THE EMERGENCE OF FUNCTIONS IN LANGUAGE

Zygmunt
Frajzyngier
&
Marielle
Butters

The Emergence of Functions in Language

The Emergence of Functions in Language

ZYGMUNT FRAJZYNGIER AND
MARIELLE BUTTERS

OXFORD
UNIVERSITY PRESS

Contents

Acknowledgments

List of abbreviations

1	first-person
2	second-person
3	third-person
AAM	additional argument marker
ACC	accusative
ADDR	addressee
ADJ	adjective
ADV	adverb
AFF	affected
ALL	allative
ANAPH	anaphora
APPL	applicative
Ar.	Arabic
ASSC	associative
ASSOC	associative
AT	stative locative preposition
ATT	attributive
AUX	auxiliary
AVP	adverbial phrase
C.FOC	contrastive focus
CAUS	causative
COL	collective
COLL	collective
COMM	comment marker
COMP	complementizer
COMPL	completive
COND	conditional
CONJ	conjunction
CONT	continuative
CONTRAST	contrastive
COP	copula
CQ	content question marker
D	dependent (aspect)
D.	distal
DAT	dative
DAT.OR	dative orientation

DEB	debitive
DED	deduced reference
DEF	definite
DEM	demonstrative
DEST	destinative
DET	determiner
DIMIN	diminutive
DU	dual
DUB	dubitative
EE	end-of-event marker
EF	required complement
EFFR	efferential
EMPH	emphatic
EP	epenthetic
EX	existential
EXCL	exclusive or exclamation
EXT	verbal extension
F	feminine
F.	Fula (Fulfulde)
FOC	focus
FR.	French
FREQ	frequentative
FUT	future
GEN	genitive (relationship or genitive case)
GER	gerund
GO	goal
gr	grade (for Hausa verbs)
H.	Hausa
HAB	habitual
HL	human locative
HUM	unspecified human subject
HYP	hypothetical
IMP	imperative
IMPER	imperative
IMPF	imperfective
IN	inner space
INCEPT	inceptive
INCL	inclusive
INDEF	indefinite human subject
INF	infinitive
INN	inner space
INTENS	intensifier
INTERJ	interjection

INTNS	intensifier
K.	Kanuri
L	logophoric
L.A	locative anaphor
LAM.	Lamang
LOC	locative
M	masculine
MOD	modal
N	noun; nasal consonant
NEG	negative
NG.	Ngambay
NOM	nominalizer or nominative
NP	noun phrase
NUM	numeral
O	object
OBL	oblique
ON	extension 'on'
ONOM	onomatopoeic
OPT	optative
OUT	extension 'out'
P	proximate
PART	partitive
PAST	past
PB	phrasal boundary
PL	plural
PNCT	punctual
PO	potential object
POL	polite
POSS	possessive
PRED	locative predicator
PREP	preposition
PRES	presentative
PRF	perfective
PRO	pronoun
PROG	progressive
PROH	prohibitive
PROX	proximate
PRS	presentative
PURP	purpose
PVG	point-of-view of goal
PVS	point-of-view of subject and point-of-view of source
Q	question
QUANT	quantifier

R	root and remote
R1	first reduplicant
R2	second reduplicant
RE	reverse
REC	recent (past)
RECIP	reciprocal
REDUP	reduplication
REF	referential
REFL	reflexive
REL	relative
REM	remote
RQ	rhetorical question
S	source; subject; sentence
SEQ	sequential
SG	singular
SPEC	spatial specifier
STAT	stative marker
SUBJ	subjunctive
SVC	serial verb construction
T	target
TAG	tag question
TEMP	temporal
TO	destinative
TOG	together (plural participants)
TOP	topicalization
TOT	totality
TQ	question about the truth
TR	transitive
UH	unspecified human (subject)
UNSP	unspecified
V	verb
VENT	ventive
VN	verbal noun
/	short pause
//	long pause
XXX	unidentified morpheme

1
Introduction

1.1 Introduction

1.1.1 Basic assumptions

The grammatical system of a language is composed of a set of functions and the forms that encode those functions. The importance of the functions encoded in the grammatical systems of an individual language is manifold. Here are just two points: (1) Speakers of individual languages must attend to the functions that are encoded in those languages (Jakobson 1959),[1] whenever communication involves one or more of those functions. (2) the realization of the functions encoded in individual languages determines, to a large degree, the forms of utterances in that language. The justification of this basic assumption and its implications can be found in Frajzyngier with Shay (2016).

