bare *que*) free relative. I will turn now to issues concerning the status of the cleft pronoun.

Recent studies specifically dedicated to the history of clefts in French such as Rouquier (2014) search for constructions with the form *ce* + copula + X + relative. I am not aware of any study also considering clefts without the cleft pronoun *ce* (or a variant of this form) in French. Dufter (2008) points out their existence in

Old French but explicitly notes that he excludes them from his investigation in the assumption "that the great majority of clefts is introduced by *ce* from early Old French onwards" (p. 15). Dufter supports this methodological decision referring to Jochimsen's (1907) investigation of clefts in Old French.

Jochimsen (1907: 22) notes that clefts without a cleft pronoun occur with some frequency in older texts.[97] Consider the following examples (from a translation of the Bible into Old French). Note that the sentence with a cleft pronoun (119) and the one without (120) are taken from the same source:

(119) Old French (12th c.; Le Roux de Lincy (ed.) 1841: 73; cited in Jochimsen 1907: 22)
Quant li prince des Philistiens vindrent encountre ces de Israel, ço fud David [*ki as esturs plus vertuusement se contint e plus pruusement que tuit li autre*].
'When the princes of the Philistines came to encounter those of Israel, it was David [who then behaved more virtuously and more capably than all others].'

(120) Old French (12th c.; Le Roux de Lincy (ed.) 1841: 164; cited in Jochimsen 1907: 22)
Tis frère est [*qui ço t'ad fait*].
'Your brother is [the one who has done that to you].'

As far as I can see, all examples of clefts without *ce* in Jochimsen (1907) are taken from texts from the 12th century (i.e., from the earliest texts he considers in his investigation). At this stage of the language, the pattern involving a cleft pronoun and the pattern without one seem to alternate. As far as I am aware, there is no investigation of the relative frequency of the patterns in the texts in which they. At least in the case of the examples presented by Jochimsen (1907), the relevant factor conditioning the presence or absence of the pronoun seems to be the position of the copula, which never occurs, as far as I can see, clause initially (as it does, e.g. in clefts following the Romance pattern).

Historical accounts of clefting in French and other Romance languages (e.g. Dufter 2008 for French; Roggia 2012 for Old Italian) point out the continuity of the phenomenon between Latin and Romance languages. Hoffmann (2016) reports that in Latin the pronouns *is* and *ipse* can be used in a similar way as cleft pronouns but that these cases are rare. Other accounts (e.g. Löfstedt 1966; Goria 2013) present examples involving the distal demonstrative *ille* but these are sometimes problematic as clefts. (It is often not clear if there is a specificational relation, and, if so, whether the nominalization is the VARIABLE expression.) Hoffman (2016) argues that there is, in any case, no specialized cleft pronoun in Latin.

Unlike similar constructions in other Romance languages, French *c'est…qu-…* constructions involve a cleft pronoun. A subject pronominal form (*ce* or *ço*) is attested in clefts (and other extraposition constructions) since the early stages of French.[98] Such forms do not, however, develop in other Romance languages. Unlike other Romance languages, present-day French does not regularly permit null subjects. There is some disagreement in the literature concerning the chronology of the development of this constraint. Null subjects occur in Old French texts. Balon & Larrivée (2014) argue on the basis of studies of legal (rather than literary) texts that the loss of null subjects does not start as late as the 13th century as often assumed in the literature

[97] "nicht ganz selten finden sich im Altfranzösischen Beispiele, welche logisch durchaus als Hervorhebung gefaßt werden müssen, in denen aber das deiktische Pronomen nicht gesetzt ist" (Jochimsen 1907: 22)

[98] In the earliest stages of French, the pronoun *ce* can be used as a regular demonstrative. Gradually, *ce* loses the status of an autonomous pronominal form and substituted by complex forms such as *ceci* 'this' and *cela* 'that' from the 14th century onward. This process is almost complete by the 16th century (Wunderli 1980: 250; Pierrard 1995: 115). In present-day French, *ce* is used mainly as a determiner and as a "light head" in free relative clauses (*ce que…*) on the one hand, and in combination with the copula (*c'est, ce sont, ce fut, ce furent,* etc.) on the other.

(null-subjects becoming marginal in literary texts only in the 15th century). The authors suggest that null subjects may not have been a regular option in the spoken language already as early as the 12th century. Zimmermann (2009) finds Old French texts with a high frequency of subject pronouns, which would be unexpected for a language with regular null subject. The development of the cleft pronoun thus arguably coincides with that of the ban of null subjects in the language. These developments predate that of the restriction applying to free relatives introduced by bare *que* (and other relative pronouns). The later fact provides an argument against the analysis of the cleft clause as forming a discontinuous constituent with the cleft pronoun and favors an analysis of the cleft clause being (or originating as) a right-dislocated free relative clause. The relation between the cleft pronoun and the cleft clause can be described in terms of cataphora.

There is an historical development in cleft sentences in French (and in Romance in general) tending to a simplification of the cleft clause. This simplification can be described as shifting the locus of function/role marking from the cleft clause (i.e. from the relative pronoun) to the clefted constituent. Creissels (2021) describes this trend as a process tending towards the simplification of cleft sentences from biclausal to monoclausal constructions. This process corresponds to a general tendency, proposed by Harris & Campbell (1995), for complex sentences to simplify (see also Section 1.3.2. in Chapter 1). In the case of clefts, the clefted constituent is treated (with respect to case marking) like the argument of the predicate of the subordinate clause rather than a term of the matrix copular clause. I will discuss below the change of locus of the role/function marking from the cleft clause to the clefted constituent in Romance.

The French *c'est…qu-…* cleft is commonly described in terms of three types "old", "redundant", and "modern" (Muller 2003). The patterns are respectively illustrated in the examples below. Note the diverging distribution of the preposition *à* 'to', which occurs in the "old" type (121)(a) in the cleft clause only. The preposition occurs both in the cleft clause and the clefted constituent in the "redundant" type (121)(b). Finally, the preposition occurs in the clefted constituent only in the "modern" type (121)(c):

(121) French (Muller 2003: 102)

a. *C'est <u>ma mère</u>* [*à qui tu as parlé*].
'It's my mother to whom you have spoken.'

b. *C'est <u>à ma mère</u>* [*à qui tu as parlé*].
'It's to my mother to whom you have spoken.'

c. *C'est <u>à ma mère</u>* [*que tu as parlé*].
'It's <u>to my mother</u> [that you have spoken].'

The "old" type (121)(a) is common in the older stages of the language and corresponds to the pattern found in Latin (where the relative pronoun inflects for the case corresponding to the referent's role/function in the subordinate clause). The "redundant" pattern (121)(b) is regular in Classical French and the "modern" type (121)(c) is considered the regular pattern in the present-day language. But note that the "modern" type is (possibly) found already in Old French[99] and the "old" and "redundant" types are still used in the present-

[99] Though rarely, and the attested constructions are problematic. See discussion in Muller (2003: 111). An example from the 13th century is presented below:

(i) Old French (*Lancelot du Lac*; in Kunstman 1999: 289; cited in Muller 2003: 111)
car a Lancelot fu ce que nos jostames
'car ce fut contre Lancelot que nous combatîmes'
'for it was against Lancelot that we fought'

day language. In some accounts, the use of the "old" and "redundant" types is described as an archaism in the present-day language (but see Muller (2003) for a critique of this position). The historical development sketched out above leads to a pattern in which the cleft clause is introduced by an invariable subordinator regardless of the role/function of the clefted constituent. The pattern alternations just described are not exclusive to French and are reflected in other Romance languages but to different degrees (as was illustrated in the discussion on Spanish at the end of 5.2.2).

5.5 "Clefts" vs. "pseudo-clefts" outside Germanic and Romance

As mentioned in the beginning of this chapter, constructions described as clefts which do not straightforwardly correspond to specificational sentences are reported to occur in several languages outside of Germanic and Romance. In this section, I will discuss data from languages for which a contrast is reported between constructions that can straightforwardly be described as clefts—clearly corresponding to specificational sentences—and constructions in which the cleft clause (or a construction described as such) does not fully resemble an oriented clausal nominalization, rendering its analysis as a term of a specificational sentence problematic The data to be discussed is rather heterogeneous. In 5.5.1 I will discuss cleft patterns in Standard Modern Persian. In this (Indo-European) language, a distinction between "clefts" and "pseudo-clefts" very similar to that described for Romance and Germanic can be observed. This is to some extent also the case in the Bantu language Kirundi, but not in Makhuwa-Enahara, another Bantu language (both to be discussed in 5.5.2). The crucial difference between these languages concerning the "cleft" vs. "pseudo-cleft" distinction is that in the former case free relative clause constructions are formally distinct from adnominal ones. In some clefts with VALUE-VARIABLE order in Kirundi, the cleft clause resembles an adnominal relative clause (and differs from a free relative). In Makhuwa-Enahara, the relevant distinction is absent altogether. In 5.5.3, I will discuss an anomalous construction in Santa María Peñoles Mixtec (Otomanguean; see also discussion of the regular cleft patterns in this language in 3.2.2-3.2.4). The subordinate clause in this type of constructions resembles complement clauses rather than relative clause constructions in the language (which are formally distinct).

5.5.1 Persian

For Modern Standard Persian, Faghiri & Samvelian (2021) report the existence of both clefts corresponding to the English *wh-/th*-cleft type as well as clefts with a cleft clause with the form of an adnominal relative clause. The first type is illustrated in the examples in (122) and (123).

(122) Modern Persian (Kormai and Shahbaz 2010: 54; cited in Faghiri & Samvelian 2021: 192)
[*ānče ke mo'arref=e nazariye=ye me'yār mi-bāš-ad*] *in ketab ast*
what that introducer=EZ theory=EZ standard IPFV-COP.PRS.3SG this book COP.PRS.3SG
'What introduces the Standard Theory is this book.'

(123) Persian Mahootian 1997:118; cited in Faghiri & Samvelian 2021: 192)
[*kas=i ke asb dust dār-e*] *mina=st*
person=RSTR that horse friend have.PRS-3SG Mina=COP.PRS.3SG
'The one who likes horses is Mina.'

These are clefts construed following the regular pattern for copular sentences in the language. They consist of a free relative clause in sentence-initial position (the VARIABLE expression) and a nominal expression (the VALUE expression) immediately preceding a copula verb, which occurs sentence-finally. Free relative clauses in Modern Persian can be introduced by a complex form involving the demonstrative *ān* or an

interrogative pronoun such as *ānče* ‘what’, or *ānke* ‘the one’, and the relativizer *ke* (which can be dropped in this context) or a noun heading a relative clause (the authors mention the general nouns *kes* ‘person’ and *ciz* ‘thing’ as examples). The authors refer to the respective types as *ānče* and *čiz-i ke* type clefts.

Faghiri & Samvelian (2021) mention that an inverted order is only possible in constructions with a relative clause construction headed by a noun (and not by a *wh*-element):

(124) Persian (Khormai & Shahbaz 2010:54; cited in Faghiri & Samvelian 2021: 192)

in	*ketab*	[*čiz=i*]	*ast*	[*ke*	*mo’arref=e*	*nazariye=ye*	*me’yār*]	
this	book	thing=RSTR	COP.PRS.3SG	that	introducer=EZ	theory=EZ	standard	
mi- bāš-ad								
IPFV-COP.PRS.3SG								

‘This book is what introduces Standard Theory.’

Note that the relative clause forms a constituent with the pre-copular noun *čiz* ‘thing’. But it follows the copula forming a discontinuous constituent. The authors mention the use of the “restrictive” clitic *i* on the head noun in relative clauses and the fact that relative clause extraposition is frequent in the language.

The pattern similar to the *it*-cleft consists involves an (optional) proximal demonstrative pronoun (*in* ‘this’), a clefted constituent, a copula, and a clause introduced by the relativizer *ke* in post-copular position:

(125) Persian (Moezzipour 2010: 182; cited in Faghiri & Samvelian 2021: 188)

(*in*)	*farhād*	*bud*	[*ke*	*širin=rā*	*dust*	*dāšt*].	
this	Farhad	COP.PST.3SG	that	Shirin=DOM[100]	friend	have.PST.3SG	

‘It was Farhad who loved Shirin.’

The authors mention that there are no expletive pronouns in the language. The status of the demonstrative remains unclear, and its use is optional.

As in Romance languages and in English, relative clause constructions and complement clauses in Persian involve the same marker of subordination. Consider the examples in (126) and (127). The former illustrates a complement clause (optionally) introduced by the subordinator *ke.* The latter is described by Mahootian (1997) as a cleft sentence in which and adverbial expression is clefted.

(126) Persian (Mahootian 1997: 29)

Fekr-mi-kon-æm	(*ke*)	*šiva*	*emšæb*	*mi-res-e*
thought-DUR-do-1SG	(that)	Shiva	tonight	DUR-arrive-3SG

‘I think (that) Shiva will arrive tonight.’

(127) Persian (Mahootian 1997: 119)

ba	*eštiaq-e*	*tæmam*	*bud*	*ke*	*færiborz*	*be*	*iran*	*bærgæšt*
with	eagerness-EZ	complete	was	that	Fariborz	to	Iran	returned

‘It was with complete eagerness that Fariborz returned to Iran.’

The sentence in (127) illustrates an example of a Persian sentence in which the equational-specificational nature of relation between the clefted constituent and the cleft clause is not immediately clear (the English translation is arguably a sentence with an extraposed non-oriented clausal subject rather than a cleft).

[100] The original gloss is RA (*rā*).

