Master Thesis

Implementing Georgian Polypersonal Agreement through the LinGO Grammar Matrix

Irina Borisova

August 27, 2010

European Masters Program
in Language and Communication Technologies

University of Groningen & Saarland University

Supervisors: Antske Fokkens and PD Dr. Tania Avgustinova and
Prof. Hans Uszkoreit
Saarland University
Co-supervisor: Assoc. Prof. Gertjan van Noord
University of Groningen

Contents

Abstract

This thesis analyzes Georgian polypersonal agreement through the implementation of a Georgian grammar fragment using the LinGO Grammar Matrix customization system. The complex interaction between the argument marking patterns in Georgian is captured in a verbal morphology that is implemented through a set of lexical rules. Characteristics of Georgian polypersonal agreement, such as morpheme blocking, zero and ambiguous morphemes, varied expressions of grammatical functions are analyzed through the customization page. The implemented grammar is evaluated against a test suite which is developed to represent the key aspects of Georgian polypersonal agreement.

Acknowledgements

I owe a lot to my supervisors in Saarbrücken and Groningen. I am indebted to Antske Fokkens for her thorough guidance, for always finding time and spending it on consultations, debugging, reading and advising, for her patience and motivation. Without her this project would have neither started nor finished. I am grateful to Tania Avgustinova for encouraging this work and bringing clarity in the linguistic data and approaches to them. Thanks to Gertjan van Noord for providing valuable suggestions and encouraging me to be clear and concise. Thanks to the participants of the Monday meetings in CoLi, Yi Zhang and Maria Sukhareva, for asking questions and motivating me for more inquires and details.

I cannot express my full gratitude to the people who consulted me about Georgian, read through the hundreds of not very exciting sentences and tirelessly answered my questions: Anne Nemsitsveridze-Daniels, Mari Rosti, Robert, Irakli, Tinatin Bolkvadze, Lili Jologua, Tamuna Arahamia, Salome Bobokhidze, Salome Khetsuriani, Tamuna Morchadze, Gvanca Turiashvilii, and Givi Edisherashvilii. Thanks to Tinatin Bolkvadze and Damana Melikishvili, the professors of Iv. Javahishvili University in Tbilisi, for teaching me Georgian and explaining the peculiar details.

This work could be never done without the work of the LinGO Grammar Matrix community, in particular the Matrix developers. Thanks to Emily Bender for working on the LinGO Grammar Matrix and for her suggestions on animacy hierarchy; to Michael Goodman for being extremely responsive to my needs, for fixing the bugs and being very open-minded to my suggestions for the Matrix development; to Safiyyah Saleem for implementing the argument optionality library; to Laurie Poulson for valuable suggestions on the tense hierarchies.

This Masters program is a big journey, and I am grateful to Gisela Redeker, Gosse Bouma and Bobbye Pernice for making this journey smooth and enjoyable. Thanks to my peers in Groningen and Saarbrücken for their intellectual and psychological support, and especially to Xuchen Yao for motivating me with a question 'What's your progress in percentage?' and being extremely helpful with all sorts of issues.

Finally, thanks to my family and friends, who keep on asking questions, diving into linguistics, listening to me and believing that I am doing right. Thanks to John for hours of proofreading, correcting, and consulting, and above all - for support and happiness.

List of Tables

List of Figures

Abbreviations

Tenses
aor - aorist
cond - conditional
fut - future
fut_conj - future conjunctive
imp - imperfect
non-neg_imp - non-negative imperative
opt - optative
pluperf - pluperfect
pres_perf - present perfect
pres - present
pres_conj - present conjunctive
sr - series

Tables
arg - argument
constr - constraint
iccm - imperfect-conditional-conjunctive marker
intrans - intransitive
numb - number
req - require
no - forbid
trans - transitive

Gloss
1/2 - first or second
dat - dative
erg - ergative
nom - nominative
obj - object
partic - participle
perf - perfect
pl - plural
prv - preverb
sf - stem formant
subj - subject
sg - singular
ts - thematic suffix
ver - version (pre-radical) vowel

1. Introduction

In 1947, Warren Weaver, one of the pioneers of machine translation, wrote to Norbert Wiener:

> When I look at an article in Russian, I say: "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode." (Weaver, 1955, 17)

From the moment, when this fragment was included in the renown memorandum on machine translation, the field of computational linguistics has been seeking ways for the better 'decoding' the properties of one human language and 'coding' it to another human language. The consequent extensive research on language universals tried to amplify 'decoding' procedure, but indeed challenged the whole matter of this operation, due to the diverse and complex realizations of seemingly universal phenomena.

This thesis aims to 'decode' Georgian polypersonal agreement. Agreement, although not universal, but exceedingly frequent language phenomenon, refers to "some systematic covariance between a semantic or formal property of one element and a formal property of another" ((Steele, 1978, 610) after (Corbett, 2006)). This is a general yet structural definition, and throughout this research it is linked to a complex, varied and extremely representative language phenomenon.

1.1. Georgian polypersonal agreement

Georgian is the only written language of the South Caucasian linguistic group and the official language of Georgia. It is often classified as a polysynthetic language with free word order and split-ergativity. The Georgian verb is the focus of the current investigation: it is notable in its ability to express an exceptionally wide range of meaning; tense-aspect modifications, verbal constructions, and morphology can be unusually complex. In particular, all arguments of a predicate, such as subjects, direct and indirect objects, can be marked on the verb that they agree with. This subject-object marking on the verb form is called polypersonal agreement, or polypersonalism, and is regularly seen in non-Indo-European languages, such as Basque, Abkhaz, Chukchee, Ostyak.

The Georgian verb carries an exceptional semantic load in the sentence; verbal markers, given that Georgian is a pro-drop language[1], can be the only source of establishing the argument functions. The complexity of verbal meaning can be viewed from at least three perspectives: from a surface realization point of view, they are rooted in a morpheme, in the absence of a morpheme, or in the vague boundary between the morphemes;

[1]The term 'pro-drop' refers to the languages that allow pronoun arguments to be omitted.

semantically, they might be determined by a high aptitude of transformations; syntactically, they are constrained by the verb class, the applied tense-aspect model and the realized syntactic construction. The disambiguation of agreement markers incorporates the interaction of multiple components between the aforementioned levels, which made a particularly challenging and interesting task.

Georgian polypersonal marking contrasts sharply with verb-argument relations in Indo-European languages and particularly in subject-oriented languages which often appear as the core material in computational linguistics. Research on verb-argument relations in Georgian is not only a key step for any inquiries in the Georgian grammar, but can also provide valuable insights for research in other areas, such as syntactic-semantic inference and relations extraction.

1.2. Scope of the thesis

This thesis aims to analyze Georgian polypersonal agreement through an implementation of a fragment of the Georgian grammar using the LinGO Grammar Matrix customization System (Bender et al., 2002), (Bender et al., 2010). The LinGO Grammar Matrix is a multilingual grammar engineering platform developed on the theoretical grounds of Head-Driven Phrase Structure Grammar (Pollard and Sag, 1994) and Minimal Recursion Semantics (Copestake et al., 2005). The LKB grammar engineering environment (Copestake, 2002) and [incr tsdb()] test environment (Oepen, 2001) are used to evaluate grammar performance. Data are collected from the native speakers through questionnaires on grammaticality and a translation set from English and Russian into Georgian.