1.1.2 The main question

A fundamental question of the present study is why grammatical systems in some languages code functions that are not encoded in the grammatical systems of other languages, regardless of the formal means by which they are encoded. The commonly invoked motivations for grammaticalization are communicative need and creativity (Heine and Kuteva 2002, Heine 2014, Kuteva et al. 2019). Communicative need, while a plausible explanation, begs the question of why speakers of different languages, especially languages spoken in the same geographic environment, differ in their communicative needs, despite sharing similar economic conditions, social organization, and cultural characteristics. Culture-specific conditions, such as religious and political beliefs, social structure, and history may affect language structure, but these usually affect one functional domain (Frajzyngier 2019c). Creativity

---

[1] 'Languages differ less in what you can express in them than in what you must express in them' (Jakobson 1959, reprint 1971, p. 492).

is one of the abilities that is involved in solving problems, but it is too broad a concept to provide a falsifiable explanation about differences in the functions encoded across languages. These two factors, communicative need and creativity, need to be supplemented by more specific reasons whose validity can be checked through use of rigorous methods. Hence the main question is what the motivation is for encoding different meanings in the grammatical system.

The present book offers a theoretical framework for the study of the emergence of functions and a methodology to conduct such a study. The individual chapters discuss various motivations for the emergence of functions. Each chapter consists of several case studies supporting the proposed motivation. We do not claim that all possible motivations for the emergence of functions have been discovered; the study leaves open the possibility that there may be other motivations that we have not yet explored. Having a reasonably large list of motivations for the emergence of functions will constitute a starting point for the potential expansion or reduction of the list. Such a list will be an important contribution to an explanation of why languages are similar and why they are different in all areas other than phonology. In the approach we take, phonology is a coding means rather than a functional domain or a function.

1.1.3 Theoretical assumptions: The meaning of a linguistic form

Frajzyngier with Shay (2016) claim that the grammatical system of every language encodes a unique semantic structure. At any given time, this structure is composed of a finite number of functional domains. Each functional domain has a finite number of functions. The functional domains and functions encoded differ across languages. Even if languages have similar functional domains, the internal structure of the domains may differ in the types and number of the functions encoded.

At any given time, each language has a finite number of formal coding means, which may include: lexical categories and subcategories; derivational morphology to change a lexical category or to derive new lexical items within an existing category; linear orders (Frajzyngier 2011, Frajzyngier with Shay 2016); use of lexical items to code functions, e.g. serial verb constructions and auxiliaries; use of nouns to code a variety of semantic functions, such as spatial relationships; prepositions and postpositions (often derived from verbs and nouns); phonological means such as pauses, intonations, and other prosodic

means are the markers coding phrasal, clausal, or sentential boundaries (Frajzyngier 2016); inflectional morphology used to indicate relationships between members of the utterance or to directly code elements from the semantic structure; repetition of lexical items and phrases; and potentially other formal means. The number and types of coding means cannot be determined now, as thousands of languages remain undescribed and many languages have been described only with the conceptual apparatus developed in the description of Indo-European languages. Furthermore, the formal means within each language can be combined, leading to an even larger number of coding means and thus creating more forms to code more functions.

The role of the formal means is to (1) code functions that compose the semantic structure of the language, and (2) ensure the principle of functional transparency, i.e. a principle that states that the role of every constituent in the utterance must be transparent to the listener (Frajzyngier 2004b). Lexical categories, linear orders, the use of lexical items to code functions, adpositions, and particles are components of what is traditionally called 'syntax.' Within the proposed approach herein, the formal means in a language interact in the coding of functions composing the semantic structure of the language. The autonomy of syntax as a system of rules for forming large structures and the process of combining lexical items into larger units as an outcome of the lexical properties of the heads, as proposed in Chomsky (1995), cannot be maintained in the proposed approach.

The grammatical functions of the coding means may include indicating relationships among the elements of the utterance and the coding of other functions that are part of the semantic structure of the language, such as modality, tense, aspect, directionality of movement and spatial orientation with respect to the deictic center. Also included are the relationship between speaker and listener, the social status or affiliation of the speaker and possibly the social status of the listener, and the social relations between the speaker and the listener. There is no a priori limit as to how many and what kind of functional domains are encoded in each language.

In the process of language use, a speaker chooses to produce functions encoded in her/his language and lexical items that represent the activities and entities that the speaker wants to talk about. The only rules of syntax that exist are the rules that allow the speaker to convey the functions encoded in the language. The application of these rules takes into consideration the lexical items chosen by the speaker.

Within the proposed approach, the fundamental object of semantic investigation is not a sentence, a clause, or an utterance. The main object of

semantic investigation is the semantic structure encoded in the grammatical system of the language. Much of the past literature on semantics is based on questions about the meaning of a sentence in a language or about the meaning of a word in a language. Moreover, the answers to such questions were usually based on inferences about reality stemming from the implication of the utterance. These types of questions are essentially irrelevant to the task of describing the semantic structure encoded in the grammatical system. On the other hand, having a list of functions encoded in the grammatical system of a given language will make it easier to answer questions about the meaning of individual utterances.[2] The meaning of an utterance is then the sum of all the functions encoded in the utterance.