5.5.2 Bantu (Kirundi and Makhuwa-Enahara)

In this section I will discuss data from the Bantu languages Kirundi (Lafkioui et al. 2016) and Makhua-Enahara (van der Wal 2009). Kirundi is reported to exhibit cleft sentences in which the cleft clause resembles an adnominal relative clause, which cannot function on its own as a nominal expression. In Makhuwa-Enahara, there is no formal distinction between free and adnominal relative clauses and thus a contrast between "clefts" and "non-pseudo-clefts" is absent.

Kirundi exhibits an alternation resembling that commonly described in terms of a distinction between "clefts" and "pseudo-clefts". As in the case of Germanic and Romance languages (as well as in Persian), there is in Kirundi a construction with a VALUE-VARIABLE configuration in which the VARIABLE expression does not, by itself, have the potential distribution of a nominal expression. It is formally identical to an adnominal relative clause. This construction contrasts with another one with a VARIABLE-VALUE configuration, in which the VARIABLE expression may be either a free or a noun-anchored ("headed") relative clause construction. There is a further construction in the language with a VALUE-VARIABLE order, in which the VARIABLE expression is described as a "light-headed" relative clause construction. In this latter case, the VARIABLE expression seems to be formally coded as a predicate. Before examining these patterns, I will briefly discuss the characteristics of copular sentences and relative clause constructions in the language as reported by Lafkioui et al. (2016).

Equational and nominal-predicate sentences in Kirundi exhibit the same basic structure. In a nominal-predicate sentence, the subject is followed by the nominal predicate. The latter is preceded by a copula. (Note that the basic word order in the language is SVO.) There are two distinct copular patterns in the language. In main clauses, the copula is a non-verbal element distinguishing polarity. The forms are *ní* (positive) and *sí* (negative), both realized with a high tone. They are written as separate words by convention but must immediately precede the predicate word and form a tight syntactic unit with it. There is also a verbal copula (*rí*), which occurs in TAM other than the present indicative, and also in subordinate clauses. In other contexts, the verb *rí* functions as a locative copula. The examples below illustrate the use of the non-verbal copulas in an adjectival-predicate sentence (128), in a sentence described by the authors as specificational (129), and in a sentence described by the authors as "equative" (130).[101]

(128) Kirundi (Lafkioui et al. 2016: 75)
Iyo ng'iingo nínshaásha.

i-i-o	n-giingo	ní	n-shaásha
AUG_9-PP_9-DEM_{II}	NP_9-measure	COP	NP_9-new

"This measure is new."

(129) Kirundi (Lafkioui et al. 2016: 75)
Umugorési Frida gusa.

u-mu-goré	sí	Frida	gusa
AUG_1-NP_1-woman	NEG.COP	Frida	only

'The woman is not just Frida.'

[101] The subindexed numerals in the Kirundi examples indicate noun-class indexation (unless indicating a singular/plural distinction, in which case they indicate first and second person). The Roman numerals refer to different demonstrative series.

(130) Kirundi (Lafkioui et al. 2016: 75)
Mushíkiwé ní umubáanyi waanje.
mushíkiwé ní u-mu-báanyi u-aanje
his.sister COP AUG_1-NP_1-neighbour PP_1-$POSS_{1SG}$
"His sister is my neighbour."

I will turn now to a brief description of relative clause constructions in the language. Relative clause constructions in Kirundi involve a special verbal form distinguished by "a high tone with a relatively variable surface realization" (Lafkoui et al. 2016: 83). The high tone usually (but not always) occurs on the syllable following the root.[102] The examples in (131) and (132) illustrate "headed" relative clause constructions. In the former, the relative clause construction is subject-oriented, in the latter it is oriented towards the object.

(131) Kirundi (Lafkioui et al. 2016: 83)
(…) *abakózi baávyiize néezá*
[a-ba-kózi {ba-á-bi-íig-ye^{H} néezá}]
AUG_2-NP_2-employee SP_2-REM.PST-OP_8-learn-PFV.REL well
'(…) employees who learned well.'

(132) Kirundi (Lafkioui et al. 2016: 83)
Impanuuro umwaámi yaháaye abaruúndi bíiwé.
[i-n-hanuuro {u-mu-aámi a-á-há-ye^{H} a-ba-ruúndi ba-íiwé}]
AUG_{10}-NP_{10}-advice AUG_1-NP_1-king SP_1-REM.PST-give-PFV.REL AUG_2-NP_2-Rundi NP_2-$POSS_1$
"The advice that the king gave to his Burundians."

A headless relative clause construction requires the prefixation of a morpheme described as "augment" in Bantu linguistics. The pattern is illustrated in (133). Note that this prefix occurs in nominal expressions generally; it is not a marker of nominalization.

(133) Lafkioui et al. 2016: 84)
(…) *abasába ubuhuunzi.*
[a-ba-Ø-sab-a^{H} u-bu-huunzi]
AUG_2-PP_2-PRS-search-IPFV.REL AUG_{14}-NP_2-asylum
"… those who seek asylum."

The pattern illustrated in (133) is available only for subject-oriented clausal nominalizations. In (134) below, the relative clause construction is oriented towards the object. In this case, the augment is not prefixed directly on the the verb. Instead, a specialized form described as "precessive" precedes the verb. The augment prefix attaches to this form.

[102] In the Kirundi examples, the "post-radical" high tone marking verbs in relative clauses is represented by a superindexed H.

(134) Kirundi (Lafkioui et al. 2016: 84)

(…) *Abaruúndi bateze ivyómuzóobaázanira* (…).

a-ba-ruúndi ba-Ø-tég-ye [i-bi-ó

AUG_2-NP_2-Rundi SP_2-PRS-wait-PFV AUG_8-PP_8-PRCS

mu-zoo-ba-əzan-ir-a^{H}]

SP_{2PL}-FUT-OP_2-bring-APPL-IPFV.REL

'(…) Burundians wait for what you will bring them (…)'

I will now turn to Lafkioui et al.'s (2016) account of the different cleft constructions in Kirundi. These are described by the authors as "basic clefts", "pseudo-clefts", and "inverted pseudo-clefts". The latter two are by far the ones most frequently attested in the corpus investigated by Lakioui et al. (2016).[103] The construction described as "pseudo-cleft" exhibits a VARIABLE-VALUE order and is more or less transparently a regular copular construction (following in principle the pattern illustrated in (128)-(130) above). Note that, in cleft constructions, however, the authors gloss the elements corresponding to the non-verbal copulas as "cleft markers". Formally, there is a difference in tone. While the copulas *ní* and *sí* are realized with a high tone (indicated by the acute diacritic) in non-cleft sentences, in clefts the tone is low. An example of a "pseudo-cleft" sentence is given in (135).

(135) Kirundi (Lafkioui et al. 2016: 87)

Icaánzanye ni ukuroondera ubumwé.

[i-ki-á-n-əz-an-ye^{H}]

AUG_7-PP_7-REM.PST-OP_{1SG}-come-CAUS-PFV.REL

ni u-ku-roond-er-a u-bu-mwé

CM AUG_{15}-NP_{15}-look-APPL-FV AUG_{14}-NP_{14}-one

"What made me come was the quest for friendship."

The construction in (135) consists of an oriented clausal nominalization in sentence initial position followed by the copula (or "cleft marker") *ni* (with low tone), which immediately precedes (or encliticizes to) the VALUE expression. In the case at hand, the VALUE expression is also a nominalization (note the augment).

The example in (136) illustrates a sentence described as "basic cleft" by Lafkioui et al. (2016). The order is VALUE-VARIABLE. A copula (in this case, a verbal copula inflected for the remote past tense) precedes the VALUE expression and the (augment-less) cleft clause has the form of an adnominal relative clause lacking an antecedent noun.

(136) Kirundi (Lafkioui et al. 2016: 82)

Baári abatuutsi baároonderwa.

ba-á-ri a-ba-tuutsi [ba-á-roond-er-u-a^{H}]

SP_2-REM.PST-be AUG_2-NP_2-Tutsi SP_2-REM.PST-look-APPL-PASS-IPFV.REL

'It was the Tutsi who were targeted.'

The following examples below illustrate two sentences contrasting the patterns discussed so far. In (137)(a), a "pseudo-cleft" (VARIABLE-VALUE), and in (b), a "basic cleft" (VALUE-VARIABLE). Note the presence of the augment prefix on the VARIABLE expression in the former case and its absence in the latter.

[103] In the corpus investigated by Lafkioui et al. (2016), "basic clefts" represent 7% (n=4) and 13% (n=9) of the total number of clefts in oral and written texts respectively. By contrast, "pseudo-clefts" represent 50% (n= 27) and 43% (n=29), and "inverted pseudo-clefts" 43% (n=23) and 44% (n=30). The corpus consists of a randomly selected sample of 982 written and 942 oral Kirundi sentences (Lafkioui et al. 2016: 74).

(137) Kirundi (Lafkioui et al. 2016: 76, 88)

a. *Ayaséesetse ni amaráso yáacu*.

[a-a-á-sees-ek-ye^{H}]	ni	a-ma-ráso	a-áacu
AUG$_6$-PP$_6$-REM.PST-shed-NEUT-PFV.REL	CM	AUG$_6$-NP$_6$-blood	PP$_6$-POSS$_{1PL}$

'What was shed is our blood.'

b. *Ni amaráso yáacu yaséesetse*.

ni	a-ma-ráso	[a-áacu	a-á-sees-ek-ye^{H}]
CM	AUG$_6$-NP$_6$-blood	PP$_6$-POSS$_{1PL}$	SP$_6$-REM.PST-shed-NEUT-PFV.REL

'It is our blood that was shed.'

I will turn now to the third cleft pattern identified by Lafkioui et al (2016), referred to by the authors as "inverted pseudo-cleft". The construction is illustrated in (138). Note that it is not simply an inverted-word-order version of the "pseudo-cleft" type. It is also not, as the authors explicitly indicate, a variant of the "basic cleft" construction with a fronted VALUE expression resumed by a pronominal form. The authors refer to the VARIABLE expression as a "light-headed" relative clause construction. The element preceding the verb, described as a "substitutive", is argued to function as the "head" of the VARIABLE expression.

(138) Kirundi (Lafkioui et al. 2016: 92)

Abayisaraámu ni bóbaámufashije caane.

a-ba-yisaraámu	[ni	ba-ó	ba-á-mu-fásh-ye^{H}	caane]
AUG$_2$-NP$_2$-Muslim	CM	PP$_2$-SBST	SP$_2$-REM.PST-OP$_1$-help-PFV.REL	much

"Muslims are the ones who helped him [Rwagasore] a lot."

Note that although the configuration in this construction is VALUE-VARIABLE (if I understand the authors correctly), the VARIABLE expression is formally encoded in a manner corresponding to a nominal predicate (minus the low tone of the copula or "cleft marker"). Recall that, as pointed out earlier, the copula forms a tight syntactic unit with the expression it precedes.

The three cleft constructions described above are reported to be associated with different pragmatic properties. As mentioned above, "pseudo-clefts" and "inverted-pseudo clefts" are reported to occur more frequently in texts. It seems that sentences following the "basic cleft" pattern (like *it*-clefts in English and the *c'est…qu-…* cleft in French[104]) do not always clearly resemble equational sentences as they may involve expressions introduced by a copula (or"cleft marker") which do not seem to be in a co-referential relation with the subordinate clause. Elements which apparently cannot be "clefted" in "pseudo-clefts" nor in "inverted pseudo-clefts" (e.g. adverbials) may be "clefted" in "basic clefts". Note, however, that also "basic clefts" are subject to restrictions with respect to the expressions which may be "clefted". Consider the following examples (the English translations of which should arguably be described as sentences with extraposed non-oriented clausal subjects rather than clefts).

(139) Kirundi (Lafkioui et al. 2016: 82)

a. *Ariko ni keénshi tuyareénga*.

ariko	ni	keénshi	tu-Ø-ya-réeng-a^{H}
but	CM	often	SP$_{1PL}$-PRS-OP$_6$-violate-IPFV.REL

'But it is often that we violate them (the laws).'

[104] Note that both French and English are, together with Kirundi, co-official languages of Burundi.

b. *Ariko ni kera tu-á-ya-réeng-a^{H}
but CM long.ago SP_{1PL}-REM.PST-OP_{6}-violate-IPFV.REL
Intended meaning: 'But it was long ago that we violated the laws.'

Lafkioui et al. (2016: 82) point out that it is not clear what kind of elements may be "clefted" and which cannot. In constructions involving argument clefting, Lafkioui et al. (2016: 82) suggest that "basic" clefts may be more restrictive than "pseudo-clefts". Both types of constructions may in principle involve the "clefting" of subject and object arguments, but the authors note that in the case of object clefting, "basic clefts" are "less accepted" (p. 82). This may be related (the authors seem to suggest) with the general relativization patterns in the language.

I will turn now to a review of cleft constructions in Makhuwa-Enahara as described in van der Wal (20016). In this language, "relative" verb forms are identical to those of the "conjoint conjugation" (p. 28). Conjoint verb forms do not necessarily signal nominalization (nor are necessarily used as modifiers in nominal expressions). Unlike other Bantu languages, Makhuwa-Enahara lacks an "augment" prefix. According to van der Wal (2009: 32), there is evidence that the it was lost in earlier stages of the language.[105] The examples in (140) illustrate the contrast between a verbal form in a neutral sentence ("disjoint conjugation") (a), a "headed" relative clause (b), and a free ("headless") relative clause construction (c).[106]

(140) Makhuwa-Enahara (van der Wal 2009: 228)

a. *mwanámwáné* *o-hoó-khwa* (disjoint verb form)
1.child 1-PFV.DJ-die
'A/the child died.'

b. *mwanámwáné* *o-khwa-alé* *o-ri* *owáani* (conjoint/relative verb form)
1.child 1-die-PFV.REL 1-be 17.home
'The child who died is at home.'

c. *o-khwa-alá* *o-rí* *owáani* (conjoint/relative verb form)
1-die-PFV.REL 1-be 17.home
'The one who died is at home.'