The study of Georgian polypersonal agreement through the LinGo Grammar Matrix follows the main idea underlying the Grammar Matrix, that is, to provide an easy and standard start for grammar engineering and language description. In particular, the LinGO Grammar Matrix has been recently used for testing the linguistic hypotheses in several studies ((Bender, 2010),(Drellishak, 2009),(Fokkens et al., 2009)), from this perspective, the current study contributes to the development of the computational approach in linguistic typology. This research addresses the following typological properties oan predictions on Georgian:

- argument marking patterns are subject to paradigmatic changes, that is, changes across tense-aspect paradigms;
- the relations of the pronominal markers with the arguments are realized through assigning two case roles, restricted to the context of the paradigmatic characteristics of the verb.
- the choice for the argument realization might be defined by the positions that the potential arguments occupy in the animacy hierarchy.

The evaluation of these predictions through grammar engineering includes the analysis of the different constructions in Georgian that feature polypersonal agreement, detection of

the semantic and morphosyntactic ambiguities that cause the complexity in the marking patterns, and disambiguation of the subject and object markers.

The outcome of my implementation generally supports my predictions on the analysis and the customization system performance. The grammar fragment captures the interaction between different levels of the language, disambiguates multiple meanings and provides an appropriate semantic and syntactic analysis for each of them, also proving that my hypotheses were consistent with the grammar structure and the LinGO Grammar Matrix resources.

The evaluation of a linguistic hypothesis is a very important, but not the only goal of this research. The implementation of a new grammar explores and evaluates the sufficiency of the existing resources in the core grammar and the libraries of the Grammar Matrix, especially those that are responsible for morphosyntax and argument marking, i.e. the case and argument optionality libraries. The evaluation of the LinGO Grammar Matrix against the Georgian polypersonal agreement and the rich verbal inflection as such investigates plausibility of implementation of a complex morphology in the customization system. As a rule, external preprocessing resources, such as morphological analyzers, are involved in developing large scale grammars. However, morphological tools are not available for a significant number of languages. Furthermore, a morphological analyzer is not necessarily bidirectional, that is, not all morphological analyzers allow to generate strings besides parsing them. Finally, an output of the existing analyzer will not allow me to investigate the processs that I am interested in.

The implementation shows that the verbal inflection featuring intense interaction between slots can be successfully maintained by the LinGO Grammar Matrix: morphophonological interaction in Georgian verb is quite moderate and thus can be modelled through the customization system. The number of implemented morphemes has increased significantly in comparison to the descriptive verb structures due to the specificity of the customization system constraining capacities.

1.3. Outline of the thesis

There are many steps that have to be accomplished throughout the research in frames of grammar engineering for linguistic hypothesis testing: it starts with acquiring necessary grammar and research sources, defining the phenomenon and collecting language data. Next, the analysis has to be developed and implemented in the customization system: it requires a thorough understanding of the grounds, conventions and limitations of the engineering platform, as well as correcting and debugging the customization system output. The grammar fragment is then evaluated against a test suite. These steps are represented in the structure of the thesis in the following manner. Chapter 2 provides an overview of the relevant research on the Georgian language, and the necessary theoretical foundations in HPSG and MRS. Chapter 3 describes the phenomenon of polypersonal agreement in Georgian and the data. Chapter 4, first, discusses the LinGO Grammar Matrix and the ways for implementing morphology through the customization system. Second, it describes the first implementation steps - the choices for Georgian in the cus-

tomization system, and the description of the noun inflection. The implementation of the Georgian noun inflection illustrates the constraining patterns in the customization system. In Chapter 5 implementation of the Georgian verbal morphology is given: I argue for my analysis of the subject-object pronominal marking and its implementation. The additional work on the indirect object and animacy hierarchy support is also discussed in this chapter. The results of the evaluation are presented and discussed in Chapter 6. Finally, Chapter 7 summarizes the results of this study and describes potential future work.

1.4. A note on terminology

There is a number of grammatical terms that are specific for Georgian: the term *screeve* defines a tense-aspect verbal category in Georgian, however, I largely use *tense* for the sake of reader's convenience and consistency with typological research. The screeves or tenses constitute *Series* (same term for plural and singular).

Two names for one morpheme - pre-radical vowel and version vowel - are frequently used interchangeably in the research literature. The morpheme will be referred to as 'pre-radical vowel' in the thesis, nevertheless, the reader should not be confused with the corresponding gloss term. In the gloss, the morpheme is abbreviated as VER from 'version': the expected abbreviation PRV from 'pre-radical vowel' could not serve since another morpheme, preverb, has one conventional abbreviation PRV.

The literature on Georgian commonly uses a Latin transliteration of the Georgian script. I am following a transliteration scheme from (Gurevich, 2006). The abbreviations that appear in the thesis are listed in the Abbreviations section.

2. Theoretical background

This chapter introduces the relevant theoretical and computational accounts of Georgian grammar and presents some necessary essentials of HPSG and MRS theories that underlie the LinGO Grammar Matrix (Section 2.3). Section 2.4 describes the grammar engineering approach for linguistic hypothesis testing. The relevant aspects of these accounts will be discussed throughout the thesis.

2.1. Research on Georgian

2.1.1. Theoretical inquiries

There is significant theoretical research on Georgian: the early descriptive reviews date back to the 1870-s; however, the first fundamental Georgian grammar by Akaki Shanidze (1953) is still a commonly used resource in research and teaching. A recent detailed description of the Georgian verb proposed by Damana Melikishvili (2001) contrasts sharply with Shanidze's account. Melikishvili suggests a division of the Georgian verbs into three diatheses[1] motivated by the requirements for the derivational morphemes, whereas Shanidze's analysis is based on the syntactic properties, such as valency and types of subjects.

In western linguistics, the grammars by H. Vogt (1971), H. Aronson (1990) and G. Hewitt (1995) cover substantial fields of Georgian morphology and syntax. A number of works on the Georgian syntax have been considered in the current analysis. N. Amiridze focuses on such aspects of syntax, as reflexive and anaphoric constructions (Amiridze, 2005, 2006). The works by S. Skopeteas (Skopeteas and Fanselow, 2007, Skopeteas et al., 2009) investigate the informational structure of a Georgian sentence and the syntactic phenomena performing on a discourse level. A number of works address the Georgian verbal system, including research on particular aspect kind by A. Holisky (1981) and the argument relations by K. Tuite (1987, 1988, 1994, 1996, 2007, 2009). Two works by K. Tuite are particularly important for my research: his dissertation on number agreement (Tuite, 1988) provides important observations on the potential of the direct objects to trigger plurality marking, and on the role of animacy hierarchy in these processes; and his earlier work on case controllers (Tuite, 1984) influenced my implementation decisions.