The semantic structure of every language needs to be discovered, i.e. linguists need to determine what are the domains encoded in any given language and what are the functions composing each of its domains. Neither the domains nor the functions are available through speakers' introspections, in the same way in which, just because we are human beings, an untrained person cannot describe the functioning of our anatomy, physiological, or mental processes. Moreover, the discovery of the meanings encoded in the grammatical system cannot be based on inferences about individual utterances or, as is done in some philosophical literature, on inferences from individual sentences.

1.1.4 How to describe functions

All functions within the domain have one common feature which is the defining feature of the domain. For example, in a language that has the domain 'aspect,' every function within this domain must characterize the event with respect to its internal structure (or whatever function of 'aspect' has been discovered and postulated for that language). Some functional domains may have subdomains. Thus, the domain of aspect may have subdomains of perfective, imperfective, etc. In several African languages, there is a further distinction within some aspects and tenses of coding pragmatically dependent

---

[2] We use the term 'utterance' for any instantiation of linguistic production, rather than the term 'sentence.' The universality of the category 'sentence' is very much in doubt, as linguists cannot provide a universal definition of the category 'sentence.' Even in individual languages one sometimes cannot provide a satisfactory characterization of the notion 'sentence' (see Mithun 2004 and Frajzyngier, Gurian, and Karpenko (in press a. and b.)). When we use the term 'sentence' it is in references to examples representing whole sentences in languages in which such entities have clear formal characteristics, whether in the spoken or in the written variety.

as opposed to pragmatically independent clauses (Frajzyngier 2004a, Frajzyngier and Shay 2010). Languages differ in what functional domains are encoded in the grammatical system and in the composition of functional domains.

The encoding of a functional domain in the grammatical system has important consequences for speech production in a given language. If a language codes some functional domain, and the content of speech is, in one way or another, subsumed by this functional domain, the speaker must address the categories encoded in the functional domain. As an illustration, consider the function encoded by the past tense in English (all labels of grammatical categories are as used in standard English grammars without any commitment to the function they encode). If the event took place in the past time, the speaker must use the past-tense marker, unless the clause represents the historical present. If the speaker does not, then the speaker violates Grice's cooperative principle. Similarly, if the speaker in English talks about an event that involved many objects, the speaker must encode the plurality of the objects, lest he or she violate the cooperative principle or produce an ungrammatical utterance.

1.2 Universal versus non-universal functional domains

1.2.1 Universal functions

We can assume that some functions encoded in the grammatical systems are universal, in that they reflect the basic communicative function of the language. The term universal implies that either every language has formal means to encode a given function or that a given function is the default value of some formal means, e.g. of a linear order, of a lexical category, or of an intonation pattern. The existing descriptions of individual languages indicate that every language has some means to code negation; polar questions, i.e. questions about the truth of a proposition; and content questions. Moreover, every language has a reference system with some means to identify referents of the constituents in the utterance. It is also likely that all languages have the function of assertion, which in many languages is the unmarked modality of the clause. There are, however, languages where the assertive modality must be overtly marked by a marker that is not found in modalities other than the assertive, as is the case in Swahili (Bantu, Niger Congo) and Bele (Kru, Niger Congo, Rebecca Burns, Ms.). It is very likely that all languages encode the

modality of obligation in their grammatical structure. Whether there are other universal domains across languages remains to be discovered. The addition of other universal functions does not in any way affect the theoretical foundations of the present work.

The universal functional domains may differ in the subdomains and functions coded. For example, within the domain of interrogative modality, some languages may encode a speaker's assumption about the potential answer to a polar question and other languages do not. In English, this function is coded by tag questions, while in some other languages there is no such function. Within the domain of modality of obligation, some languages make a distinction between the imperative, which we take to code obligation with the expectation of immediate realization, and optative, i.e. an obligation without such an expectation. In other languages, there may be only one function within the modality of obligation which does not distinguish between obligation with respect to the listener as opposed to other participants and does not distinguish between obligation with the expectation of immediate realization and other obligations.

Within the domain of relations between the predicate and noun phrases, some languages code the semantic relations of all participants and other languages do not. Some languages code benefactive function while other languages do not. Some languages code indirectly affected argument, while other languages do not. Some languages code affected subject, while other languages do not.

The other way in which languages differ are in the functional domains encoded. The standard illustrations include the fact that some languages code the domain of tense and aspect, some code only tense, some code only aspect, and some code neither tense nor aspect. Here are some less often cited examples. The single-generation Sino-Russian idiolects do not appear to code the relationships between the predicate and noun phrases, a function that is considered by some a core function of the grammatical system. Chadic languages encode a domain of locative predication, with several functions composing the domain, including andative, ventive, stative, and directional. No such domain is coded in the grammatical systems of Indo-European languages.