Note that, unlike in Kirundi, the form of the verb in a relative clause construction is identical whether occurs with an antecedent noun or not.

A cleft with a VARIABLE-VALUE configuration in Makhuwa-Enahara is shown in (141).

(141) Makhuwa-Enahara (van der Wal 2009: 229)
[*o-khw-aalé*] *mwanamwáne*
1-die-PFV.REL 1.child.PL
'The one who died is a/the child.'

As suggested in the English translation, the sentence in (141) is in principle ambiguous. It allows both a nominal-predicate sentence reading and an equational-specificational one. In the former case the clausal nominalization is interpreted as a fully referential nominal expression and a property is predicated of its

[105] It seems, for instance, that only items that involved the "augment" prefix in previous stages of the language can undergo "predicative lowering" (i.e. the marking of predicative expressions by means of the substitution of high tones by low ones in a word) (van der Wal 2009: 32).
[106] Arabic numerals indicate noun-class membership/agreement. Small caps Roman numerals indicate different demonstrative series.

referent. In the latter case, it can be interpreted as a VARIABLE expression, whose reference is specified by the VALUE expression in second position. The structure is comparable to the pseudo-cleft in Kirundi. In this case, the nominalization takes a special conjugation pattern. Note that in the case at hand the VALUE expression is marked with a special tonal pattern described as "predicative lowering", also used with nominal predicates.[107] Not all nominal expressions, however, allow predicative lowering. In these cases, a copula may be used. Note that, unlike other Bantu languages, in which the "augment" prefix may signal definiteness, in Makhuwa-Enahara, the referential status of a nominal expression is interpreted contextually (van der Wal 2009: 174).

An example involving a VALUE-VARIABLE configuration is shown in (142). Note that the use of the copula is not allowed if predicative lowering is available:

(142) Makhuwa-Enahara (van der Wal 2009: 121)

a. *oravó* [*o-thum-aly-áaka*]
14.honey 14-buy-PFV.REL-POSS.1SG
'it is honey which I bought'

b. **ti* *orávo* *o-thum-aly-áaka*
COP 14.honey 14-buy-PFV.REL-POSS.1SG

There is another pattern with a VALUE-VARIABLE configuration, in which a copula-marked expression occurs in second position. Consider example (143).

(143) Makhuwa-Enahara (van der Wal 2009: 123)
ekanétá [*t'* *í-ki-vah-aly-ááwe* *Aléksi*]
9.pen COP 9-1SG-give-PFV.REL-POSS.1 1.Alex
'a pen is what Alex gave me'

Note that van der Wal (2009: 261) explicitly notes that sentences following the pattern illustrated in (143) may be interchangeable with clefts with a VARIABLE-VALUE order. Consider the possible answers to the question in (144).

(144) Makhuwa-Enahara (van der Wal 2009: 261)
X: "Who is sleeping inside? Abdul?"
Y: "No, it's not Abdul, …

a. [*o-n-rúpá*] *ti Coána*
1-PRS-sleep.REL COP 1.Joana
lit. '… the one who sleeps is Joana'

b. *Coáná* [*t'* *í-ń-rupa*]
1.Joana COP 1-PRS-sleep.REL
lit. '… Joana is the one who sleeps'

[107] The contrast is illustrated in the following examples. The example in (i) shows the underlying tonal pattern of a noun with the meaning 'child'. The example in (ii) illustrates the pattern involving "predicative lowering".

(i) *mwanámwáne* 'child' (LHHL)

(ii) *mwanamwáne* 'it is a child' (LLHL)

The reported pragmatic equivalence between the clauses in (144)(a) and (b) seems to confirm the VALUE status of the expression in clause-initial position in (b). (Note that in this case the VARIABLE expression is formally marked as a predicate, as it is preceded by a copula.) It is not quite clear under which conditions a construction with a VALUE-VARIABLE order such as that illustrated in (142) above (where the VARIABLE expression is not introduced by a copula) may be preferred to those such as the one shown in (144)(b). To judge from examples presented elsewhere by the author, it may be the case that the pattern illustrated in (144)(b) is preferred in cases in which the VALUE expression involves proper names (and perhaps also pronominal forms; see f.n. 55 in Section 4.1, Chapter 4). In any case, though clefts in Makhuwa-Enahara may follow different patterns, none of these result in a construction in which the cleft clause exhibits a form which does not have the potential distribution of a nominal expression. Unlike in Kirundi, a "cleft" vs. "pseudo-cleft" distinction similar to that observed in Germanic and Romance is absent in the language.

5.5.3 Santa María Peñoles Mixtec

Clefts in Santa María Peñoles Mixtec were discussed in 3.2.2 in Chapter 3 in the context of a discussion of nominalization patterns. With the exception of locative-denoting clausal nominalizations, the regular nominalization pattern in this language involves either a noun-anchoring strategy or the introduction of the clausal nominalization by means of a nominal classifier. (The difference between both strategies is not always clear, however, given the overlap between common nouns and classifiers). There are two alternative word order patterns described by Ramírez Pérez (2014) for (regular) cleft sentences in Santa María Peñoles Mixtec. One involves a VALUE-VARIABLE order and in the other the order is VARIABLE-VALUE. These patterns are illustrated in (145) and (146) respectively.

(145) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 7)
ndi²ku²tu³ *kuu²³* [*kɨ²tɨ³* *nde²ni²³*]
bull RLS.be.ICMPL animal RLS.tie.ICMPL
'El animal amarrado es el toro.'
'The animal that is tied up is the bull.'

(146) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 160)
[*kɨ²tɨ³* *nde²ni'i²³*] kuu²³=dɨ³ ndi²ku²tu³
animal RLS.tie.ICMPL RLS.be.ICMPL=3SG.AGT bull
'El animal amarrado es el toro.'
'The animal that's tied up is the bull.'

The VALUE expression may be placed in pre-copular position and the VARIABLE expression in canonical, post-verbal subject position as in (145). Alternatively, The VARIABLE expression is topicalized, with a resumptive pronominal clitic occurring in post-verbal position and the VALUE expression following as in (146). Note that verb in the oriented clausal nominalizations may inflect for aspect and mood and involve the expression of arguments other than the one they are oriented to. With the exception of the obligatory omission of the argument toward which the relative clause is oriented in post-verbal position, and the obligatory presence of a noun or a classifier preceding the verb, the structure of an oriented clausal nominalization is in principle identical to that of a main clause. The following examples illustrate a cleft sentence featuring an object-oriented relative clause construction (147), and one with a relative clause construction oriented toward an intransitive subject (148), respectively.

(147) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 161)
ku²chi³ *kuu²³* [*kɨtɨ²³* *ni³-sa'³-ni³=yu³*]
pig RLS.be.ICMPL animal CMPL-RLS-kill=3PL.GENR
'¿Marrano es el animal que ellos mataron?'
'The animal they killed is/was a/the pig?'

(148) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 163)
xi²chi³ aa¹² kuu²³ [*xi²chi³* *ni³-ndua²¹*]
girl DEM.MED RLS.be.ICMPL girl COMPL-RLS.fall
'La muchacha que se cayó es esa muchacha.'
'The girl who fell is/was that girl.'

The constructions in (145) to (148) may be regarded simply as variants of specificational sentences in which two nominal expressions occur in a subject-predicate like relationship. As mentioned at the beginning of this section (5.5), there is a further construction described by Ramírez Pérez (2014) where this is not the case. The construction in question is illustrated in (149). A corresponding canonical sentence is illustrated in (150) for comparison.

(149) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 168)
chua²¹ *kuu²³* [*saa¹²* *ni³-xi²-te²*]
broth RLS.be.ICMPL SUB COMPL-RLS-spill
'Caldo es lo que se regó.'
'What spilled was broth.'

(150) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 169)
ni³-xi²-te² *chua²¹*
COMPL-RLS-spill broth
'El caldo se regó.'
'The broth spilled.'

The construction in (149) is reported to be pragmatically similar to the clefts illustrated in (147) and (148) but, as will be discussed in the following, differs from these considerably. Perhaps the most salient difference is the form of the post-copular constituent. Unlike the sentences in (147) and (148), the constituent corresponding to the clausal nominalization in the cleft sentences presented earlier does not have the form of a relative clause construction, which, as was discussed earlier, necessarily involves a noun or a nominal classifier (and in the case of locative orientation, a relational noun). Rather, the construction at hand is introduced by the morpheme *saa¹²*, described by Ramírez Pérez (2014) as a subordinator. This morpheme is used to introduce complement clauses, as illustrated in (151) to (153).

(151) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 44)
ka²-chi³=u³de³ [*saa¹²* *a¹dɨu²¹* *mpee²³* *kuu²³=i³*]
RLS-say.ICMPL=3SG.M SUB NEG sheep RLS.be.ICMPL=1SG
'Ellos dicen que no soy borrega.'
'They say that I'm not a sheep.'

(152) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 63)
xi2-ni3=o² [*saa¹²* *ndau¹²=o²¹*]
RLS-see.ICMPL=1PL.INCL SUB poor=1PL.INCL
'Sabemos que somos pobres.'
'We know that we are poor.'

(153) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 123)

ni³-xi²-nde'e²³=i³	[*saa¹²*	*ni³-x-i'i²³=xi³*	*ndɨ³dɨ³*]
COMPL-RLS-see=1SG	SUB	COMPL-RLS-drink=3SG.F	mezcal

'Yo vi que ella tomó mezcal.'
'I saw that she drank mezcal.'

Note, however, that not all verbs that might be expected to take complement clauses introduced by the subordination marker *saa¹²* do so. Ramirez Pérez (2014: 123) points out that while the verb *nde'e²³* 'see' requires the subordinator, the verb *te²ku²-lo'o²³* 'hear' does not allow its use. Consider the sentences in (154).

(154) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 123)

a.

ni³-te²ku²-lo'o²³=xi³	[*ni³-xi³-ta¹-yu'u²³=yu³*]
COMPL-RLS.hear-ear=3SG.F	COMPL-RLS.dance-mouth=3PL.GENR

'Ella escucho que ellos cantaron.'
'She heard that they sang.'

b.

**ni³-te²ku²-lo'o²³=xi³*	[*saa¹²*	*ni³-xi³-ta¹-yu'u²³=yu³*]	
COMPL-RLS.hear-ear=3SG.F	SUB	COMPL-RLS.dance-mouth=3PL.GENR	

Intended: 'She heard that they sang.'

The morpheme *saa¹²* is also used in relative clause constructions, where it functions like a relative particle linking a noun with the subordinate verb. The pattern deviates from the dominant noun-anchoring/ support (and classifier marking) strategies, in in which no subordination marker is present. Ramírez Pérez (2014: 63) reports, however, that this strategy is only possible with inanimate referents. Compare (155) and (156).

(155) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 62)

ni³-xi²-ni³	[*ma¹chi²ti³*	*saa¹²*	*ni³-tna¹nu²*]
COMPL-RLS-see	machete	SUB	COMPL-RLS.break

'Vi el machete que se quebró.'
'I saw the machete that broke.'

(156) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 63)

a.

ni³-x-i²ni²=i³	[*i³na²*	*ni²-xi²´i³*]
COMPL-RLS-see=1SG	dog	COMPL-RLS.die

'Vi al perro que se murió.'
'I saw the dog that died.'

b.

**ni³-x-i²ni²=i³*	[*i³na²*	*saa¹²*	*ni²-xi²´i³*]
COMPL-RLS-see=1SG	dog	SUB	COMPL-RLS.die

Intended: 'I saw the dog that died.'

But note that the pattern seems to be merely optional for inanimate referents. Consider the following sentence, involving a referent of the same sort as that in (155):[108]

[108] The relevant examples in Ramírez Perez (2014) involve, as far as I can see, matrix predicates of the kind otherwise allowing complement sentences introduced by *saa¹²*. But if the matrix predicate were the relevant factor allowing the pattern, the unacceptability of (156)(b) would be difficult to explain.

(157) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 68)
ni³-tna¹nu³ [*ma¹chi²ti³* *ni³-xe²-nde²-ndɨ'ɨ²³=n* *yu²tnu²³*]
COMPL-RLS.break machete COMPL-RLS-cut-APPL=2SG tree
'Se quebró el machete con que cortaste el árbol.'
'The machete you cut the tree with broke.'

Now, unlike in relative clause constructions involving an antecedent noun just discussed, in cleft (or cleft-like) constructions involving the subordinator *saa¹²* the animacy-related restriction does not apply. Consider the following sentence:

(158) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 170)
ña¹dɨ'ɨ²³ *kuu²³* [*saa¹²* *ni³-sa²ña'²¹a-ta¹u²=n* *di²ta¹*]
woman RLS.be.ICMPL SUB COMPL-RLS.give-give.present=2SG tortilla
'Mujer es a quien le regalaste tortillas.'
'The one you gave tortillas is (a) woman.'