Specific syntactic questions are thoroughly discussed in A. Harris (1981, 1982), W. Boeder (2002), and K. Vamling (1989). The account to Georgian syntax provided by A. Harris is very valuable for the present research: Harris approaches the Georgian syntax in terms of Relational Grammar (Blake, 1990) paying special attention to expressed and

[1] Each diathesis is subdivided in two groups, overall splitted in 68 verbal paradigms.

unexpressed arguments. A number of claims from A. Harris have not been proved in my data, as discussed in Chapter 3.

The difficulties in defining which argument is actually a subject or an object led to an intense discussion among researchers. Particular attention has been paid to the necessity to establish permanent parameters and statuses for arguments or a clear division between grammatical and logical argument. Proposals on the paradigmatic meaning of the argument roles in Georgian and an animacy hierarchy of arguments marking have been appearing as such argument parameters.

The work by Cherchi (1997) evolves on the interaction of the morphosyntactic structures and particular verbal categories, representing the merged nature of the grammatical layers in Georgian. O. Gurevich (2003, 2006) analyzes the role of the pre-radical vowel in Georgian in light of Construction Grammar (Goldberg, 1995). She approaches Georgian morphology arguing for a broader and less semantic understanding of a morpheme, related to word and paradigm approach to inflectional morphology, and strengthened by a special attention to psycholinguistic plausibility. Interestingly, Georgian complex morphosyntactic structures illustrate the analyses in major accounts for inflectional morphology: item and arrangement (Lieber, 1992), item and process (Steele, 2002), stem and paradigm (Stump, 2001), and distributed morphology (Halle and Marantz (1994) after Gurevich (2003)). M. McGinnis (1995, 1996, 1998)approaches Georgian morphosyntax from the generative perspective, arguing for disjunctive properties in the morphological competition in Georgian, and for a lexical turn in treating some syntactic phenomena. The findings on the scopal limitations of verbal agreement morphemes as expressed, among others, by (Gurevich, 2006), have a clear reference to the proposed analysis.

Besides this, Georgian data have frequently appeared in the literature on issues of general linguistics and typology, such as works on ergativity and morphological structure by S. Anderson (1976,1992), aspect and language universals by B. Comrie (1995), and others.

2.1.2. Computational models for Georgian

The first application of computational approaches to Georgian data goes back to 1967: Thamarashvili (1967) considers the issues of morphological analysis in Georgian machine translation. Antidze and Mishelashvili (2006)propose specific syntactic and morphological analyzers that can be applied to Georgian. The authors provide some details on the morphological rules for the Georgian verb: each processed morpheme is indexed with the slot of the verbal structure that it represents; in case of morphological fusion, the morpheme is tagged with two indices. The morpheme-slot correspondence is given in the lexicon; the root follows the same annotation, though some additional features are added. Though the evaluation results are unknown, this study is interesting at least because it implements the classification by Melikishvili (2001)rather than the classification in four classes based on Shanidze' works (which is also the most popular in the English literature on Georgian) by shedding light on advantages and disadvantages of both approaches.

The Georgian Grammar Project[2] (Meurer, 2009) consists of a morphological analyzer, LFG Georgian Grammar, and a demo Treebank. The Georgian Grammar Project is currently under development. In the morphological analyzer, all words are annotated with an LFG-lexicon reference number and morphosyntactic features. The reference numbers aim to disambiguate the possible meanings. The features for nouns are: case, number, double declension case and number, full or reduced case inflection, postposition and clitics. The features for verbs include tense/mood as well as person and number marking, represented through Subj and Obj features. The argument structure as well as conjugation are annotated for each lexical entry. Thus, grammatical functions are mapped through the verbal morphology, and most relevant information seems to be available. The lack of the other details on the verb is compensated by the size of the transducer's lexicon: 74.000 nouns and adjectives and 3.800 verb roots (Meurer, 2009).

Kapanadze (2009), like Meurer (2009), builds a morphological analyzer for Georgian on the base of Xerox finite-state morphology tools. Also as in (Meurer, 2009), flag diacritics are widely applied for memorizing and marking purposes in order to recognize and generate the circumfixes, which are quite frequent in the Georgian verb. The lexicon of this FST comprises much smaller number of entries (150 verbs), and has a specified system of lexical rules: they are either considered general characteristics of the verb form or are applied to particular morphological slots.

2.2. Theoretical foundations

HPSG and MRS The LinGO Grammar Matrix is developed on the theoretical grounds of Head-Driven Phrase Structure Grammar (HPSG) (Pollard and Sag, 1994) and Minimal Recursion Semantics (MRS) (Copestake et al., 2005). HPSG is a lexical highly constrained and sign-based grammar formalism: sign properties are expressed through types (Carpenter, 1992), consisting of features and their values, and the feature structures are represented as attribute-value matrices. Inheritance and unification operations play a crucial role in the way types function classifying them into subtypes and supertypes: more specific subtypes inherit features from more general supertypes. If the feature structures have compatible values, they can be unified: unification is the key operation to support grammatical relations in HPSG.

A sign in HPSG consists of a PHON feature with a phonetic representation of a sign, and a SYNSEM feature that comprises relevant syntactic and semantic information. SYNSEM distinguishes LOCAL and NON-LOCAL features with respect to the processed unit: LOCAL comprises syntactic features, such as part-of-speech and valence, expressed in CATEGORY, and semantic features, represented in CONTENT and CONTEXT.

The syntactic processes are expressed through a set of lexical and phrase-structure rules. Lexical rules are applied to words in lexicon and support derivation processes. Phrase-structure rules regulate the combinations of lexemes that are allowed in the language. Phrase-structure rules are supported by the immediate dominance (ID) schemata,

[2] http://maximos.aksis.uib.no/Aksis-wiki/Georgian_Grammar_Project

a. Every big white horse sleeps.
b. every(x, $\bigwedge$(big(x), $\bigwedge$(white(x), horse(x))), sleep(x))
c.

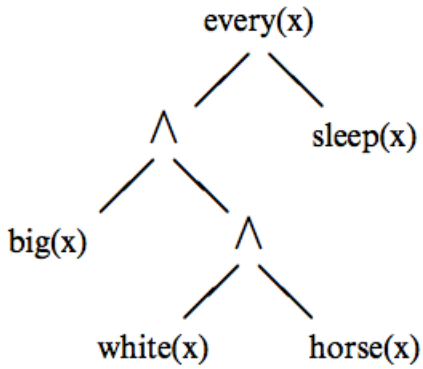

Figure 2.1.: An MRS representation (Copestake et al., 2005)

conveying the applicable constituent structures, and linear precedence (LP) statements that introduce the order of constituents.

MRS is a computational semantics framework; it is based on the typed feature structure formalism, so it can be easily integrated with HPSG. MRS is based on the grounds of semantic compositionality: it provides flat representations of semantic relationship between words in a sentence. The semantic relations are viewed in terms of predicates, labels, and arguments, consisting of individuals and events, which guarantees the support of scopal dependencies. An example of the MRS representation, including scopal dependencies, is presented in Figure 2.1.

LKB, TDL, and [incr tsdb()] are the platforms developed for grammar engineering and evaluation, and generally used with HPSG grammars. Due to the space limitations I describe them briefly below; all of the systems are freely available and are provided with detailed documentation, so an interested reader should follow the reference material.