1.2.2 Rephrasing the main question: How do non-universal functions come about?

Assuming that all humans have identical physical makeup and have similar biological and social needs, why do languages differ in the functions encoded

in their grammatical systems? It is a fundamental question for understanding the differences and similarities among languages. This is also a difficult question. One way to answer this question is by studying how various functions have emerged in different languages. The question of such a study is not where the markers of various functions come from (as this has been thoroughly covered in numerous studies of grammaticalization), but rather why a given function came to be encoded in the grammatical systems of a given language.

1.3 State of the art

Current historical studies provide ample information about the derivation of grammatical markers from lexical sources, in the Humboldtian tradition. Studies that continue this tradition, e.g. Meillet (1912) who apparently coined the term 'grammaticalization' as cited in Lehmann (1982/2002),[3] Kuryłowicz (1965), Samuels (1975), Lehmann (1982/2002), Heine and Kuteva (2002), Kuteva et al. (2019), focus on the lexical origins of grammatical markers. Studies of grammaticalization from sources other than lexical items, e.g. Frajzyngier (2010b), provide insight into how semantically empty formal means, such as tone or linear order can become formal means coding semantic functions. The notion of exaptation, adopted for linguistics by Lass (1998, 1990), has turned out to be an explanation for the emergence of several functions (Norde and Van de Velde 2016). Kuryłowicz (1965) and in his other works invoked analogy as a mechanism in grammaticalization. Analogy is indeed a process through which new functions can emerge.

Two external motivations for the emergence of functions may be responsible for the coding of various functions. One is the influence of the geographical environment in which a language is spoken. The salient characteristics of the environment are sometimes reflected in the coding of directionality and spatial relations, as in Karuk (alternative spelling (Karok) in Bright 1957) and many other languages.

The other external influence are social relations, religious beliefs, and political manipulations whereby some societal norms or desires of various groups can filter into the grammatical system. Such influences are reflected in the honorific systems of such languages as Japanese, forms of address and reference in many Indo-European languages, and the entrenchment of various forms and functions that are believed to represent higher social register. For

---

[3] We are grateful to an anonymous referee for reminding us about Meillet's work.

further explanation of social factors that play a role in the coding of various functions see Hagège (1993).

These external factors cannot be the causes of difference in functions coded in languages spoken in the same geographical environment and by societies that have the same economic base, e.g. grow the same crops, have the same social system, and the same religious beliefs. Within any given society there may exist language use norms and associated function that apply to some members of the society and not to others.

The bulk of the functions that are coded in the grammatical systems of some languages but not in others remain unexplained. In sum, the present state of the knowledge does not allow an explanation of why a given function emerged in one language and not in another. Consequently, we cannot predict what non-universal functions the grammatical system will encode.

1.4 The structure of the book

Chapter 1 and Chapter 2 address the theoretical assumptions and the methodology in the current work. In Chapter 2 we discuss the methodologies required to answer the basic question of this book. One is the methodology for the discovery and determination of functions coded in the grammatical system and the other is the methodology for the discovery of how functions coded in the grammatical system emerge.

The next two chapters address the existence of specific motivations. Chapter 3 describes forced interpretation where a speaker narrows down possible interpretation of utterances, and Chapter 4 demonstrates the emergence of a function as a means to avoid systematic ambiguity.

The initial state as a motivation for the emergence of functions is addressed in three chapters. Chapter 5 describes the emergence of benefactive functions because of narrowing from an initial state, Chapter 6 discusses the emergence of the point-of-view of the subject, and Chapter 7 describes the emergence of the function goal-orientation.

Chapter 8 deals with the role played by the principle of functional transparency in the emergence of functions. The specific function discussed in this chapter is the coding of the presence in the proposition of an additional argument, i.e. in the semantic content of the clause. The semantic role of that argument may be, but does not have to be, overtly marked.

Chapter 9 describes the role of the lexicon in the emergence of locative predication.

Chapter 10 describes the emergence of functions coding the relationship between propositions.

The opportunistic emergence of functions, whereby the presence of formal means without an assigned function is a formal niche for the emergence of new functions, is addressed in the next three chapters. Chapter 11 describes the emergence of the complex action, Chapter 12 describes the emergence of coding of gender and number in content questions in the domain de dicto, and Chapter 13 describes the emergence of grammatical relations.

Chapter 14 describes the emergence of a functional domain, not merely a function, through language contact.

Chapter 15 contains conclusions, implications, and open questions.

In what follows we briefly describe the importance of each motivation for the emergence of functions.