Ramírez Pérez reports that the elements that may be "clefted" in these constructions must be nominal. Relational noun constructions (159)(a) and adverbs (b) may not be involved. Note that a goal-denoting nominal expression (159)(c) and a temporal noun (d) are, by contrast, acceptable:

(159) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 171)
a. **nuu²* *xi³to²* *kuu²³* [*saa¹²* *ni³-ke²xio³²=i³*]
RN.face bridge RLS.be.ICMPL SUB COMPL-RLS.arrive=1SG
Intended: 'En el puente es donde llegué.'
Intended: 'The bridge is where I arrived.'

b. **ndɨɨ¹³* *kuu²³* [*saa¹²* *ni³-ke²xio²³=de²*]
fast RLS.be.ICMPL SUB COMPL-RLS.arrive=3SG.M
Intended: 'Rápido es que él llegó.'
Intended: 'It's fast that he arrived.'

c. *be'e²³=yu³* *kuu²³* [*saa¹²* *ke²xio³²=i³*]
house =3PL.GENR RLS.be.ICMPL SUB RLS.arrive.ICMPL=1SG
'En casa de ellos es donde llegué.'
'At their house is where I arrived.'

d. *i³ku²* *kuu²³* [*saa¹²* *ni³-ke²xio³²=i³*]
yesterday RLS.be.ICMPL SUB COMPL-.RLS.arrive=1SG
'Ayer es cuando llegué.'
'Yesterday is when I arrived.'

The "clefted" expression may be possessed (160) and take quantifiers (161)(a-b)

(160) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 172)
ñu'u²³ *ta¹ta¹²* *kuu²³* [*saa¹²* *na¹nde²=u³de²*]
land father RLS.be.ICMPL SUB RLS.fight.ICMPL=3SG.M
'El terreno de papá es lo que pelean ellos.'
'Dad's land is what they are fighting about.'

(161) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 172)

a. *bai*^{12} *dɨ'u*^{23} *kuu*^{23} [*saa*^{12} *ni*^{3}*-da*^{1}*kwi*^{23}*ta*^{1}*=yu*^{3}]
<u>much money</u> RLS.be.ICMPL SUB COMPL-RLS.lose=3PL.GENR
'Mucho dinero es lo que ellos perdieron.'
'Much money is what they lost.'

b. *bai*^{12} *la*^{1}*ndu*^{23} *kuu*^{23} [*saa*^{12} *ni*^{3}*-x-i'i*^{23}]
<u>much children</u> RLS.be.ICMPL SUB COMPL-RLS-die
'Muchos niños fueron los que se murierón'
'Many children was what died.'

Ramírez Pérez, reports, however, that the "clefted" expressions may not involve either adjectives or demonstratives. Consider the examples in (162). (Note that the sentence in (162)(a) stems from a naturally occurring text.)

(162) Santa María Peñoles Mixtec (Ramírez Pérez 2014: 172-173)

a. *ku*^{2}*ru*^{2}*xi*^{3} *kuu*^{23} [*saa*^{12} *ne'e*^{23}*=u*^{3}*de*^{2}]
cross RLS.be.ICMPL SUB COMPL.RLS.have=3SG.M
'Cruz es lo que tienen ellos.' {txt004}
'(A) cross is what they have.'

b. **ku*^{2}*ru*^{2}*xi*^{3} *aa*^{12} *kuu*^{23} [*saa*^{12} *ne'e*^{23}*=u*^{3}*de*^{2}]
cross DEM.MED RLS.be.ICMPL SUB COMPL.RLS.have=3SG.M
Intended: 'Esta cruz es lo que tienen ellos.'
Intended: 'This cross is what they have.'

c. **ku*^{2}*ru*^{2}*xi*^{3} *tnuu*^{23} *kuu*^{23} [*saa*^{12} *ne'e*^{23}*=u*^{3}*de*^{2}]
cross black RLS.be.ICMPL SUB COMPL.RLS.have=3SG.M
Intended: 'La cruz negra es la que tienen ellos.'
Intended: 'The black cross is the one they have.'

Ramírez Pérez argues that the "clefted" elements are necessarily interpreted as non-specific. This unlike specificational sentences, in which the VALUE expression is typically specific if not definite. Now, some of the Spanish translations provided by Ramírez Pérez could be taken to suggests that the relations between the expressions in the *saa*^{12}-constructions might be predicational rather than specificational. However, Ramírez Pérez (2014: 175-177) explicitly notes that, whereas in nominal-predicate sentences the copula may express a variety of aspect-mood oppositions, in the cleft (or cleft-like) constructions involving a post-copular subordinate clause introduced by *saa*^{12} it is restricted to the realis mood and incompletive aspect.

To sum up, the construction is reminiscent of a "non-pseudo-cleft" in that it involves a subordinate clause resembling an adnominal (not a free) relative clause, and also a complement clause. Unlike *it*-clefts in English, *c'est…qu-… clefts* in French, or bare *que*-clefts in Spanish, however, *saa*^{12}-clefts in Santa María Peñoles Mixtec do not tend to resemble extraposed clausal subject sentences. Recall that the "clefted" constituent must be a nominal expression, though apparently one with non-specific reference.

5.6 Summary

This chapter was dedicated to the discussion of cleft constructions of the type often described as "clefts" in accounts distinguishing "clefts" and "pseudo-clefts" as two major types of clefting constructions. The latter correspond clearly to the cross-linguistically oriented definition adopted in this investigation. The former,

do not (or at least not as clearly) because the cleft clause does not exhibit the formal characteristics of an oriented clausal nominalization and thus these constructions cannot be straightforwardly described as equational sentences involving two nominal expressions. An overview of the problems "non-pseudo-clefts" present with respect to the definition adopted in this investigation was provided in Section 5.1. For the most part, the discussion in this chapter concentrated on cleft constructions in Germanic and Romance. I distinguished two major patterns. Clefts corresponding to the Germanic pattern (which include French *c'est…qu-…* clefts) consist of a cleft pronoun, a clefted constituent in post-copular position and a cleft clause following the latter. In principle, the cleft pronoun, the copula, and the clefted constituent exhibit the structure of a complete equational clause (although the cleft pronoun and the copula in combination may be argued to function as a "predicator" or a "focus marker" rather than as a subject and a verb). Clefts following the Romance pattern lack a cleft pronoun. Languages in which this pattern occurs generally do not require the overt expression of a subject argument separate from person inflection on the verb. The status of the cleft clause is problematic in both patterns. Formally, it resembles an adnominal relative clause, and, in many languages, it also resembles a complement clause (though not in all; see Section 5.2.3). The main problem addressed in this chapter concerned the relationship between clefts of the Germanic and Romance type, and clefts more clearly exhibiting the structure of specificational sentences—which also occur in all of the languages examined. There is some consensus in the literature that the former and the latter are related historically. I this respect, two main hypotheses have been proposed for clefts following the Germanic pattern. I referred to these as the "discontinuous-constituent" hypothesis, and the "extraposition" hypothesis. Under the first hypothesis (discussed in 5.3.2), the cleft pronoun and the cleft clause are regarded as a discontinuous relative clause construction. Under the second, a cataphoric relation between the cleft pronoun and the cleft clause is assumed. Obviously, the "discontinuous-constituent" hypothesis is not relevant to clefts of the Romance type, as the constructions following this pattern lack a cleft pronoun.

Under an extraposition analysis, the structure of a cleft of the Germanic type is in principle analogous to that of (non-oriented) clausal subject constructions with *it*-insertion in English. In fact, as was mentioned in Section 5.3.3, there are accounts in the literature (e.g. Ball 1991; 1994a) arguing that the development of (at least some types of) *it*-clefts in English might have involved a blending of clefts and such constructions. In 5.3.3, I discussed two alternative accounts on the history of the English *it*-cleft: Patten (2012a; 2012b), following a "discontinuous-constituent" approach, and Ball (1991; 1994a), whose approach shows some affinity to the "extraposition" hypothesis (though the author describes the cleft pronoun as an expletive, rather than as a pronoun cataphorically linked to the cleft).

In 5.4 I discussed the "extraposition" hypothesis with respect to the French *c'est…qu-…* cleft. Unlike in English, the "discontinuous-constituent" hypothesis does not seem to have attracted much attention in the literature to explain the development of this construction. The appeal of the "extraposition" approach in French is arguably motivated by the fact that the alternative clefting pattern in the language (the *ce qu-…c'est…* cleft) may be analyzed as a correlative (or correlative-like) construction involving a pronominal form anaphorically linked to a left-dislocated VARIABLE expression. The difference between the VARIABLE-VALUE and the VALUE-VARIABLE cleft variants in French, as far as the cleft pronoun is concerned, can be described in terms of the direction in which the pronoun is linked to the VARIABLE expression. In the former case anaphorically, and in the latter cataphorically. The particularity of the cleft pattern involving a VALUE-VARIABLE order in French lies in the anomalous form of the oriented clausal nominalization involved, as the variable expression resembles an adnominal relative clause (but one lacking an antecedent), or even a complement clause.

Independently of the possible historical explanations to account for the development of clefts of the Romance and Germanic types, it is rather clear that these constructions may be regarded to a considerable extent as constructions sui generis. As suggested on various occasions throughout this chapter, some "non-pseudo-clefts" are more problematic (with respect to their status as specificational sentences) than others. The same holds cross-linguistically.

In chapters 2 through 4, the discussion concentrated on constructions which more or less straightforwardly conform to the definition of clefts adopted in this dissertation. The focus on Germanic and Romance in discussing constructions which are more or less clearly related to specificational sentences but exhibit idiosyncrasies distinguishing them from more regular clefts should not be taken to suggest that such constructions ("non-pseudo-clefts") occur exclusively in these language groups. As mentioned earlier, comparable constructions are reported outside Germanic and Romance, and in fact outside of Indo-European. Three case studies were presented in 5.5 ranging from patterns very similar to those found in Romance (in Persian, and to a more limited extent, in Kirundi) to a highly idiosyncratic construction reported for Santa María Peñoles Mixtec.

6 Conclusions

In this dissertation, I examined cleft constructions in typologically diverse languages. The comparative concept on which the present investigation is based corresponds to the following cross-linguistically applicable definition:

A cleft is a specificational sentence in which the VARIABLE expression involves an oriented clausal nominalization.

This definition is based on two notions: specificational sentence, and oriented clausal nominalization. The understanding of these was clarified in Chapters 2 and 3 respectively. These notions may be briefly characterized as follows: A specificational sentence is a sentence consisting of two nominal expressions (plus, in many languages, a copula) in which the reference of one expression (the VARIABLE) is specified by another (the VALUE), understanding these terms in the sense used in Akamajian (1970) and much subsequent literature. An oriented clausal nominalization is a nominal expression referring to an entity described in terms of a state of affairs it is involved in.

I followed an approach to specificational sentences under which these are regarded as a particular kind of equational sentences (the other kind being identity statements or "equative" sentences). Equational sentences were characterized as sentences consisting of two referential nominal expressions in a subject-predicate-like relationship. They are to be distinguished from nominal-predicate sentences. Sentences of the latter type consist of a referential nominal expression as a subject and a non-referential nominal predicate. While referential nominal expressions are used to refer to an instance of the kind of entities in question, nominal predicates are not used to refer, but to ascribe the properties associated with such an entity to the referent of the subject expression. Though in many languages nominal-predicate sentences and equational ones may be formally identical (in some cases, allowing in isolation either interpretation), there are cross-linguistically recurrent differences distinguishing both sentence types. Dryer (2007) identifies three main potentially distinctive features: word-order flexibility, copularization patterns, and the encoding of (non-referential) nominal predicates vs. referential nominal expressions. I examined data from genealogically and areally diverse languages reported to exhibit the relevant distinctions. Contrasting copularization patterns were discussed mainly in Sections 2.2.2.1, 2.2.2.2, and 2.3.1.2 on Thai, Akan/Twi, and Modern Hebrew respectively. A discussion on the distinction between equational and nominal-predicate sentences based on the coding of the expression in predicate position is found in Sections 2.3.2.1 and 2.3.2.2 on Upper Nicola Okanagan and Rapa Nui respectively. It was argued that an assessment of the referential status of an expression in predicate position on the basis of its formal coding is problematic. A marker assumed to signal referentiality may in fact, in the relevant context, be used to mark distinctions concerning specificity or definiteness.

In Section 2.3.1, I examined linguistic phenomena reported in the literature reflecting the special referential status of the VARIABLE expression in specificational sentences. Though arguably referential, the VARIABLE expression in a specificational sentence is not fully, but "weakly" referential (more or less in the sense of Declerck 1983; 1988). The phenomena in question can be subsumed under the rubric of "pronominalization". It has been noted that in some languages, pronominal forms anaphorically linked with the VARIABLE expression in specificational sentences allow for sortal incongruence (or underspecification), which is precluded in anaphoric reference to fully referential nominal expressions. This contrast is illustrated with English, French, and German data in Section 2.3.1.1 in Chapter 2. The discussion is continued with data from Modern Hebrew in 2.3.1.2, where pronominal forms from two different series (distal and proximal

demonstratives, the former also used as personal pronouns) are used in a copula-like manner and are reported to distinguish specificational sentences from identity statements (the latter, involving fully referential pre-copular expressions, partly pattern like nominal-predicate sentences). In 2.3.1.3, I examine an arguably related phenomenon observed in Korean and Japanese. In these languages, fully referential oriented clausal nominalizations require a common noun "head" sortally congruent with the referent. In clefts, the VARIABLE expression may be "headed" by elements otherwise compatible with non-human referents only, regardless of the referent's animacy status.