Linguistic Knowledge Builder (LKB) LKB (Copestake, 2002) is a grammar engineering and lexicon development environment supporting the typed feature structures formalisms. Although considered theory independent, the LKB is used mostly for developing HPSG grammars. The LKB supports parsing and generation, allowing to evaluate and edit an existing grammar and create a new grammar employing types and feature structures descriptions for grammar modelling.

Type Description Language (tdl) TDL (Krieger and Schäfer, 1994) is a typed feature formalism and a declarative language used for writing grammars. As a grammar development environment, TDL consists of UDiNE feature solver and *tdl* language. The

language allows users to define type hierarchies composed of isolately processed type or feature constraints or to apply macros - the existing templates. It also distinguishes different types, such as avm, sort, and built-in types, and atoms. A transparent transfer from the feature structures to the conventional expression and use of multiple operators make *tdl* a visible and powerful development tool.

[incr tsdb()] [incr tsdb()] (pronounced 'tsdb++') (Oepen, 2001) is a tool for evaluating competence and performance of constraint-based grammars. It employs a profiling method that creates a profile - a unit of the system knowledge about a grammar in its current state applied to specific data. The competence is tested in terms of grammatical coverage and overgeneration, and the performance is evaluated through such parameters, as the number of readings, time and memory used, spurious ambiguity, phenomena represented in the test data.

LKB, TDL and [incr tsdb()] can be mutually integrated, which means that a grammar written in *tdl* can be loaded and explored in both LKB and [incr tsdb()]. The access to multiple development and evaluation resources amplifies the steps throughout the engineering cycle.

2.3. Grammar engineering for hypothesis testing

Grammar engineering through the LinGO Grammar Matrix for linguistic hypothesis testing has been recently applied in linguistic research. The growing interest is enhanced by several reasons: first, every implementation of a grammar requires a precise syntactic or morphosyntactic analysis which can be easily evaluated in progress by running a grammar. Second, the LinGO Grammar Matrix libraries capture phenomena in their typological diversity across many languages, thus offer a broad linguistic perspective for a user. Third, the grammar rules can be easily applied to the large data sets; finally, the LinGO Grammar Matrix structure allows to model interaction between different phenomena, which is of a particular value for a researcher.

Grammar engineering requires an engineer to provide a set of linguistically motivated and machine readable rules of a language, thus an analysis that comprises the rules has to be precise and consistent with an underlying system knowledge. The ability to test the analysis on a large grammatical coverage is also an attractive component of grammar engineering for hypothesis testing.

There are several studies accomplished through grammar engineering for linguistic hypothesis testing, all of them aim not only to verify the syntactic hypotheses, but to argue for methodological value of grammar engineering for linguistic tasks as well. Constructions with auxiliaries in Wambaya are analyzed through aux+verb cluster in (Bender, 2008, 2010); several aspects of verb phrase coordination in Turkish are analyzed in (Fokkens et al., 2009). Complex case and agreement phenomena are tested on the material of Hindi and Tagalog, as well as so-called direct-inverse languages, such as Cree and Fore (Drellishak, 2008, 2009). The ongoing work on phenomena coverage of the customization system includes evaluation through elaborative analyses on morphological

constraints (O'Hara, 2008) and argument optionality (Saleem, 2010) in many languages.

This emerging field opens a new perspective for linguistic typology and syntax research; furthermore, it serves a broader interaction between computational and theoretical linguistic accounts. However, it has to be noted that grammar engineering for linguistic hypotheses validation does not substitute, but facilitates the syntactic research. As shown in (Bender, 2010), grammar engineering allows for evaluation of precision of the analysis and its compatibility with the language data, but it does not neglect the linguist's input behind the analysis.

3. Description of the phenomenon and data

This chapter provides the necessary background in Georgian polypersonal agreement and the data used for my research. Section 3.1 presents the tense-aspect paradigms, case system, and verb classes in Georgian. Section 3.2 describes the argument representations in the verb form. Section 3.3 explains the morphological structure of the Georgian verb most commonly used in the previous research and addresses its benefits and drawbacks. Section 3.4 introduces the data - the development of the test suite, its collection and its analysis.

3.1. Introduction

There are two dimensions in classification of the Georgian verbal system - verb classes, or conjugations, and series, or sets of tenses. As we should see here and in the following chapters, they closely interact between each other in order to support polypersonal agreement.

Series are tense-aspect-mood paradigms that share verb stems and certain syntactic properties. There are three series in Georgian with the following tenses:

Series I (future/present): present, future, imperfect, conditional, present conjunctive, and future conjunctive
Series II (aorist): aorist, optative, non-negative imperative
Series III (perfect): present perfect (evidential), pluperfect

The tenses that constitute the series are called *screeves* (after Georgian *mc'krivi* 'row'); this term is frequently applied since Georgian screeves can specify modal and aspectual information, but do not necessarily bear temporal meaning.

There are four main verb conjugations, or classes, in Georgian, which can be defined in terms of morphological, syntactic, and semantic similarity. Harris (1981) suggests the following morphological criteria for the verb class groupings: word formation for future/aorist series and subject markers - third person singular and plural markers in future tense and third person plural subject in aorist tense (see Table3.1).

Syntactically, verb classes can be divided according to the case marking patterns for subject, object, and indirect object, chosen in every series[1] (Tables 3.2 and 3.3). The

[1] Another approach for classification of the Georgian case marking and agreement is in division into lexical classes of active and passive verb stems (Tuite, 1987). The underlying difference is in the case shift in some series that experience only active verbs. For the purpose of the current work I do not see yet the benefits of the lexical classification over the conjugation-based classification.

Table 3.1.: Verb classes morphology

	FUT/AOR	FUT 3SG/PL SUBJ	AOR 3PL SUBJ
Class 1	preverb	*-s / -es*	*-es*
Class 2	preverb/*e-*	*-a / -an*	*-nen*
Class 3	*i- -(eb)*	*-s/ -en*	*-es*
Class 4	*e-*	*-a*	-

Table 3.2.: Case marking patterns

	Subject	Direct Object	Indirect Object
Pattern A	ERGATIVE	NOMINATIVE	DATIVE
Pattern B	NOMINATIVE	DATIVE	DATIVE
Pattern C	DATIVE	NOMINATIVE	*tvis*-Oblique

ergative appears only in one pattern (A) which is applied in Series 2 to verbs in Class 1 and 3. Pattern B with the nominative and dative cases for the subject and the direct object, respectively, and pattern C, an inverted version of pattern B, dominate the case marking system.

Table 3.3.: Class/Series case marking

	Series I	Series II	Series III
Class 1, 3	B	A	C
Class 2	B	B	B
Class 4	C	C	C

Harris (1981, 130)

The verbs of Class 1 are transitive, with the verb forms per se marking the subjects and objects. The Class 1 verb *c'er* 'write' is shown below in Series (1a), Series 2 (1b), and Series 3 (1c).