1.5 Communicative motivation

Communicative need and creativity have been the only motivation postulated in the literature to be responsible for the emergence of grammatical morphemes. One can claim with good reason that emergence of every function fulfils some communicative need. If communicative need were indeed to be the main motivation, then one would also need to explain why different languages code different functions. Do the speakers of those languages have different communicative needs, and if so, why? Given that such questions cannot be answered at the present time, one needs to find out which communicative needs are motivations for the emergence of functions. For the communicative need to be a motivation for the emergence of functions one needs to narrow it down to such situations where indeed the communicative process is in one way or another affected. The Chapters 3 and 4 describe two such processes. The first one deals with what we call forced motivation. The listener can interpret the connection between the sequence of clauses in discourse in several ways. And yet the speaker may want to constrain this interpretation to just the one interpretation the speaker has in mind. This forced interpretation is the motivation for the emergence of various markers which tell the listener how to interpret the connection between clauses in discourse.

The other communicative need that we describe deals with a situation where the existing structures create a systematic ambiguity with respect to some function encoded in the language. The specific case involves the reference of the subjects of the complement clause of verbs of saying. Such

situations create a need to distinguish between the participants in the ongoing discourse, participants in the reported discourse, participants in the matrix clause, and participants in the embedded clause. Some languages code co-reference using pronouns and disjoint reference using full noun phrases (English), some languages code co-reference using the coding on the verb ('agreement') and disjoint reference using pronouns, and some languages have grammaticalized a special set of logophoric pronouns to code co-reference with the participants of the matrix clause and use the matrix clause pronouns to code disjoint reference. Some languages (Pero, Frajzyngier 1989) have encoded in their grammatical system a special set of pronouns to refer to participants in the reported discourse.

1.6 The role of the initial state

Any stage of the grammatical system represents for its speakers the set of forms and functions that can undergo a change. We call this state an 'initial state,' in analogy with the stages in crystallography: A function F1 encoded at the initial state X constitutes a stimulus and a base for the emergence of a function F2 with a different value within the domain of the function F1 (Frajzyngier 2010b). Thus, if a language has only one or two tenses, e.g. future (marked) and non-future, one can expect the emergence of other tenses within the domain of tense. The new functions may be coded by alteration of formal means that code functions at the initial state or by the deployment of other formal means. Over time, the new function may become the default, i.e. unmarked, value of some form. An interesting and important factor in the process of the emergence of a function is that the function present at the initial state does not have to be overtly marked and therefore may not have been observed by linguists analyzing the language. The fact that speakers of the language associate a given form with a given function allows this form to become a means of coding a different function. Consider SV order, without rising intonation, in English. At one time this order came to be interpreted as indicating assertive modality, i.e. indicating what the speaker wanted the listener to believe (see Davidse 1998: 283). In contemporary English, the interrogative modality is derived from the assertive modality by inversion to VS, if V represents the verb 'to be,' or by insertion of the verb 'do' before SV. Changing the assertive modality into the negative modality is achieved through negation of the verb 'to be' or through negation of the auxiliary 'do.'

Any function coded at the given stage constitutes a nucleus from which other function(s) within the same functional domain can emerge. What follows is the list of case studies that we include in the book to document the emergence of functions within an existing domain.

1.6.1 The benefactive function in English

This chapter describes the emergence of the benefactive function from an initial state that emerged after the collapse of case marking in English.

The benefactive function emerged in English as a product of narrowing after the erstwhile indirect object function was lost following the loss of the dative case marker. The noun phrases marked by the dative case could be interpreted in several ways, and benefactive was only one of them. The linear order V NP1 NP2, frequently found in phrases in which NP1 used to be marked for the dative case, came to code only the benefactive function.

1.6.2 Point-of-view of the subject

Point-of-view of the subject instructs the listener to interpret the event in regards to how it concerns the subject. The function can be coded with both transitive and intransitive verbs. This chapter discusses the emergence of the point-of-view of the subject function from several types of initial states. In Romance, Germanic, and Slavic languages, this function, coded by reflexive markers/pronouns, emerged from an initial state in which the verbs inherently represent goal-orientation. In some Chadic languages, the point-of-view of the subject emerged in languages where the verb did not inherently indicate any point-of-view. If the clause is coded for the point-of-view of the subject, no goal can be added. Thus, no indirect object can be added to a clause with the point-of-view of the subject in Chadic languages. The importance of this section is that it demonstrates that identical functions can emerge because of different motivations created by different initial states.

1.6.3 Goal-orientation

Goal-orientation instructs the listener to interpret the event as oriented toward some goal. The goal could be an equivalent of a second argument with an

intransitive verb, an argument other than the second argument with a transitive verb, or a locative complement. Many Chadic languages whose verbs inherently represent point-of-view of the subject ('anticaustive' in Schuh 2017) have grammaticalized point-of-view of the subject.