As was made explicit in Section 2.1 and 2.3, the referential vs. non-referential (and the "weakly" vs. fully referential) status of an expression is not to be conflated with notions such as specificity (reference to a particular individual) or definiteness (reference to a unique entity or to one uniquely identifiable by virtue of its familiarity). There is certainly an affinity between referentiality and specificity (and definiteness), and most of the available data (especially for less-well described languages) on clefts and specificational sentences generally tend to involve VALUE expressions referring to specific (if not definite) referents. It is necessary, however, to account for cleft sentences (and specificational sentences generally) that do not involve VALUE expressions with specific (or definite) referents, and also for sentences in which the notions of specificity or definiteness with respect to the status of the VALUE expression is problematic (or irrelevant). Consider the example in (1).

(1) English (www)
[*Another thing you don't need*] *is* <u>*exclusivity*</u>*, whether it is for your store in a particular area or for your supplier.*

As discussed in 1.1.3 in Chapter 1 and throughout Chapter 3, there is a tendency in the literature on clefting to restrict the kinds of constructions described as clefts. In some accounts on cleft constructions (e.g. Lambrecht 2001; discussed in 1.1.3), a distinction is made between clefts and non-cleft (or "simple") specificational sentences on the basis of the presence or absence of a lexical noun in the VARIABLE expression. If one is involved, the sentence in question is not considered a cleft. This decision is often not made explicit but is reflected in the kind of expressions discussed. One of the consequences of such an approach is the association of clefting with pragmatic effects such as "exhaustive listing" or "exhaustive identification", which arguably may be best regarded as resulting from the fact that the constructions usually accepted as cleft clauses such as free relative clause constructions introduced by *wh*-pronouns, relative clause constructions headed by *the one* in English, and comparable constructions in other languages generally trigger definite (unique/maximal) readings. Exhaustivity implicatures may even be precluded in clefts, as is arguably the case in example (1) above.

It was pointed out in Chapter 3 (see especially Section 3.4) that there is considerable cross-linguistic variation in the patterns involved in the composition of oriented clausal nominalizations. In many languages the use of lexical nouns is required to form expressions describing a referent in terms of a state of affairs it is involved in. Under a restrictive approach to clefting, one might consider only relative clause constructions headed by nouns with very general meanings and which, in many cases, do not exhibit sortal congruence with the referents in question (e.g. Korean *kes*, Rapa Nui *me're,* or Hausa *abu/abin*, all with the basic meaning 'thing'). Such a position would allow the presence of items (also) used as common nouns as long as their status in the relevant constructions might be comparable to that of a marker of subordination. I reject this position, however. To be clear, in the approach I adopt, a sentence such as that illustrated in example (2) qualifies as a cleft.

(2) English (www)
I took the el train to work and it was eerie as hell. There was no one else in the car I was in but a cop. There were cops at every station too. [*The train I took*] *was* <u>*the one that ends its line at O'Hare airport*</u> *but they were stopping it at the second to last stop (which was where I would stop) and emptying the few of us who were on it*.

In the cleft sentence illustrated in (2) the sortal information concerning the referent (what kind of thing it is) is provided in the VARIABLE expression (i.e. the expression in square brackets) and not in the VALUE expression (underlined). Now, describing a sentence such as the one illustrated in (2) as a cleft is clearly incompatible with accounts in which cleft sentences are viewed as constructions possibly related to equational/specificational sentences but as being fundamentally different from these. Such an approach can be found in Lambrecht's (2001) highly influential account of cleft constructions (discussed in 1.1.3 in Chapter 1), where the author insists on the semantically empty nature of information-structure manipulating elements involved in cleft sentences. In this dissertation I adopted an approach under which clefts are viewed first and foremost as specificational sentences. Under this approach, the presence (or absence) of material providing sortal information about the referent in the VARIABLE expression is not definitory of a cleft sentence. In the account of cross-linguistic variation in nominalization patterns in Chapter 3, I avoided a dichotomy of "headless" vs. "headed" relative clause constructions (or a trichotomy expanding the classification with "light headed" ones) in favor of an (incomplete) classification of nominalization strategies, including what I called the noun anchoring/support strategy.[109] I follow the idea argued for in Dryer (2004) that the use of lexical nouns in nominal expressions (though possibly required in some nominal expressions in some languages) may be viewed as a device to include sortal information about a referent (which may be obviated if this information is recoverable, irrelevant, or unknown). Sortal information about the referent of a nominal expression does not necessarily involve the use of lexical nouns. To a more limited degree, such information is (or may be) encoded in pronominal forms, nominal classifiers, and articles.

Now, the idea that the VARIABLE expression in a cleft sentence may encode sortal information about its referent goes against the grain of an intuition reflected in the literature—to judge from the kinds of constructions (not) discussed in most accounts—that the cleft clause should encode exclusively information concerning a state of affairs, which is backgrounded in order to highlight a participant involved in it. My position is the following: An oriented clausal nominalization describes a referent in terms of a state of affairs in which it is involved but I do not take this to preclude that an oriented clausal nominalization may encode information about a referent beyond the description of a state of affairs it is involved in. On the contrary, I argue that one might as well take such information as evidence for the oriented-nominal status of the expressions in question. It is precisely the equational-specificational relationship between the cleft clause and the clefted constituent (which presupposes the nominal status of the cleft clause), that distinguishes clefts from other sentence types in which a constituent is singled out against the rest of the sentence.

In Chapter 4, I discussed word order patterns in cleft sentences drawing from accounts on individual languages. Section 4.1 may be regarded as a continuation of Section 2.2 in Chapter 2, examining more closely the word order alternations possible in clefts (and specificational sentences generally) in contrast to those in nominal-predicate sentences. The data suggest a greater word order flexibility in specificational

[109] For an account in a similar spirit, see Epps (2012), who describes nominalization patterns in the Naduhup (Makú) language Hup as involving a continuum between "headless" and "headed" relative clause constructions. Unfortunately, her account does not contain much information concerning clefts in this language.

sentences, permitting in most cases VARIABLE-VALUE and VALUE-VARIABLE configurations. Word order flexibility may result from the possible inversion of terms flanking an equative copula (as in Akan/Twi, discussed in 4.2.1), which is precluded in nominal-predicate sentences. A contrast between possible word order patterns in specificational and nominal-predicate sentences may also result from restrictions against non-referential expressions sentence initially, as in Hausa or Korean (discussed in 4.1.3 and 4.1.4 respectively). In Amharic (Section 4.1.5), a language with basic verb-final (SOV) word order, two different configurations are reported. The VARIABLE-VALUE configuration follows the basic word order in the language, placing the VARIABLE expression in subject position and the VALUE expression immediately preceding the verbal copula. In the VALUE-VARIABLE configuration, the VARIABLE expression follows the copula, possibly outside the core clause. Section 4.2 examines languages that require the placement of the VARIABLE expression in clefts (though not necessarily in simple specificational sentences) outside the core clause and involve a resumptive pronominal element. The examination of the data in this section anticipates the discussion in Chapter 5.

Chapter 5 is dedicated to the discussion of constructions frequently referred to as "clefts" (proper), as opposed to "pseudo-clefts". While the latter may be straightforwardly described in the terms proposed in this dissertation, corresponding to "plain" clefts in the sense of Creissels (2021), the former may present some difficulties with respect to the definition of clefts as specificational sentences with a VARIABLE expression involving an oriented clausal nominalization. The discussion was centered on two of these constructions, the English *it*-cleft and the French *c'est…qu-…* cleft. In these constructions, the cleft clause (i.e. the VARIABLE expression) generally does not resemble a nominal expression, but rather an extraposed adnominal relative clause whose status as a term in an equational sentence is not obvious. Additionally, given the tendency to mark the cleft clause with an element which may be used as a general subordinator (*that* in English and *que* in French, as well as cognates of the latter in other Romance languages), the status of the subordinate clause as a participant-oriented one becomes uncertain. This is especially the case in constructions involving the "clefting" of non-argument expressions, some of which may be better described as complex sentences featuring (non-oriented) extraposed clausal subjects. As argued in Section 5.3 and 5.5, the *it*-cleft in English and the French *c'est…qu-…* cleft can be regarded as the outcome of a historical development giving rise to idiosyncratic constructions distinct from "plain" clefts. To the extent that such constructions (and similar constructions in other languages) can be analyzed as variants of specificational sentences involving oriented clausal nominalizations as VARIABLE expressions, it seems adequate to describe them as clefts. From a cross-linguistic perspective, however, it is preferable to take less idiosyncratic constructions as canonical instances of cleft constructions.

7 References

Abdoulaye, Mahamane L. 2007. Profiling and identification in Hausa. *Journal of Pragmatics* 39. 232-269.

Abdoulaye, Mahamane L. & Barmou, Salifou & Souley Bida, Saoudé. 2020. The one-term predication for deictic identification in Hausa. *Studies in African Linguistic*s, 49(2). 187-202.

Adamou, Evangelina & Costaouec, Denis. 2010. Connective constructions in the world's languages: a functionalist approach. *La Linguisitque* 46. 43-80.

Adelaar, K. Alexander. 1992. *Proto Malayic: The reconstruction of its phonology and parts of its lexicon and morphology*. Canberra: Dept. of Linguistics, Research School of Pacific Studies, the Australian National University.

Akamajian, Adrian. 1970. *Aspects of the grammar of focus in English*. Cambridge: Massachusetts Institute of Technology. (Doctoral dissertation.)

Álvarez González, Albert. 2012. Relative clauses and nominalizations in Yaqui. In Comrie, Bernard & Estrada-Fernández, Zarina (eds), *Relative clauses in languages of the Americas: A typological overview*, 67-96. Amsterdam: Benjamins.

Álvarez González, Albert. 2016. The evolution of grammatical nominalizations in cahita languages. In Chamoreau, Claudine & Estrada-Fernández, Zarina (eds), *Finiteness and nominalization*, 107-140. Amsterdam: Benjamins.

Amfo, Nana Aba Appiah. 2010. Noun phrase conjunction in Akan; The grammaticalization path. *Pragmatics* 20(1). 27-41.

Amuzu, Evershed Kwasi. 2005. *Ewe-English codeswitching; A case of composite rather than classic codeswitching*. Canberra: Australian National University. (Doctoral dissertation.)

Apothéloz, Denis. 2012. Pseudo-clivées et constructions apparentées. In Groupe de Fribourg (eds), *Grammaire de la période*, 202-232. Frankfurt am Main et al.: Peter Lang.

Ball, Catherine N. 1991. *The historical development of the it-cleft*. Philadelphia: University of Pennsylvania. (Doctoral dissertation.)

Ball, Catherine N. 1994a. The origins of the informative-presupposition it-cleft. *Journal of Pragmatic*s 22. 603–628.

Ball, Catherine N. 1994b. Relative pronouns in it-clefts: The last seven centuries. *Language Variation and Change* 6. 179-200.

Balon, Laurent & Larrivée, Pierre. 2014. L'ancien francais n'est déjà plus une langue à sujet nul - nouveau témoignage des textes légaux. *Journal of French Studies* 26(2). 1-17. doi:10.1017/S0959269514000222

Banti, Giorgio. 2013. Clefts, epexegetic focus constructions, and information in Classical and Koine Greek. In Mancini, Marco & Lorenzetti, Luca (eds), *Le lingue del Mediterraneo antico; Culture, mutamenti, contatti*, 37-68. Rome: Carocci.

Bentivoglio, Paola & De Stefano, Luciana & Sedano, Mercedes. 1999). El uso del que galicado en el español actual. In Rojas Mayer, Elena M. (ed.), *Actas del VIII Congreso internacional de la Asociación de*

Lingüística y Filología de la América Latina, 104-111. San Miguel de Tucumán: Universidad Nacional de Tucumán.

Bhatt, Rajesh. 2015. Relative clauses and correlatives. In Kiss, Tibor & Alexiadou, Artemis (eds), *Syntax: Theory and analysis*, vol. 1, 708–749. Berlin: De Gruyter.

Blanche-Benveniste, Claire. 2002. Auxiliaires et degrés de verbalité. *Syntaxe et sémantique* 3. 75-97.

Boadi, Lawrence Addai. 1974. Focus Marking in Akan. *Linguistics* 140. 5–57.

Bolinger, Dwight. 1972. A Look at Equations and Cleft Sentences. In Firchow, Evelyn S. (ed.) *Studies for Einar Haugen*, 96–114. The Hague: Mouton de Gruyter.

Bombi, Carla. 2018. Definiteness in Akan: Familiarity and uniqueness revisited. *Proceedings of SALT* 28. 141–160. doi:10.3765/salt.v28i0.4406

Bombi, Carla & Grubic, Mira & Renans, Agata & Duah, Reginald Akuoko. 2019. The semantics of the (so-called) clausal determiner *nó* in Akan (Kwa). *Proceedings of Sinn und Bedeutung* 23(1). 181–199.

Calude, Andrea S. 2008. Clefting and extraposition in English. *ICAME Journal* 32. 7–34.

Campbell, Akua. 2020. Relative Clauses as Grammatical Nominalizations: Evidence from Akan. *SKASE Journal of Theoretical Linguistics* 17 (2). 23-58.

Carlier, Anne. 2004. Ce sont des Anglais: un accord avec l'attribut? (Première partie). *L'Information grammaticale* 103. 13-18.

Carlier, Anne. 2005. Ce sont des Anglais; un accord avec l'attribut? (Seconde partie). *L'Information grammaticale* 104. 4-14.