(1) Class 1:

(a) Series I

deda	*c'erils*	*c'ers*
deda	c'eril-s	c'er-s
mother-nom	letter-dat	write-pres3sgsubj

'Mother is writing a letter'

(b) Series II

dedam	*c'erili*	*dac'era*
deda-m	c'eril-i	da-c'er-a
mother-erg	letter-nom	prv-write-aor3sgsubj

'Mother wrote a letter'

(c) Series III

dedas	*c'erili*	*dauc'eria*
deda-s	c'eril-i	da-u-c'er-i-a
mother-erg	letter-nom	prv-ver-write-perf3sg

'Mother apparently wrote a letter'

The verbs of Class 2 are intransitive and largely derived from the Class I verbs to build forms corresponding to the English passive (*dacers* - 'He will write it', *daicereba* - 'It will be written') or inceptive (inchoative) forms, marking a change of state (*galamazdeba* 'It will become beautiful') (Aronson, 1990, 61). The verb stem *yl/yal* 'get tired' in (2), however, is not derived from a Class 1.

(2) Class 2:

(a) Series I

deda	*iyleba*
deda	i-yl-eb-a
mother-nom	ver-tire-sf-pres3sg

'Mother is getting tired'

(b) Series II

deda	*daiyala*
deda	da-i-yal-a
mother-nom	prv-ver-tire-aor3sg

'Mother got tired'

(c) Series III

deda	*daylila*
deda	da-yl-il-a
mother-nom	prv-tire-ts-perf3sg

'Mother apparently got tired'

Class 3 incudes verbs, largely intransitive, referring to ongoing activities, often related to motion, weather, light and noise (Aronson, 1990). One of the verbs of this Class, *cxovr* 'live' has the following forms in the Series, as shown in (3a-c).

(3) Class 3:

(a) Series I

nino	*mosk'ovshi*	*cxovrobs*
nino	mosk'ov-shi	cxovr-ob-s
Nino-nom	Moscow-in	live-sf-pres3sgsubj

'Nino lives in Moscow'

(b) Series II

ninom	*mosk'ovshi*	*icxovra*
nino	mosk'ov-shi	i-cxovr-a
Nino-erg	Moscow-in	ver-live-aor3sgsubj

'Nino lived in Moscow'

(c) Series III

ninos	*mosk'ovshi*	*ucxovria*
nino-s	mosk'ov-shi	u-cxovr-i-a
Nino-dat	Moscow-in	ver-live-ts-subj3sg

'Nino apparently lived in Moscow'

(King, 1994, 93-94)

Finally, Class 4 verbs denote the "affective" meaning related with an emotion or sensation (*miqvarxar* 'I love you'). They are also called inverse or indirect after indirect case marking across all Series: the subject of these verbs bears the dative, and the direct object bears the nominative case (Table 3.2-3). The verb form, however, changes in different tenses: *uqvars* 'loves' in Series 1, *uqvarda* 'loved' in Series 2, and *uqvardia* 'has loved' in Series 3, as appears in (4).

(4) Class 4:

(a) Series I

gelas	*uqvars*	*nino*
gela-s	u-qvar-s	nino
Gela-dat	3subj-love-pres3sgobj	nino-nom

'Gela loves Nino'

(b) Series II

gelas	*uqvarda*	*nino*
gela-s	u-qvar-d-a	nino
Gela-dat	3subj-love-ts-3obj	Nino-nom

'Gela loved Nino'

(c) Series III

gelas	*uqvardia*	*nino*
gela-s	u-qvar-d-i-a	nino
Gela-dat	3subj-love-ts-perf-3obj	Nino-nom

'Gela apparently loved Nino'

The division in the classes is largely derivational, and incorporates a number of properties, such as voice and modality. Causatives belong to Class I, synthetic and analytic passives, along with passives of state and inceptives are in Class 2, and desideratives - in Class 4. The verb *dac'ers* 'he will write it' of Class 1 produces such forms, as a causative *daac'erinebs* 'he will make him write it' (Class 1), passive forms of Class 2 *dac'erili ikneba* 'it will be written', *dac'erili ikna* 'it got written', *ec'ereba* 'it will stand written'. *icek'vebs* 'he will dance' of Class 3 creates a causative *acek'vebs* 'he will make him dance' of Class 1, a Class 2 inceptive *acek'vdeba* 'he will begin to dance', and a desiderative *ecek'veba* 'he feels like dancing' (Harris, 1981, 261). In essence, the variation in case marking patterns is a key reason for the concept of a class grouping.

3.2. Argument representations and verb forms[2]

In Chapter 1, Georgian was introduced as a language with an intense semantic and syntactic load on the verb that is supported by language characteristics, such as optional arguments and free word order. The complex interaction between two different sets for object and subject marking in Georgian is regulated by morphosyntactic and syntactic rules. Additional number agreement rules, as well as other syntactic processes (e.g. object raising, passivization, reflexivization) and the specificity of the verb type, further determine the use of the object and subject markers on the verb form. This results in a number of phenomena discussed below.

3.2.1. Person and number agreement

Subjects, objects, and indirect objects mark the verb form according to the paradigms presented in Table 3.4-6.

[2]In this work, I will follow the definitions of initial and final terms - subject, direct object, indirect object - established in relational grammar and applied by (Harris, 1981) for the description of the grammatical relations in Georgian (rather than grammatical and real subject, or internal and external terms).

Table 3.4.: Subject marking

	singular	plural
1 Person	*v-*	*v- -t*
2 Person	*ø*	*-t*
3 Person	*-s/a/o*	*-en/es/nen*

Table 3.5.: Direct object marking

	singular	plural
1 Person	*m-*	*gv-*
2 Person	*g-*	*g- -t*
3 Person	*ø*	*ø*

Table 3.6.: Indirect object marking

	singular	plural
1 Person	*m-*	*gv-*
2 Person	*g-*	*g- -t*
3 Person	*s/h/ø-*	*s/h/ø- -t*

Initial observations of these sets show that, despite the presence of three marking sets, the ambiguity in verbal agreement is not resolved. First, the direct and indirect object marking sets are fully compatible for the first and the second person. Second, *-t* appears as a plurality marker for different persons in all sets, sometimes acting as a sole marker (e.g. third person plural indirect object). Third, the direct and indirect object in third person may be expressed by the same morpheme. Finally, ambiguity of agreement markers is increased by the following morphophonemic and syntactic rules.