1.7 The principle of functional transparency

The principle of functional transparency, as formulated in Frajzyngier (2004b) and Frajzyngier and Shay (2003), states that the role of every constituent in the utterance must be transparent to the listener. The role of the constituent in the utterance is quite distinct from the role of the referent of the constituent in extra-linguistic reality. Functional transparency is satisfied by the deployment of the coding means existing in a given language, which may include: lexical categories, derivational markers, linear orders of constituents, deployment of lexical items to code functions, e.g. serial verb constructions, adpositions, inflectional coding on all lexical categories, repetition of phrases, and several other means not listed above. The principle of functional transparency is responsible for the presence of prepositions and case marking. If the role of a constituent to be included in the utterance is inherently not transparent, one of the coding means existing in the specific language needs to be deployed. The coding of the role of constituent in the utterance is thus a motivation for the emergence of the function.

The principle of functional transparency as described above is expanded in the present work to include functions that are part of the semantic structure encoded in the grammatical system. The listener expects that the utterance contains markers to encode those functions when the situation invoked may include such functions. To give a simple illustration, in English indicative mood the clause must include information about tense. In a language that does not have tense in its grammatical system, such information is not required.

1.8 The role of the lexicon in the emergence of functions

It has long been widely assumed that in individual languages lexical items have inherent properties that affect or even determine the formation of larger entities (syntax) (see Chomsky (1995) to this effect with respect to verbs). The inherent properties of lexical items constitute an initial state and may trigger the emergence of a function. In this study, we demonstrate how the

inherent properties of verbs and nouns can lead to the emergence of a function. The emergence of a function from lexical properties of nouns and verbs is demonstrated by locative predication.

1.9 Metonymic extensions as a motivation

One type of emergence of a new function occurs when a form that codes one function comes to be associated with another function through implication. The new function may then receive its own coding means. This type of metonymic extension led to the emergence of three different functions coding relations between propositions in Lele (East Chadic). None of these relations includes the clausal coordination widely known in Indo-European languages.

   The chapter dealing with metonymic extension also presents a methodology to discover the emergence of functions through internal reconstruction. A function in the given domain may also emerge as a result of language contact. In the same language, there appears a fourth function coding relations between propositions, this time resulting from language contact.

1.10 Opportunistic emergence of functions

Opportunistic emergence of functions means that there are some formal means of coding that either do not have a function assigned or that provide new possibilities for the recombination of the formal coding means. Such an abundance of means produces formal niches that can be exploited for the coding of new functions. Three unrelated cases demonstrate the opportunistic emergence of functions.

1.10.1 The emergence of an associated event

Verbs in Bantu and Chadic languages have verbal affixes known as 'verbal extensions.' As in Amazonian languages, these affixes often code spatial relationships that include the directionality of movement and the spatial configuration of participants with respect to some locative center. In some languages, verbal extensions have come to encode associated movement, resulting in complex propositions whose semantic content corresponds some-what to 'proposition 1 & proposition 2,' e.g. 'he ate and left' or 'he ate and

came.' In some languages, the verbal extensions gave rise to complex activities beyond movement. We explicitly formulate the semantic correspondence as approximation rather than identity because the functions encoded by verbal extensions constitute a function of their own as already postulated by Koch (1984), Wilkins (1991), and Guillaume (2016) and his earlier writings, a function which is unrelated to clausal coordination. Nevertheless, the importance of the emergence of this function is that it has the potential to explain why there is no coordinated clausal conjunction in some languages that have these structures.

1.10.2 Coding gender in content questions

In English, the question words 'who' and 'what' code the distinction between human and non-human participants. In Gidar (Central Chadic) the speakers must include gender and number in their assumptions about the potential participant, namely whether the participant is masculine singular, feminine singular, or plural. The reason for including this assumption is as follows: The content interrogative phrase in Gidar includes a copula, i.e. is equivalent of 'who is'; copulas in Gidar derive from pronouns; and third-person pronouns in Gidar distinguish gender and number. Tibetan does something similar in that speakers will repeat the interrogative 'who' if they anticipate a plural answer with respect to humans.

1.10.3 Coding grammatical relations

Unlike various theoretical approaches, generative and functional alike, that take the dichotomy of the subject and the object to be basic components of clausal structure, this study demonstrates that in some languages this dichotomy is a product of functional changes from two different initial states. One is a state whereby no semantic or grammatical relations have yet been coded, represented by basic variety languages as described by Klein and Perdue (1997), and Frajzyngier, Gurian, and Karpenko (in press a. and b.). The other initial state is represented by languages whereby the rich system of coding semantic and grammatical relations between the predicate and noun phrases has collapsed through the reduction of the inflectional coding means. After the collapse of the inflectional coding means on the verbs and noun phrases, languages are rebuilding the system of coding relations between the predicate and noun phrases using the only means that is readily available:

linear order. This is the situation that is represented by English after the collapse of case marking and the coding on the verb. In the present study, this situation is illustrated by the grammatical system of Mupun (West Chadic).