Carochi, Horacio. 1892[1645]. *Arte de la lengua mexicana con la declaración de los adverbios della*. Mexico City: Museo Nacional.

Chafe, Wallace L. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Li, Charles N. (ed.), *Subject and topic*, 25-55. New York: Academic Press.

Chung, Han-Byul. 2016. *On the syntax/semantics of Korean nominal particles*. New York: The City University of New York. (Doctoral dissertation.)

Citko, Barbara. 2004. On headed, headless, and light-headed relatives. *Natural Language and Linguistic Theory* 22. 95-126.

Clech-Darbon, Anne & Rebuschi, Georges & Rialland, Annie. 1999. Are there cleft sentences in French? In Rebuschi, Georges & Tuller, Laurice (eds), *The grammar of focus*, 83-119. Amsterdam & Philadelphia: Benjamins.

Collins, Peter. 1991. *Cleft and Pseudo-cleft Constructions in English*. London/New York: Routledge.

Constantino, Ernesto. 1965. The sentence patterns of twenty-six Philippine languages. *Lingua* 15. 71–124.

Coyaud, Maurice. 1975. Emphase, nominalisations relatives (Typologie). *La Linguistique* 11(2). 39-59.

Craig, William Earl. 1971. *Investigation of syntactic-semantic relationships in the selected writing of students in grades 4-12*. Ohio State University. (Doctoral dissertation.)

Creissels, Denis. 2014. Existential predication in typological perspective. (Unpublished ms. / revised version of a paper presented at the 46th Annual Meeting of the Societas Linguistica Europaea, Split, 18-21 September 2013.)

Creissels, Denis. 2021. Remarks on the grammaticalization of identificational clefts. *Faits de Langues* 52(1). 13-31.

Croft, William. 1991. *Syntactic categories and grammatical relations*. Chicago: University of Chicago Press.

Cruschina, Silvio. 2018. Setting the boundaries: Presentational *ci*-sentences in Italian. *Belgian Journal of Linguistics* 32. 53-85. https://doi.org/10.1075/bjl.00016.cru

Custis, Tonya. 2004. *Word order variation in Tongan: A syntactic analysis*. University of Minnesota. (Doctoral dissertation.)

Davidse, Kristin & Van Praet, Wout. 2019. Rethinking predicative clauses with indefinite predicate and specificational clauses with indefinite variable: A cognitive-functional account. *Leuven Working Papers in Linguistics* 6(38). 1–36.

De Cat, Cécile. *French dislocation; Interpretation, syntax, acquisition*. Oxford: University Press.

Déchaine, Rose-Marie. 2002. Decomposing focus: evidence from Yorùbá. (Paper presented at the Triggers for Movement Workshop, Tilburg University, October 24-26.)

Declerck, Renaat. 1983. 'It is Mr. Y' or 'He is Mr. Y'? *Lingua* 59. 209-246.

Declerck, Renaat. 1988. *Studies on Copular Sentences, Clefts and Pseudo-clefts*. Leuven: Leuven University Press; Dordrecht: Foris.

Delahunty, Gerald P. 1984. The analysis of English cleft sentences. *Linguistic analysis* 13(2). 63-113.

Den Dikken, Marcel. 2017. In Everaert, Martin & van Riemsdijk, Henk (eds), *The Wiley Blackweell compoanion to syntax* (second edition), 292-409. Hoboken NJ: John Wiley & Sons.

Di Tullio, Ángela & Kailuweit, Rolf. 2012. Las oraciones enfáticas copulativas enfáticas del español y sus varias realizaciones. In Lebsanft, Franz & Mihatsch, Wiltrud & Polzin-Haumann, Claudia (eds), *El español, ¿desde las variedades a la lengua pluricéntrica?* 141-160. Madrid: Iberoamericana / Frankfurt: Vervuert.

Diessel, Holger. 1999. *Demonstratives; Form, function, and grammaticalization*. Amsterdam & Philadelphia: Benjamins.

Dik, Simon C. 1980. *Studies in functional grammar*. London: Academic Press.

van Dommelen, Wim A. & Beermann, Dorothee. 2019. Tonal properties of the Akan particle *na*. In Calhoun, Sasha & Escudero, Paola & Tabain, Marija & Warren, Paul (eds), *Proceedings of the 19th International Congress of Phonetic Sciences, Melbourne, Australia 2019*. Canberra: Australasian Speech Science and Technology Association Inc.

Dryer, Matthew S. 2004. Noun phrases without nouns. *Functions in Language* 11(1). 43-76.

Dryer, Matthew S. 2007. Clause types. In Shopen, Timothy (ed.), *Language typology and syntactic description* (second edition) *Volume 1: Clause structure*, 224-275. Cambridge: University Press.

Dryer, Matthew S. 2013. Order of relative clause and noun. In Dryer, Matthew S. & Haspelmath, Martin (eds), *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/90, Accessed on 2021-05-27.)

Dufter, Andreas. 2008. On explaining the rise of c'est-clefts in French. In Detges, Ulrich & Waltereit, Richard (eds), *The paradox of grammatical change: Perspectives from Romance, 31-56.* Amsterdam & Philadelphia: Benjamins.

Dufter, Andreas. 2010. El que galicado: distribución y descripción gramatical. In Sinner, Carsten & Zamorano Aguilar, Alfonso (eds), *La excepción en la gramática española, 255-280.* Madrid: Iberoamericana / Frakfurt am Main: Vervuert.

Ellis, Jeffrey & Boadi, Lawrence. 1969. To 'Be' in Twi. In Verhaar, J.W.M. (ed.), *The Verb 'Be' and its Synonyms*, Part 4, 1–71. Dordrecht: Reidel.

Epps, Patience. 2012. Between headed and headless relative clauses. In Comrie, Bernard & Estrada-Fernández, Zarina (eds), *Relative clauses in languages of the Americas: A typological overview, 67-96.* Amsterdam: Benjamins.

Erguvanlı, Eser Emine. 1984. *The function of word order in Turkish grammar*. London: University of California Press.

Faghiri, Pegah & Samvelian, Pollet. 2021. A corpus-based description of cleft constructions in Persian. *Faits de langue*s 52(1). 183-206.

Falk, Yehuda N. 2007. *Hebrew present-tense copular constructions: The view from parallel syntax*. Jerusalem: Hebrew University. (Unpublished ms.) <https://pluto.mscc.huji.ac.il/~msyfalk/StrucFunc/Copula.pdf> (last accessed 22.10.2020)

Fiedler, Ines & Schwarz, Anne. 2005. Out-of-focus encoding in Gur and Kwa. In Ishihara, Shinichiro & Schmitz, Michaela & Schwarz, Anne (eds), *Interdisciplinary Studies on Information Structure* 3, 111-142. Potsdam: Potsdam University

Francis, Elaine J. 2010. Grammatical weight and relative clause extraposition in English. *Cognitive linguistics* 21(1). 35-74.

Geist, Ljudmila. 2014. Bare predicate nominals in German. In Veselovská, Ludmila & Markéta, Janebová (eds), *Complex visibles out there*; *Proceedings of the Olomouc Linguistics Colloquium 2014: Language Use and Linguistic Structure*, 83-106. Olomouc: Polacký University.

Gheorge, Mihaela. 2017. (Pseudo)cleft constructions in Old Romanian. *Diacronia* 6. 1-10.

Givón, Talmy. 1979. *On understanding grammar*. Orlando: Academic Press.

Givón, Talmy. 2001. *Syntax*, vol. 2. Amsterdam: J. Benjamins

Goria, Eugenio. 2013. Towards a taxonomy of Latin cleft sentences. *Journal of Latin Linguistics* 22(2): 147-172.

Green, Melanie. 2007. *Focus in Hausa*. Oxford: Blackwell.

Greenberg, Yael. 2008. Predication and equation in Hebrew (nonpseudocleft) copular sentences. In Armon-Lotem, Sharon & Danon, Gabi & Rothstein, Susan D. (eds), *Current Issues in Generative Hebrew Linguistics*, 161-196. Amsterdam: Benjamins.

Grønvik, Ottar. 1991. "Utbrytning" i eldre norrøn diktning og i andre gammelgermanske språk. *Arkiv för nordisk filologi* 106. 1-56.

Grubic, Mira & Renans, Agata & Duah, Reginald Akuoko. 2019. Focus, exhaustivity and existence in Akan, Ga and Ngamo. *Linguistics* 57(1). 221-268.

Guadalupe Joaquina, Amadeo. 2007. *Análisis morfosintáctico de la estructura oracional del mixteco de Yoloxóchitl, Guerrero*. Mexico City: CIESAS. (Masters thesis.)

Guasch Leguizamón, Jorge. 1951. *Galicismos aceptados, aceptables y vitandos*. Buenos Aires: Kapeluz.

Guerrero, Lillián. 2012. On relative clauses and related constructions in Yaqui. In Comrie, Bernard & Estrada-Fernández, Zarina (eds), *Relative clauses in languages of the Americas: A typological overview*, 97–126. Amsterdam: Benjamins.

Guesser, Simone & Trianon, Rafael. 2020. Clivadas, pseudoclivadas, português brasileiro e abordagem cartográfica. Medeiros, Paulo Jr. & Guesser, Simone & Lunguinho, Marcus Vinicius & Guerra Vicente, Helena (eds), *Relativização e clivagem no pb: sintaxe, aquisição, diacronia e experimentação*, 9-48. Campinas: Pontes.

Güldemann, Tom. 2016. Maximal backgrounding = focus without (necessary) focus encoding. *Studies in Language* 40(3). 551-590.

Gutiérrez-Bravo, Rodrigo. 2013. Free relative clauses in Yucatec Maya. *STUF Language Typology and Universals* 66(1). 22-39.

Gutiérrez-Bravo, Rodrigo. 2017. Clefts and focus in Yucatec Maya. *Cuadernos de Lingüística de El Colegio de México* 4(1). 5–47.

Halliday, M.A.K. 1967b. Notes on transitivity and theme in English: Part 2. *Journal of Linguistics* 3(2). 199–244.

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

Halliday, M.A.K. & Matthiessen, M.I.M. 2014. *Halliday's introduction to functional grammar*. London & New York: Routledge.

Harada, Masashi. 2018. The Existence of Pseudoclefts in Japanese. *Buckeye East Asian Linguistics* 3. 43-52.

Harries-Delisle, Helga. 1973. Contrastive Emphasis and Cleft Sentences. *Working Papers on Language Universals* 12, 85-144. Stanford: University of California.

Harris, Alice & Campbell, Lyle. 1995. *Historical syntax in cross-linguistic perspective*. Cambridge: University Press.

Hartmann, Jutta M. & Hegedűs, Veronika & Surányi, Balázs. 2013. Pseudoclefts in Hungarian. In Brandtler, Johan & Molnár, Valéria & Platzack, Christer (eds), *Approaches to Hungarian. Volume 13: Papers from the 2011 Lund conference,* 67-96. Amsterdam & Philadelphia: Benjamins.

Hartmann, Jutta M. & Heycock, Caroline. 2020. (Morpho)syntactic Variation in Agreement: Specificational Copular Clauses Across Germanic. *Frontiers in Psychology* 10:2994. doi:10.3389/fpsyg.2019.02994

Hartmann Katharina & Veenstra, Tonjes (eds). 2013. *Cleft structures*. Amsterdam & Philadelphia: Benjamins.

Hedberg, N., & Potter, David. 2010. Equative and predicational copulas in Thai. *Annual Meeting of the Berkeley Linguistics Society* 36(1). 144-157. https://doi.org/10.3765/bls.v36i1.3908

Heine, Bernd & Reh, Mechthild. 1984. *Grammaticalization and Reanalysis in African Languages*. Hamburg: Helmut Buske.

Heller, Daphna. 2005. *Identity and Information. Semantic and Pragmatic Aspects of Specificational Sentences. Rutgers University*. (Doctoral dissertation.)

Hendery, Rachel. 2012. *Relative clauses in space and time: a case study in the methods of linguistic typology*. Amsterdam & Philadelphia: Benjamins.

Hengeveld, Kees. 1992. *Non-verbal predication: Theory, typology, diachrony*. Berlin: Mouton de Gruyter.

Higgins, Francis R. 1979. *The pseudo-cleft construction in English*. London & New York: Routledge.

Himmelmann, Nikolaus P. 1997. *Deiktikon, Artikel, Nominalphrase: Zur Emergenz syntaktischer Struktur*. Tübingen: Hans Niemeyer Verlag.

Himmelmann, Nikolaus P. 1999. The lack of zero anaphora and incipient person marking in Tagalog. *Oceanic Linguistics* 38. 231-269.

Himmelmann, Nikolaus P. 2008. Lexical Categories and Voice in Tagalog. In Austin, Peter K. & Musgrave, Simon (eds), *Voice and Grammatical Relations in Austronesian Languages*, 247–293. Stanford: CSLI Publications.

Himmelmann. Nikolaus P. 2015. Notes on 'noun phrase structure' in Tagalog. In Fleischhauer, Jens & Latrouite, Anja & Osswald, Rainer (ed.), *Explorations of the Syntax-Semantics Interface*, 315-337. Düsseldorf: University Press.

Hiraiwa, Ken. 2003. Relativization in Buli. In Akanlig-Pare, George & Kenstowicz, Michael (eds), *Studies in Buli grammar* (MIT Working Papers on Endangered and Less Familiar Languages Vol. 4), 45-84. Cambridge: MITWPL.