Morphophonemic rules

There are three morphophonemic rules that are relevant for subject and object marker interaction, and all of them license different cases of morpheme blocking:

i. The morpheme v is deleted before the morpheme g

ii. The morpheme s or its variant h is deleted before the morpheme v

iii. The morpheme s is deleted before the morpheme t

Applied to the aforementioned marking paradigms, (i) implies that the first person subject marker is deleted in presence of the second person direct or indirect object marker. From (ii) follows that the third person indirect object marker is deleted if it is followed

by the first person subject marker. Rule (iii) determines that the third person subject marker is eliminated if there is a plurality marker. The realization of these rules leads to a higher ambiguity of verb forms, as illustrated by the multiple interpretations of the form *mo-g-kl-av-t*:

(5) (a): the subject marker *v* is blocked by the first part of the object marker *g*

mo-g-kl-av-t
prv-obj2-kill-sf-obj2pl
'I will kill you-pl'

(b) the subject marker *s* is blocked by the first part of the object marker *g*

mo-g-kl-av-t
prv-obj2-kill-sf-obj2pl
'He will kill you-pl'

(c) the first part of the subject marker *v* is blocked by the object marker *g,* the plurality marker belongs to the subject marking pair

mo-g-kl-av-t
prv-obj2-kill-sf-subj2pl
'We will kill you-sg'

(d) the first part of the subject marker *v* is blocked by the object marker *g,* the plurality marker belongs to the object marking pair

mo-g-kl-av-t
prv-obj2-kill-sf-obj2pl
'We will kill you-pl'

Syntactic rules

Tav-Reflexivization[3] is a syntactic rule derived by Harris (1981), it is concerned with the potential of person verbal agreement in a clause:

> To indicate the reflexive verb meaning, the reflexive pronoun *tav-* is used, and it is coreferent with the subject.

This means that a sequence of the first person subject and first person object marker in one verb form is excluded: the grammatical properties of the object are expressed by the appearance of a special pronoun *tav*, as in the example (6):

[3]I follow the graphics of the rules nominations as presented in (Harris, 1981)

(6)

is	*xat'avs*	*tavis*	*tvis*
is	xat'-av-s	tav-is	tvis
he-nom	paint-sf-subj3sg	himself-dat	

He is painting himself.

The rule extends to the second person subject prohibiting the verb form marking the second person object, and also applies when subject and object are not in the same number. This principle leaves only third person subject markers to appear with co-referent person object markers, which is practically not possible since the third person direct objects are expressed by zero morphemes.

The person marking constraints are also reflected in the number agreement rules described below.

Number agreement rules

There are two constraints that regulate number agreement in Georgian:

iv. First and second person nominals trigger Number Agreement in the verb of which they are final terms (i.e. they are marked arguments - I.B.)

v. A third person nominal triggers Number Agreement in the verb if it is its final term and if it is the first subject of that verb (in the same clause - I.B.) that is the final term, and it is not outranked by the first or second person nominal in this clause. (Harris, 1981, 219)

The examples below clarify possibly vague formulations from Harris (1981):

(7)

turme	*st'udent'ebs*	*gamougzavniat*	*gela.*
turme	st'udent'-eb-s	ga-mo-u-gzavn-i-a-t	gela
apparently	student-pl-dat	prv1-prv2-subj3-send-perf-obj3-pl	gela-nom

'Apparently the students (have) sent Gela'

(Harris, 1981, 217)

Here *st'udent'ebs* has a plurality marker *t* on the verb form, since there is no first or second person nominal that will suppress it, and it is the first subject of the verb *gamougzavniat*. The next example illustrates how the second person object marker outranks the final term subject, resulting in a form without a plural marker:

(8)

turme	*st'udent'ebs*	*gamougzavnixar*	*(shen)*
turme	st'udent'-eb-s	ga-mo-u-gzavn-i-xar	
apparently	student-pl-dat	prv1-prv2-subj3-send-perf-obj2sg	

'Apparently the students (have) sent you-sg'

(Harris, 1981, 218)

Overall, the number agreement rules along with the morphophonemic and syntactic rules shape the person hierarchy of argument marking with a preference for the first and second person over the third person arguments, as further discussed in the proposed analysis in Chapter 5. The argument roles that stay behind the marked persons play a crucial role in another designed analysis, which will be introduced in Chapter 7.

3.3. Morphological structure of the Georgian verb

As seen from the observations in the previous section, an agreement marker can be either blocked by a following morpheme or outranked due to the specificity of a syntactic construction, leaving certain arguments unmarked. In both cases, the task of modelling agreement properties should rely on building a system of morpheme-to-morpheme and morpheme-to-argument relations. This system requires a clear verb morpheme structure.

Hewitt (1995) suggests the following ordering of the verb morphemes:

Position	Morpheme
1.	Preverb I
2.	Preverb II
3.	Preverb III
4.	Prefixal pronominal marker
5.	Pre-radical vowel
6.	Root
7.	Passive marker
8.	Thematic suffix
9.	Causative
10.	Extension marker
11.	Screeve marker
12.	Suffixal person marker

The preverbs are distinguished on the basis of their semantic properties, and extension and thematic suffixes are derived according to their grammatical function. The table below shows the morpheme distribution of the verb *dagac'erinebdat* '(s)he would make you(pl) write it' (Gurevich (2006, 92) after Boeder (2002)):

Value	Morpheme
da	Preverb
g	Pronoun Marker1
a	pre-radical vowel
c'er	*write*
in	Thematic Suffix
eb	Existential Marker
d	Screeve Marker
a	Pronoun Marker2
t	Plurality Marker

The verb structures introduced in (Hewitt, 1995) and modified in (Boeder, 2002) de-

termine the pronoun markers by the suffix and prefix position only. As such, they do not distinguish them according to their argument function or case assigning. The complexity of the argument markers and their relations require a more precise structure which I will argue for in Chapter 5. However, I refer to this pronoun-based approach in the implemented analyses, and refine it with the functional division of the markers in another suggested analysis.

Since the implementation moves from the morphosyntactic representations to the feature properties, it is particularly important to determine the boundaries of the morphemes. The roles of derivational and inflectional morphemes in the verb structure are quite problematic in Georgian. The Georgian preverb is an example of how inflectional and derivational functions can be merged. Preverbs originate from prepositions, and some of them not only define the semantic scope of the verb's meaning, but add restrictions on the verb's grammatical functions, such as tense, and argument's relations. For example, preverbs *mi* and *mo* in *mic'era/moc'era* 'write' interact with the indirect object: first and second person require *mo*, third person requires *mi*.

(9)

vanom	*mogc'era*	*c'erili*
vano-m	mo-g-c'er-a	c'eril-i
Vano-erg	prv-2sgobj-write-aor3sg	letter-nom

'Vano wrote you a letter'

vanom	*misc'era*	*c'erili*
vano-m	mi-s-c'er-a	c'eril-i
vano-erg	prv-3sgobj-write-aor3sg	letter-nom

'Vano wrote him a letter'

(Harris (1981, 33), my gloss)

Thematic, or existential, suffixes have the same functionality: they are stem-specific, require for verb class derivations and carry temporal-aspectual meaning. A more difficult example refers to the capacity of a pre-radical vowel to express a reflexive meaning and a presence of an indirect object, among many other processes comprised in the Georgian version (more on this in Gurevich (2006)) . An indirect object can be expressed through two different constructions in Georgian. Consider the example:

(10)

gelam	*shekera*	*axali*	*sharvali*	*merabisatvis*
gela-m	she-ker-a	axal-i	sharval-i	merab-is-a-tvis
Gela-erg	prv-sew-aor3sgsubj	new-nom	trousers-nom	Merab-gen-ext-for

'Gela made new trousers for Merab.'

gelam	*sheukera*	*axali*	*sharvali*	*merabs*
gela-m	she-u-ker-a	axal-i	sharval-i	merab-s
Gela-erg	prv-ver-sew-aor3sgsubj	new-nom	trousers-nom	Merab-dat

'Gela made new trousers for Merab.'