1.11 Emergence of functions through language contact

The abundant literature on language contact demonstrates that functions can be borrowed from one language to another (Heine and Kuteva 2005). As expected, languages often borrow functions that they do not code in their own grammatical systems. Matras (2009, chapter 8) lists numerous instances of languages borrowing discourse connectors, temporal expressions, indefinite expressions. Heine and Kuteva (2005) is focused on the emergence of grammatical morphemes from lexical items and the emergence of more abstract grammatical morphemes from the less abstract grammatical morphemes. Heine and Kuteva (2005) contains dozens of examples of grammatical functions emerging in languages as a result of language contact.

In the present study, we describe a case where not a function but rather a functional domain emerges as a result of language contact. Moreover, it appears that what was borrowed was general means of coding rather than specific forms. It also appears that this borrowing took place in conditions other than bilingualism which is often described as a necessary condition for borrowing among languages.

1.12 Languages often discussed in the present volume

The data in the present study are drawn from a variety of languages familiar to most readers, including English, French, Spanish, Polish, Russian, and also a variety of less familiar Sino-Russian idiolects and West and Central African languages. To avoid repeating the same information in various chapters we include here the information on lesser known languages from which some of the data were drawn in the present study. For many languages, the analyses presented here supersede the analyses in the sources from which the data are drawn.

1.12.1 Sino-Russian idiolects

The Sino-Russian idiolects discussed here are spoken by adult Chinese immigrants to the Far East Region of Russia who have never had any formal

schooling in Russian (Frajzyngier, Gurian, and Karpenko (in press a. and b.) and references there). The data for Sino-Russian idiolects cited in this study consist of natural language narratives collected from about ten speakers of Mandarin Chinese who have resided in Russia for between two and 20 years. The typed transcription of the data is about 40 pages. Each idiolect is used by its speaker in communication with Russians and is not used in communication with other speakers of Mandarin. The idiolects are not used at home. These idiolects are not a historical continuation of the Russo-Chinese pidgin, mainly a trade language that used to be spoken along the Russian-Chinese border at the end of nineteenth and the beginning of the twentieth century. The condition of the use of the Sino-Russian idiolects is thus significantly different from the conditions in which pidgins are typically used. Because of this limited linguistic interchange, each Mandarin speaker generates his or her own language using Russian lexical items. The lexical items are most often bereft of grammatical characteristics. Though there are cases where some traces of inflectional coding appear on lexical items, this does not indicate that speakers have a productive knowledge of the inflectional system of Russian, since in the same text the relevant inflectional marking(s) may be missing or may be used in contexts where they are not used in Russian. The Sino-Russian idiolects have no inflectional or derivational system, no categories of number, tense, or aspect even though the individual lexical items have been borrowed from Russian with their inflectional markers, often coding tense, aspect, person, or number. There is no evidence, however, that these markers represent part of the grammatical structure of the Sino-Russian speakers. The various forms can be used in utterances having different persons or tenses than the one marked by inflectional markers. The phonological system of the Sino-Russian idiolects do not have a fixed system of underlying forms and the rules of phonetic realization. Consequently, the same lexical lemma may have different phonetic realizations in the same environment. Even the same speaker may produce the same word in different ways. The most stable element in the realization of words is the place of articulation for different segments. Hence, the same word may have different written representations in the present work.

Throughout the book we use the notion of 'initial state' as a starting point for the emergence of functions. There is an important difference between single-generation languages, i.e. contact languages spoken by adults without any formal education in the language which they must use, and languages transmitted from one to another generation of speakers. For the speakers of single-generation languages there is no initial state in L2, as they do not know the grammatical system of the language whose lexical items they use. There is

an initial state of L1, in our case Chinese, but then, L1 is useless as a tool of communication with the speakers of L2. Consequently, the emergence of any new function in basic variety languages must have motivations other than the initial state.

1.12.2 Chadic languages

Chadic languages, spoken in Nigeria, Cameroon, Niger, and Chad, are the largest, between 140 and 160 languages, and most typologically diversified family within the Afroasiatic phylum. Chadic languages are divided into three or four families, West, Mandara, East, and Masa (Newman 1977) and West, Central, and East by Jungraithmayr and Ibriszimow (1994). There are clear typological distinctions between West and Mandara/Central languages on the one hand, and East and Masa languages on the other.

No Chadic language has case marking on nominal categories. Many languages have rich verbal morphology coding semantic relations between the verb and arguments, point-of-view, spatial relations within the proposition and spatial relations of the events, plurality of the event or plurality of the object rather than plurality of the subject (Frajzyngier 1985b). There is no passive or any other intransitivizing form in Chadic. The existence of causative markers is controversial. In some West and Central branch languages, there is an additional argument marker which indicates presence of an additional argument with inherently intransitive verbs, an indirect object with majority of verbs, and a locative or a time adverb with all verbs.

The reason why Chadic languages are important for the present study is that the evolution of these languages from the same initial state understood diachronically, produced different typological characteristics and different initial states for many languages spoken in similar geographic, economic, and cultural conditions. These different initial states allow for a rich field of observation of the emergence of a variety of functions.