Hiraiwa, Ken. 2005. Predicate clefts in Buli: Categories and phases. *Linguistic Analysis* 32. 544-583.

Hoffmann, Roland. 2016. Latin cleft constructions, synchronically, diachronically and typologically reconsidered. *Pallas* 102. 201-2010.

Hole, Daniel & Zimmermann, Malte. 2013. Cleft partitionings in Japanese, Burmese & Chinese. In Hartmann Katharina & Veenstra, Tonjes (eds), *Cleft structures*, 285-317. Amsterdam & Philadelphia: Benjamins.

Hoppel, Paul J. 2004. Grammatical constructions and their discourse origins: Prototype or family resemblance? In Pütz, Martin & Neimeier, Suzanne & Driven, René (eds), *Applied Cognitive Linguistics 1: Theory and Language Acquisition*, 101–129. Berlin & New York: Mouton de Gruyter.

Huddleston, Rodney & Pullum, Geoffrey K. 2012. *The Cambridge grammar of the English language*. Cambridge et al.: University Press.

Iatridou, Sabine & Varlokosta, Spyridoula. 1998. Pseudoclefts Crosslinguistically. *Natural Language Semantics* 6. 3–28.

Iwasaki, Shoichi & Ingkaphirom, Preeya. 2009. *A Reference Grammar of Thai*. Cambridge: University Press.

Jaggar, Philip J. 2001. *Hausa*. Amsterdam & Philadelphia: Benjamins.

Jaratjarungkiat, Sureenate. 2019. Is the Copula /pen/ in Thai Meaningless? *Journal of Mekong Societies* 15(2). 1-17.

Jespersen, Otto. 1949[1927]. *A Modern English grammar on historical principles. Part III Syntax, second volume*. Copenhagen: Einar Munskgaard; London: Allen & Unwin.

Jespersen, Otto. 1969[1937]. *Analytic Syntax*. Chicago: University Press

Jo, Jung-Min. 2007. Word order variations in Korean copular sentences. *The Linguistic Association of Korea Journal* 15(3). 209-238.

Jochimsen, Paul 1907. *Beiträge zur Geschichte der deiktischen Hervorhebung eines einzelnen Satzteiles, bezw. eines Satzes mittelst c'est (…) que (qui)*. University of Kiel. (Doctoral dissertation.)

Jokinen, Ulla. 1986. Le rôle du « ce neutre » dans les relatives du moyen français. In Suomela-Härmä; Elina & Välikangas, Olli (eds), *Actes 9e Congrès des Romanistes Scandinaves (Helsinki. 13-17 août 1984)*, 165-179. Helsinki: Société Néophilologique.

Jones, Susie. 2006. Focus in Yorùbá: a semantic/pragmatic account. *ZAS Papers in Linguistics* 46. 143-160.

Junghare, Indira Y. 1996. Relative clauses in Hindi: a functional approach. In *Papers of the Mid-America Linguistics conference 1996*, 222-232. University of Kansas.

Kaltenböck, Gunther. 1999. Which it is it? Some remarks on anticipatory *it*. *Vienna English working papers* 8(2). 48-71.

Kandybowicz, Jason. 2009. Demystifying the predicate cleft construction. In Mohanty, Rajat & Mythili, Menon (eds) *Proceedings of GLOW in Asia VII 2009, 139-157*. Hyderabad: English and Foreign Languages University Press.

Kandybowicz, Jason. 2004. Kandybowicz, Jason. 2004. Predicate clefts, derivations, and universal grammar. In Akinlabi, Akin & Adesola, Oluseye (eds), *Proceedings of the fourth world congress of African Linguistics, New Brunswick 2003*., 211-223. Cologne: Rüdiger Köppe.

Karssenberg, Lena & Lahousse, Karen & Lamiroy, Béatrice & Marzo, Stefania & Drobnjakovic, Ana. 2018. Non-prototypical clefts; Formal, semantic and information-structural properties. *Belgian Journal of Linguistics* 32. 1-20.

Kaufman, Daniel. 2009. Austronesian Nominalism and its consequences: A Tagalog case study. *Theoretical Linguistics* 35(1). 1–49.

Kaufman, Daniel. 2017. Lexical Category and Alignment in Austronesian. In Coon, Jessica & Massam, Diane & Travis, Lisa D. (eds), *The Oxford Handbook of Ergativity*, 589–630. Oxford: Oxford University Press.

Kaufman, Daniel. 2018. Austronesian predication and the emergence of biclausal clefts in Indonesian languages. In Riesberg, Sonja & Shiohara, Asako & Utsumi, Atsuko (eds.), *Perspectives on information structure in Austronesian languages*, 207-245. Berlin: Language Science Press.

Kaznelson, Sigmund D. 1974. *Sprachtypologie und Sprachdenken*. Munich: Huber.

Keenan, Edward L. & Comrie, Bernard. 1977. Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8. 63-99.

Keizer, Evelien M. 1992. *Reference, predication and (in)definiteness in Functional Grammar; A functional approach to English copular sentences*. University of Amsterdam. (Doctoral dissertation.)

Khormai, Ali-Reza & Shahbaz, Afsaneh. 2010. Jomalāt-e gosaste: do maqule-ye motemāyez yā maquleyei vāhed? [Cleft Sentences: Two Distinct Categories or One Single Category?]. *Dastur* [Grammar] 5. 51-81.

Kieviet, Paulus. 2017. *A grammar of Rapa Nui*. Berlin: Language Science Press.

Kim, Jong-Bok. 2015. Copular constructions and asymmetries in the specificational pseudocleft constructions in Korean. *Language and Linguistics* 17(1). 89-112

Kim, Jong-Bok & Sells, Peter. 2013. Interactions between (pseudo-) cleft and copular constructions in Korean. *Linguistic Research* 30(1). 93-139.

Kiss, Katalin É. 1992. Az egyszerű mondat szerkezete. In Kiefer, Ferec (ed.), *Strukturális magyar nyelvtan I: Mondattan*, 79 -178. Budapest: Akadémiai Kiadó.

Kiss, Katalin É. 2002. *The Syntax of Hungarian*. Cambridge: Cambridge University Press.

Komagata, Nobo. 1996. *Pseudoclefts in Japanese*. University of Pennsylvania. (Unpublished ms.)

Konrad, Ingrid. 2019. *Quand la nominalisation envahit la subordination; étude sur 'ce que' dans les propositions subordonnées en français, selon une approche comparative, diachronique et expérimentale*. Paris: Université Paris Cité. (Doctoral dissertation.) <https://tel.archives-ouvertes.fr/tel-02971026> (last accessed 22.10.2020)

Kornfilt, Jaklin. 1997. *Turkish*. London & New York: Routledge.

Kramer, Ruth. & Eilam, Aviad. (2012). Verb-medial word orders in Amharic. *Journal of Afroasiatic Languages* 5. 75–104.

Kroeber, Paul D. 1999. *The Salish language family: Reconstructing syntax*. Lincoln & London: University of Nebaraska Press.

Kuno, Susumo & Wongkhomthong, Preya. 1981. Characterizational and identificational sentences in Thai. *Studies in Language* 5. 65-109.

Kuno, Susumo. 1970. Some properties of nonreferential noun phrases. In Jakobson, Roman & Kawamoto, Shigeo (eds), *Studies in general and oriental linguistics, presented to Shiro Hattori on the occasion of his sixtieth birthday*, 348-73. Tokyo: TEC Company.

Kunstmann, Pierre. 1990. *Le relatif-interrogatif en ancien francais*. Geneva: Librairie Droz.

Kuroda, Shigeyuki. 1992. *Japanese Syntax and Semantics*. Dordrecht: Kluwer

Lahousse, Karen. 2007. Specificational sentences and word order in Romance: A functional analysis. *Folia Linguistica* 41(3–4). 357–404.

Lahousse, Karen & Lamiroy, Béatrice. 2012. Word order in French, Spanish and Italian: A grammaticalization account. *Folia Linguistica* 46(2). 387–415.

Lambrecht, Knud. 2001. A framework for the analysis of cleft constructions. *Linguistics* 39(3). 463–516.

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

Langacker, Ronald W. 2000. *Grammar and Conceptualization*. Berlin & Boston: De Gruyter Mouton.

Launey, Michel. 1986. *Catégories et opérations dans la grammaire nahuatl*. Paris: Université de la Sorbonne Paris IV. (Doctoral dissertation.)

Lauwers, Peter & Tobback, Els. Predicational and (quasi-)specificational constructions with fronted adjectival predicates in French. *Transactions of the Philological Society* 118(3). 468-499.

Lazard, Gilbert & Peltzer, Louise. 2000. *Structure de la langue tahitienne*. Paris & Louvain & Sterling: Peeters.

Le Roux de Lincy, Antoine (ed.). 1841. *Les Quatre livres des rois, traduits en francais du XIIIe siècle, suivis d'un fragment de moralités sur Job et d'un choix de sermons de Saint Bernard*. Paris: Imprimerie Royale.

Lehmann, Christian. 1984. *Der Relativsatz. Typologie seiner Strukturen; Theorie seiner Funktionen, Kompendium seiner Grammatik*. Tübingen: Narr.

Lehmann, Christian. 1988. Towards a Typology of clause linkage. In Haiman, John & Thompson, Sandra A. (eds), *Clause combining in grammar and discourse*, 181-226. Amsterdam & Philadelphia: Benjamins.

Lehmann, Christian. 2020a. Nominalization. https://www.christianlehmann.eu/ling/lg_system/func_domains/specification/nominal_index.php?open=../nexion/nominalization.inc (last accessed 22.10.2020)

Lehmann, Christian. 2020b. Orientation. https://www.christianlehmann.eu/ling/lg_system/func_domains/specification/nominal_index.php?open=orientation (last accessed 20.03.2021)

Lehmann, Christian. n.d. Clefting. https://www.christianlehmann.eu/ling/lg_system/grammar/nexion/complex_index.php?open=clefting (last accessed 22.10.2020)

Li, Charles N. & Thompson, Sandra A. 1989[1981]. *Mandarin Chinese; A functional reference grammar*. Berkley et al.: University of California Press.

Littell, Patrick William. 2012. *Focus, predication, and polarity in Kwak'wala*. Vancouver: University of British Columbia. (Doctoral dissertation.)

Löfstedt, Bengt. 1966. Die Konstruktion "c'est lui qui l'a fait" im Lateinischen. *Indogermanische Forschungen* 71. 253–277.

Luo, Cheng. 2009. *Cleftability in language*. Wuhan: University Press.

Lyon, John. 2013. *Predication and equation in Okanagan Salish: The syntax and semantics of determiner phrases*. Vancouver: University of British Columbia. (Doctoral Dissertation.)

Mahootian, Sharzad. 1997. *Persian*. New York: Routledge.

Matić, Dejan. 2007. Spezifikation und Kongruenz: Die Syntax der Kopulasätze im Ungarischen und Englischen. In Geist, Ljudmila & Rothstein, Björn (eds), *Kopulaverben und Kopulasätze; Intersprachliche und intrasprachliche Aspekte, 21-45.* Tübingen: Max Niemeyer.

Matić, Dejan & van Putten, Saskia & Hammond, Jeremy. 2016. Integrated and non-integrated Left-dislocation: A comparative study of LD in Avatime, Tundra Yukaghir & Whitesands. In Fleischhauer, Jens & Latrouite, Anja & Osswald, Rainer (eds) *Explorations of the the Syntax-Semantics Interface*, 339–367. Düsseldorf: University Press.

Mikkelsen, Line. 2005. *Copular clauses; Specification, predication and equation*. Amsterdam & Philadelphia: Benjamins.

Mikulskas, Rolandas. 2016. *Copular Constructions in Lithuanian*. Amsterdam & Philadelphia: Benjamins.

Mitchell, Bruce. 1985. *Old English Syntax, Vol. 1: Concord, the parts of speech and the sentence*. Oxford: Clarendon Press.

Mitchell, Bruce. 1985. *Old English Syntax, Vol. 2: Subordination, independent elements and element order*. Oxford: Clarendon Press.

Mithun, Marianne. 2012. Questionable relatives. In Comrie, Bernard & Estrada-Fernández, Zarina (eds), *Relative clauses in languages of the Americas: A typological overview*, 269-300. Amsterdam: Benjamins.

Moezzipour, Farhad. 2010. *Cleft Constructions in Persian: Role and Reference Grammar Analysis*. Teheran: University of Teheran. (Master's thesis.)

Moltmann, Friederike. 2013. Identificational sentences. *Natural Language Semantics* 21(3). 43-77.

Moutaouakil, Ahmed. 1989. *Pragmatic functions in a functional grammar of Arabic*. Dorderecht & Providence: Foris.

Muller, Claude. 2003. Naissance et évolution des constructions clivées en « c'est...que... »: de la focalisation sur l'objet concret à la focalisation fonctionnelle. In Blumenthal, Peter & Tyvaert, Jean-Emmanuel (eds), *La cognition dans le temps*, 101-120. Tübingen: Niemeyer.

Murasugi, Keiko. 1991. *Noun Phrases in Japanese and English: A study in syntax, learnability and acquisition*. University of Connecticut. (Doctoral dissertation.)

Newman, Paul. 2000. *The Hausa language: An encyclopedic reference grammar.* New Haven & London: Yale University Press.

Nichols, Johanna. 2011. *Ingush Grammar*. Berkley & Los Angeles & London*:* University of California Press.

Nuhn, Patrick. 2019. Reversed ang-inversion and narrow focus marking in Tagalog. *Studies in Language* 43(2). 469-497.