(Harris, 1981, 91)

In the first example the indirect object 'for Merab' is expressed by a noun in the genitive case with a postposition 'for' and no object or indirect object markers in the verb form. In the second example, the indirect object is expressed by the dative form of the noun and a version vowel *u.* A similar pattern is applied to the reflexive constructions: they are built either with a reflexive pronoun followed by *tvis,* or with a pre-radical vowel on a verb form. The constructions with *tvis* are called external, and with a change of a pre-radical vowel - internal. I will return to the issues with indirect objects constructions in the description of the observations on the test suite presented in the sections below; an updated structure of the verb slots is suggested in Chapter 5.

As we could see, even though direct and indirect objects are involved in agreement, verbs mark only one object on a verb form. The second object, direct or indirect, is either outranked by multiple blocking techniques or marked by the semantics of the verb class. There is no direct mapping between a single syntactic process, such as case marking, word order, anaphoric noun phrase, and the role of a subject, object, or indirect object: arguments are represented through a number of grammatical functions. This plurality of grammatical relations encoded in a morpheme and application of blocking techniques leads to appearance of fused, zero and portmanteau morphemes. Arguments can be marked by deleted or ambiguous morphemes, so a single subject-object marking pattern can refer to multiple semantic-syntactic representations of arguments. In addition, argument roles may change throughout syntactic transformations in multiclausal constructions, object raising, indirect speech. The semantic properties of verbs are extremely influential and lead to mismatches with morphosyntactic characteristics, such as a conflict in semantic number and verbal number. Furthermore, the scope of verbal transitivity evolving in the Georgian verbal system includes both direct and indirect objects.

3.4. Test suite

This section describes the data used in my study: the development of grammatical and ungrammatical sets of the sentences (Section 3.4.1), their analysis (Section 3.4.2), and some observations on the phenomena that appear to be contradictory to those discussed in the research literature.

A so-called MRS test suite was originally developed for the representation of some Minimal Recursion Semantics phenomena in the English Resource Grammar and is frequently used for evaluation in the LinGO grammars. A Georgian translation of the MRS test suite is obtained, however, it does not fully suit the purposes of the current analysis, so a new test suite tailored to capture agreement features had to be developed.

3.4.1. Test suite development

The test suite for polypersonal agreement aims to capture the key properties of the agreement, such as differences in agreement marking for the verbs across all classes;

when possible, changes of one stem across the verb classes (e.g. *c'er* and *mal* examples in 3.1); variations across tense-aspect paradigms with the consequent changes in the case-marking patterns; the subject-object marking with overt and covert arguments; and different word order constructions.

In addition, the test suite demonstrates changes in a verb structure caused by an indirect object, by an adverb with locative or temporal meaning, in causative constructions, and cases of semantic and grammatical mismatch, such as verbal and semantic number conflict. Most of these changes demand modifications of pre-radical vowels and preverbs.

The main requirement for the test suite development is to obtain valid and representative data. Native speakers of Georgian were asked either to translate a sentence from English or Russian into Georgian, or to judge the grammaticality of a construction in Georgian. The latter was particularly convenient for checking examples from the research literature. In order to obtain constructions that allow multiple expressions (e.g. external and internal indirect object) I asked the native speakers both to translate the sentence from English into Russian and to provide the judgements on the various constructions. The native speakers were asked to write in Georgian script, which was later transliterated.

Twelve native speakers aged 24-50 consulted me in the test suite development. All of them reside in Tbilisi, Georgia. It should be noted that they are fluent in Georgian and were born in Georgia, and three of them learned Russian as their first language, beginning to speak Georgian at age of six. This is an interesting sociolinguistic fact which is not, unfortunately, specifically addressed in this research: the feedback provided by the bilingual speakers was double-checked with the monolingual native speakers and proved to be valid.

Ungrammatical examples are of a high value in the test data for grammar engineering: they are essential to demonstrate constraining capacities of a grammar. Ungrammatical examples for Georgian polypersonal agreement include the verbs with the shuffled morpheme order and inappropriate or conflicting argument and tense markers, and the arguments showing erroneous case marking.

The output from the native speakers was transliterated, split into morphemes, glossed, and translated. Each entry in the test suite includes a short description of the example, the information on the source of the example (author or research literature), the native speakers' judgement (`g` stands for grammatical, and `u` for ungrammatical), and the type of the phenomenon that the entry exemplifies (`agr` is for agreement). 'Vetted' parameter signals whether the example was checked with native speakers (`s` means that it was checked). A sample entry from the Georgian test suite is presented in Figure 3.1.

3.4.2. Analysis and observations

The test data was split into subparts according to the scope of each step of the implementation. Simple clause constructions with transitive and intransitive verbs constitute the basic test suite for the implementations from the customization system. The second test suite comprises sentences with indirect objects and different modes of transitivity, and the third test suite captures the relations relevant for the animacy hierarchy hypothesis.

```
  # Ex 1 demonstrates verb subject-object agreement; 'love' is the
verb of IV Class, notable for the inverted marking:  -m- is normally
an object marker, and -s- - a subject marker
  Source:  author
  Vetted:  s
  Judgement:  g
  Phenomena:  agr
  me miqvars c'igni
  me m-i-qvar-s c'igni
  I.Dat SUBJ1SG-OBJVER-love-OBJ3 book.Nom
  I like the book
```

Figure 3.1.: A sample entry from the Georgian agreement test suite.

They were merged at the final stage of implementation.

The data analysis starts with retrieving and classifying morphemes into slot positions, then detecting the pattern of argument marking and the way it is expressed in the morphemes. At this step it is extremely important to determine whether any grammatical functions are expressed in blocked or fused morphemes. The outcome of the data analysis generally goes in line with the research findings discussed earlier in this chapter. Nevertheless, there are some contradictory results that are described below, and the general morphological analysis for noun and verb inflection is presented in Chapters 4 and 5 respectively.

The first observation concerns the superessive vowel. Aronson (1990) and Gurevich (2006) indicate that a pre-radical vowel *a* is inserted in a verb if there is a reference to a surface on which an action is performed. For example, it distinguishes the forms *c'ers* 'write something' and *ac'ers* 'write something on something', and the latter form appears in such contexts as below:

(11) *p'ropesorma akhali sit'qvebi dapas daac'era*

The professor wrote the new words on the balckboard. Aronson (1990, 372)

However, this sentence, with and without an indirect object of a surface meaning, is translated by native speakers without a superessive vowel *a*:

(12)

masc'avlebeli	*davalebis*	*dapaze*	*c'ers*
masc'avlebel-i	davaleb-is	dapa-ze	c'er-s
teacher-nom	task-gen	board-on	write-subj3sg

'The teacher writes the task on the board.'

(13)

masc'avlebeli	*mosc'avleebisatvis*	*davalebis*	*dapaze*	*c'ers*
masc'avlebel-i	mosc'avle-eb-is-a-tvis	davaleb-is	dapaz-e	c'er-s
teacher-nom	student-pl-dat-ext-for	task-gen	board	write-subj3sg

'The teacher writes the task to the students on the board.'