Hausa (West Chadic):   Hausa is the largest Chadic language spoken by more than 20 million people in Northern Nigeria, Niger, and in scattered Hausa colonies across West Africa. It is also a vehicular language for speakers of other languages in Northern Nigeria. Hausa is also one of the more frequently described African languages with many grammars and hundreds of papers written in English, German, French, Russian, Polish, Italian, and in Hausa itself. The language is usually analyzed as having the linear order SVO. Subject

pronouns code person, number, gender, and some modal and aspectual values. The verbal system of Hausa consists of the root to which vocalic, consonantal, and tonal markers are added. For recent and most complete descriptions of Hausa see Newman (2000) and Jaggar (2001). For a new conceptualization of the Hausa verbal system see Frajzyngier and Munkaila (2004b).

Pero (West Chadic):   Pero is spoken in a few villages including Filiya, Gwandum, and Wurkum in the southern part of North-Eastern Nigeria (Frajzyngier 1989). At present, most speakers of Pero also speak Hausa and English, but as recently as 1938 most Pero people did not speak Hausa. Reverend Faust and other missionaries who worked in the area in mid 1930s reported (p.c.) that speakers of Pero asked missionaries to introduce Hausa classes in the village so that young people could learn this language to be able to communicate outside of the village. Pero has the category subject as evidenced by the existence of subject pronouns and a focus marker which marks only subjects in focus. The nominal subject may occur in either clause-initial or clause-final position. Not every clause has to have the subject overtly marked. Clauses without overt nominal or pronominal subject are frequent. Unlike in some other languages, e.g. Mandarin Chinese (Chao 1968), the absence of the subject indicates only the third-person subject in Pero. Many clauses in natural discourse begin with a verb. The focus on the subject, whether nominal or pronominal, is marked by the preposition *nin* preceding either a noun or a pronoun in clause-final position.

Pero does not mark nominal plurality. It does mark verbal plurality which may lead to inferences about the plurality of the direct or indirect object, plurality of the event, or plurality of the subject of intransitive verbs. It does not code plurality of the subject of transitive predications (Frajzyngier 1985b).

Mupun (West Chadic):   Mupun, a member of the Angas group, is spoken in Plateau Province of Nigeria. The folk tradition of Mupun states that the current population is composed of immigrants who arrived from different directions into the area. The verbal inflection of Mupun has been drastically reduced in comparison to that of other Chadic languages of the West and Central branches. Mupun has the categories subject and object marked by the linear order SVO. Mupun also has a rich system of serial verb constructions, a system more common of Niger-Congo languages than Chadic languages (Frajzyngier 1993). Mupun and other languages from the Angas group display a rich system of logophoric pronouns. The importance of Mupun is that it

illustrates the rebuilding of the functions in the grammatical system after the collapse of the rich inflectional coding on the verb.

Hdi (Central Chadic):   Hdi is spoken in Tourou (Turu) and surrounding villages in the Far-North province of Cameroon, at the border with Nigeria. The present discussion is based on the data in Frajzyngier with Shay (2002), and Frajzyngier et al. (2017). The analyses in this work differ significantly from those previous works. The present analyses benefited from the research on the project 'Applying linguistics for the study of migrations,' which was supported by a Seed Grant at the University of Colorado. Hdi is a predicate-initial language, regardless of whether the predicate is verbal or nominal. The subject follows the predicate. The object is most often marked by the preposition *tá*. The locative stative complement is marked by the preposition *tà*. The locative directional complement 'to' is marked by the preposition *dá* or *dà*, depending on whether the complement is inherently locative or not.

The verb in Hdi can be simple or reduplicated. Subject pronouns are suffixed to the verb. The verb can have many markers, called 'verbal extensions' in Chadic linguistics. These markers are suffixed to the simple form and infixed in between the reduplicated form of the verb. These markers code semantic relations of arguments, spatial relations with respect to the locative center, directionality of movement with respect to the place of speech, associated movement, and the manner of the event. The verbal piece in Hdi may consist of up to ten morphemes, only one of which is a verb.

Hona (Central Chadic):   Hona is spoken in North Eastern Nigeria. Hona has two linear orders for clausal structures: verb-initial and subject-initial. The function of the two orders remains to be explained. Whether one or another order is deployed has important consequences for verbal morphology.

When the nominal subject follows the verb, there are no subject suffixes on the verb. But if the nominal subject is moved into clause-initial position, then the verb has a subject suffix. The data for Hona come from fieldwork conducted in 1980 by Frajzyngier and a few years later by Laurie Jordan, then an MA student at the University of Colorado. The final analysis of the data is suspended until there is a possibility to gather more data and check the data already gathered. Nevertheless, the data gathered so far contain information at the level of morphology and syntax reliable enough to be used in this study.