Ofori, Seth Antwi. 2011. On the basic focus marker, and the basic focus sentence, in Akan (Twi). *Nordic Journal of African Studies* 20(3). 241–262.

Olmos, Andrés de. 1875[1547]. *Grammaire de la langue nahuatl ou mexicaine* [*Arte para aprender la lengua mexicana*], Siméon Rémi (transl.). Paris: Imprimerie Nationale.

O'Neil, Wayne. 1976. Clause adjunction in Old English. *General Linguistics* 17(4). 199–211.

O'Neill, Teresa. 2015. *The structure of English amalgam pseudoclefts*. New York: CUNY Graduate Center. (Unpublished ms.)

Palancar, Enrique L. 2021. Clefts vs. monoclausal constructions in Tilapa Otomí. *Faits de Langues* 52(1). 139-158.

Patten, Amanda L. 2012a. *The English it-cleft. A constructional account and a diachronic investigation*. Berlin & Boston: De Gruyter Mouton.

Patten, Amanda L. 2012b. The historical development of the it-cleft. A comparison of two different approaches. *Studies in Language* 36(3). 548–575.

Patten, Amanda L. 2016. Well-formed lists: specificational copular sentences as predicative inversion constructions. *English Language and Linguistics* 22(1). 77–99.

Pérez-Guerra, Javier. 1998. Integrating right-dislocated constituents: a study on cleaving and extraposition in the recent history of the English language. *Folia Linguistica Historica* 19. 7-25.

Perlmutter, David M. 1972. Evidence for Shadow Pronouns in French Relativization. In Peranteau, Paul M. & Levi, Judith N. & Phares, Gloria C. (eds), *The Chicago Which Hunt; Papers from the Relative Clause Festival April 13, 1972*. Chicago Linguistic Society.

Peters, P. Stanley & Emon Bach. 1968. *Pseudo-cleft sentences*. Austin: University of Texas. (Unpublished ms.)

Pierrard, Michel. 1995. Système et histoire du système: L'évolution de "ce que" introducteur de subordonées. *Langue Francaise* 107 (*Synchronie et diachronie: du discours à la grammaire*). 111-124.

Pinkster, Harm. 2012. Relative clauses in Latin: some problems of description. In Corrêa, Paula da Cunha & de Macedo, José Marcos Mariani & dos Santos, Marcos Martinho & Hasegawa, Alexandre Pinheiro (eds), *Hyperboreans: Essays in Greek and Latin poetry, philosophy, rhetoric and linguistics*, 377-393. São Paulo: Humanitas CAPES.

Prince, Ellen F. 1979. A Comparison of Wh-Clefts and it-Clefts in Discourse. *Language* 548(4). 883-906.

Puri, Vandana. 2011. The Influence of English on the history of Hindi relative clauses. *Journal of Language Contact* 4. 250–268

Quarcoo, Millicent Akosua. 2014. Copula verbs in Twi/English codeswitching. *Journal of West African Languages* 41(2). 3-14.

Ramhöj, Rickard. 2016. *On clausal subjects and extraposition in the history of English*. Gothenburg: University of Gothenburg. (Doctoral Dissertation.)

Ramírez Pérez, Elodia. 2014. *La predicación no verbal y las construcciones copulares en el tnu'u²³ dau²³ de Santa María Peñoles, Etla, Oaxaca*. Mexico City: CIESAS. (Master's thesis.)

Ramos, Teresita & Cena, Resty M. . 1990. *Modern Tagalog; Grammatical explanations and excercises for non-native speakers*. Honolulu: University of Hawaii Press.

Ramos Bierge, Stefanie. 2014. From Demonstrative to Nominalizer: The Suffix -gai in Northern Tepehuan. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 38. 432-447.

Rivero, María-Luisa. 1975. Referential properties of Spanish noun phrases. *Language* 51(1). 32-48.

Roggia, Carlo Enrico. 2012. Frasi scisse in italiano antico: alcune proposte. In Wehr, Barbara & Nicolosi, Frederic (eds), *Pragmatique historique et syntaxe* / *Historische Pragmatik und Syntax*, 193-221. Frankfurt am Main: Lang.

Ross, Haj. 1995. Defective noun phrases. *Papers of the Regional Meeting of the Chicago Linguistics Society* 31. 398–440.

Rouquier, Magali. 2014. *L'Émergence des constructions clivées, pseudo-clivées et liées en français*. Paris: Classiques Garnier.

Rouquier, Magali. 2018. Les contructions clivées. In *Encyclopédie Grammaticale du Français*. <http://encyclogram.fr>

Royal Institute of Thailand. 2013. *Photchananukrom chabap ratchabandittayasathan phoso song phan ha roy ha sip ha* [Thai dictionary: Royal Institute edition 2011]. Bangkok: Royal Institute of Thailand.

Ruangjaroon, Sugunya. 2007. Thai wh-expressions at the left edge of the clause: Contrastive and Identificational Wh-clefts. *Concentric: Studies in Linguistics* 33(2). 121-157.

Sahagún, Bernardino de. 1950-1982[ca. 1540-1569]. *Florentine Codex: General history of the things of New Spain* [*Historia general de las cosas de la Nueva España*] (12 volumes), Anderson, Arthur J.O. & Dibble, Charles (transl.). Salt Lake City: University of Utah Press.

Sandfeld, Kristian. 1977[1936]. *Syntaxe du français contemporain; Les propositions subordonnées*. Geneva: Librairie Droz.

Sasaki, Mitsuya. 2014. Pronominal copular sentences in Ixquihuacan Nahuatl. (Paper presented at the 8th Bunpō Kenkyū Workshop, the Research Institute for Languages and Cultures of Asia and Africa, December 6, 2014, University of Foreign Studies, Tokyo.)

Sasse, Hans-Jürgen. 1991. Predication and sentence constitution in universal perspective. In Zaefferer, Dietmar (ed.), *Semantic Universals and Universal Semantics*, 75-95. Berlin & Boston: De Gruyter Mouton.

Schachter, Paul & Otanes, Fe T. 1972. *Tagalog reference grammar*. Berkeley & Los Angeles: University of California Press.

Schachter, Paul. 1973. Focus and relativization. *Language* 49(1). 19-46.

Schuh, Russell G. n.d. *On the origins of the Hausa "relative" aspects and the "stabilizer"*. Los Angeles: University of California. (Unpublished ms.)

Schwarz, Anne & Fiedler, Ines. 2007. Narrative focus strategies in Gur and Kwa. In Aboh, Enoch C. & Hartmann, Katharina & Zimmermann, Malte (eds), *Focus strategies in African languages*, 267–286. Berlin: Mouton de Gruyter.

Seppänen, Aimo. 2004. The Old English relative þe. *English Language and Linguistics* 8(1). 71–102.

Seraku, T. 2011. Clefts and free-relatives in Japanese. In Bailey, Dina & Teliga, Victoria (eds), *Proceedings of the Thirty-Ninth Western Conference on Linguistics*. 221-239. Fresno: Department of Linguistics, California State University.

Seraku, Tohru. 2013. *Clefts, relatives, and language dynamics: The case of Japanese*. Oxford: University of Oxford. (Doctotral dissertation.)

Shagal, Ksenia. 2019. *Participles; A typological study*. Berlin & Boston: de Gruyter Mouton.

Shibatani, Masayoshi. 2019. What is nominalization? Towards the theoretical foundations of nominalization. In Zariquiey, Roberto & Shibatani, Masayoshi & Fleck, David (eds), *Nominalization in languages of the Americas*, 15-167. Amsterdam: Benjamins.

Shirtz, Ilona Spector. 2014. *The syntax of non-verbal predication in Modern Hebrew: Predicate nominals, pseudoclefts and clefts*. Jerusalem: Hebrew University. (Doctoral dissertation.)

Shor, Leon. 2020. Reassessing the third person pronominal "copula" in spoken Israeli Hebrew. *Linguistics* 58(6). 1807-1837.

Sichel, Ivy. 1997. Two pronominal copulas and the syntax of Hebrew nonverbal sentences. *Texas Linguistic Forum* 38. 295-306.

Siemund, Peter & Beal, Kalinda. 2011. Clefting in contact through space and time: an analysis of it-clefts in Irish English, Indian English and British English. In Hickey, Raymond (ed.), *Researching the languages of Ireland*, 243-267. Uppsala: Uppsala University Press.

Smits, Rik 1989. *Eurogrammar. The Relative and Cleft Constructions of the Germanic and Romance Languages*. Dordrecht: Foris.

Sneddon, James Neil & Adelaar, Alexander & Dwi Noverini Djenar & Ewing, Michael C. 2010. *Indonesian Reference Grammar*. London: Allen & Unwin.

Stassen, Leon. 1997. *Intransitive predication*. Oxford: Clarendon Press.

Suárez-Gómez, Cristina. 2006. *Relativization in early English (950–1250): The position of relative clauses*. Bern: Peter Lang.

Tagliamonte, Sali & Smith, Jennifer & Lawrence, Helen. 2005. No taming of the vernacular! Insights from relatives in northern Britain. *Language Variation and Change* 17. 75-112.

Takahashi, Kiyoko & Shinzato, Rumiko. 2003. On Thai copulas, khuu1 and pen1: A cognitive approach. *Proceedings of the 2nd Seoul International Conference on Discourse and Cognitive Linguistics.* 131-145.

Takizala, Alexis. 1972. Focus and relativization: The case of Kihung'an. *Studies in African Linguistics* 3(2). 259-287.

Tawilapakul, Upsorn. 2017. Exhaustivity in the pen33-cleft in Thai: A formal semantics account. *Journal of Liberal Arts* 17(2). 102-126.

Tham, Shiao Wei. 2008. The semantic category of the subject NP in Mandarin specificational copular sentences. *Journal of East Asian Linguistics* 17. 61-82.

Togeby, Knud. 1983. *Grammaire française*. Copenhagen: Akademisk Forlag.

Treis, Yvonne & Vanhove, Martine (eds). 2017. *Similative and equative constructions; A cross-linguistic perspective*. Amsterdam: Benjamins

Tsai, Wei-Tien Dylan. 2008. On relativizing adverbials. *Language and Linguistics* 9(4). 967-984.

Van Praet, Wout. 2019. Focus assignment in English specificational and predicative clauses: intonation as a cue to information structure? *Acta Linguistica Hafniensia* 51(2). 222-241.
DOI: 10.1080/03740463.2019.1641318

Van Praet, Wout & Davidse, Kristin. 2015. Revisiting the typology of English copular clauses: ascription and specification in categorizing and identifying clauses. *Leuven Working Papers in Linguistics* 1.

Van Praet, Wout & O'Grady, Gerard. 2018. The prosody of specification: Discourse intonational cues to setting up a variable. *Journal of Pragmatics* 135. 87-100.

Verhoeven, Elisabeth & Skopeteas, Stavros. 2015. Licensing focus constructions in Yucatec Maya. *International Journal of American Linguistics* 81(1). 1-40.

Wagener, Terje. 2017. *The history of Nordic relative clauses*. Berlin & Boston: De Gruyter Mouton.

van der Wal, Jenneke. 2009. *Word order and information structure in Makhuwa-Enahara*. Utrecht: LOT.

Ward, Gregory & Birner, Betty. 1995. Definiteness and the English Existential. *Language* 71(4). 722-742.

Weinert, Regina & Miller, Jim. 1996. Cleft constructions in spoken language. *Journal of Pragmatics* 26. 173-206.

Wongwattana, Unchalee S. 2015. Complexities of Thai Copular Constructions. *Journal of the Southeast Asian Linguistics Society* 8. 97-120.

Wu, Yicheng. 2017. *The interfaces of Chinese syntax with semantics and pragmatics*. London & New York: Routledge.

Wunderli, Peter. 1978. Ce neutre en moyen français. In Wilmet, Marc (ed.), *Sémantique lexicale et semantique grammaticale en moyen français; Colloque organisé par le Centre d'Etudes Linguistiques et Littéraires de la Vrije Universiteit Brussel (28-29 septembre 1978)*, 185-256. Brussels: Centrum voor Taal- en Literatuurwetenschap, Vrije Universiteit.

Wunderli, Peter. 1980. Ceci et cela en moyen français. In Domincy, Marc & Wilmet, Marc (eds), *Linguistique romane et linguistique française; hommages à Jacques Pohl*, 247-275. Brussels: Edition de l'Université de Bruxelles.

Zahn, Fanqiong & Traugott, Elizabeth Closs. 2015. The constructionalization of the Chinese cleft construction. *Studies in Language* 39(2). 459-491.

Zimmermann, Michael. 2009. On the evolution of expletive subject pronouns in Old French. In Kaiser, Georg A. & Remberger, Eva-Maria (eds), *Proceedings of the workshop 'Null-subjects, expletives, and locatives in Romance'*, 63-92. Konstanz: Fachbereich Sprachwissenschaft der Universität Konstanz.

Zobel, Sarah. 2017. The sensitivity of natural language to the distinction between class nouns and role nouns. In Burgdorf, Dan & Collard, Jacob & Maspong, Sireemas & Stefansdottir, Brynhildur (eds), *Proceedings of Semantics and Linguistic Theory (SALT)* 27, 438–458. Ithaca: CLC. doi:10.3765/salt.v27i0.4182.

Zobel, Sarah. 2021. Zur Determiniererlosigkeit bei prädikativ verwendeten zählbaren Nomen im Deutschen: Korpusdaten und ihre Konsequenzen. *Linguistische Berichte* 268. 455-504.