These sentences have also triggered another observation that contradicts the description of indirect object agreement. In Section 3.3, the change of preverbs with respect to the person of the indirect object is introduced. Unlike this observation, my test data have not demonstrated such a function of the morpheme: in the example below, the indirect object *mosc'avleebisatvis* 'for the students' does not cause a preverb change in the verb form.

(14)

masc'avlebelma	*davaleba*	*mosc'avleebisatvis*	*dac'era*
masc'avlebel-ma	davaleb-a	mosc'avle-eb-is-a-tvis	da-c'er-a
teacher-erg	task-nom	student-pl-dat-ext-for	prv-write-subj3sg

'The teacher wrote the task for the students.'

Yet another remark is related to the appearance of an indirect object in a construction. Harris (1981) derives the Object Camouflage syntactic rule, it plays one of the key roles in defining term relations in Georgian. It states that "in the presence of an indirect object in a clause, first or second direct objects are realized as a possessive pronoun and *tavi*. The possessive pronoun will then bear the information on the person and number of the object" (Harris, 1981, 51), and works as in the example below:

(15)

vano	*chems*	*tavs*	*adarebs*	*givis*
vano	chem-s	tav-s	adar-eb-s	givi-s
vano-nom	my-dat	self-dat	compare-sf-subj3sg	givi-dat

'Vano compares me to Givi.'

Harris (1981, 49)

Here 'me' is expressed in a compound form *chems tavs* and means 'myself' verbatim. However, the native speakers provided different translation to this and similar constructions with indirect objects, where the first or second person object is expressed with a personal pronoun in the dative:

(16)

vano	*me*	*mas*	*madarebs*
vano	me	mas	m-adar-eb-s
vano-nom	me-dat	him-dat	obj1sg-compare-tf-subj3sg

'Vano is comparing me to him.'

The construction with reflexive and possessive pronouns as defined in the rule appears in my test suite only in the sentences with the reflexive subject, such as 'I compare myself to Vano', and is valid for all person values. In general, the ditranstitve constructions show quite inconsistent patterns for object marking including variations in singular and plural marking (*madarebs* - *gvadarebs* with *gv* marking object first plural), and the appearance of third person plural object instead of the second singular one. More data are definitely needed to investigate this aspect of polypersonal agreement.

Finally, in translation tasks the native speakers tended to avoid the present perfect series (present perfect and pluperfect screeves) and to use aorist and (less frequently) imperfect instead. Thus the examples for this series are borrowed exclusively from (Aronson, 1990).

These observations challenge the questions of the scope of the syntactic transformations that influence the verb morphemes: as we could see, the appearance of an indirect object or a location reference did not lead to the changes claimed in the research literature. These facts definitely require a further exploration which is left out of the scope of the current research.

4. Implementations in the LinGO Grammar Matrix

This chapter introduces the LinGO Grammar Matrix (Section 4.1) and focuses on one of its parts, the customization system (Section 4.2) - the platform for the implementation in this research. In this section, the steps for creating a grammar - filling out a questionnaire, validation and creation of a grammar - are described. Particular attention is given to the Lexicon section, where the agreement features have to be specified. Section 4.3 highlights my investigation on the interaction of the input and constraint specifications in the lexicon, and their effects in the output grammar. Section 4.4 provides the choices in the customization system made for Georgian; in particular, the application of inflectional and constraint rules to the Georgian noun system is presented in Section 4.5.

4.1. Overview

The LinGO Grammar Matrix (Bender et al., 2002,2010) is a multilingual grammar engineering environment that develops starter grammars based on a user-linguist input in the web-based questionnaire. The grammars created through it are lexicalized and precise[1], and they aim to capture deep linguistic analysis. The description of the language properties is based on two principles - support of language features that might be universal, especially in terms of a use in an HPSG grammar, and widespread phenomena that are more tailored to specific linguistic cases, or phenomena whose particular behavior differs across languages. The former are implemented in the core matrix, and the latter in the libraries.

The core matrix, or *matrix.tdl*, comprises the descriptions of the types and the basic type hierarchies that are considered to be cross-linguistically useful. The core type hierarchies outline the head-phrase combinations with arguments and adjuncts and build links for the appropriate semantic and syntactic descriptions and relations between arguments (Bender, 2008). Libraries provide the type description and the linguistic analysis for particular phenomena (one phenomenon per library) that might be common in many languages. Gender and case are the most typical examples: they are extremely common, although not all languages make use of them. Also, case systems, along with agreement patterns, vary a lot across languages.

There is no agreement library in the customization system. It defines agreement across libraries, supporting two types: between verbs and their arguments, and between determiners and nouns. Agreement between verbs and arguments can be either specified on a

[1]By this I refer to the preference for precision over recall in developed grammars.

HEAD or INDEX of an argument or a verb; the choice for HEAD or INDEX depends on whether the feature is semantic or syntactic. Agreement between determiners and nouns is specified on the SPEC list of the determiners. Agreement is described in the lexical types or inflection sections of the lexicon.

The tight interaction between the core matrix and the libraries defines both user's experience with the customization system and output grammars, as discussed below.

4.2. Customization system

The key to creating a grammar through the customization system lies in filling out a web-based questionnaire[2]. In the following section I will describe the processes of providing an analysis, validating it and creating a grammar.

4.2.1. Filling out the questionnaire

The questionnaire consists of sixteen sections: General Information, Word Order, Number, Person, Gender, Case, Direct-Inverse, Tense and Aspect, Other Features, Sentential Negation, Coordination, Matrix Yes/No Questions, Argument Optionality, Lexicon, Test Sentences, Test by Generation Options. Four of them must be filled out (General Information, Word Order, Person and Case), and the rest can be left unspecified. At least one noun type, one transitive and one intransitive verb have to be added to the lexicon.

The way the phenomena are approached in the library influences the type of the user's input in every section. A user might be asked to choose one option out of many possible options, for example, in the Word Order section a user checks a box for one word order from a list, and answers yes/no questions on auxiliaries and determiners. This is valid for the other obligatory sections, Person and Case; the latter, however, allows a user to submit more cases in addition to the selected case-marking pattern. A structure of the Person section is shown in Figure 4.1. In other sections, such as Gender and Number, a user should create types for his own hierarchies for respective features. The Tense and Aspect section allows a user to choose among semantic tenses and create his own hierarchy, and to specify the syntactic tense-aspect properties. Hierarchies for additional features can be added in Other Features, and in the Direct-Inverse section, a user can introduce a grammatical scale for the features dependent on a grammatical function. For each scale entry a user can choose a feature and its value if it has been previously described in other sections, since they interact. Then a user can define the verb form, direct or other, that appears when this feature is realized on an argument.

It has to be noted that the user's analysis is restricted to, first, phenomena coverage and, second, to the approach to these phenomena realized in the libraries and in the interface. For example, the customization system does not support adjectives, adverbs or indirect objects, so a user cannot include them in the implementation in the customization system. These points go in line with the purpose of the customization system to provide a jump start for grammar engineering with a possible manual extension of a grammar,

[2] `http://depts.washington.edu/uwcl/matrix/customize/matrix.cgi`

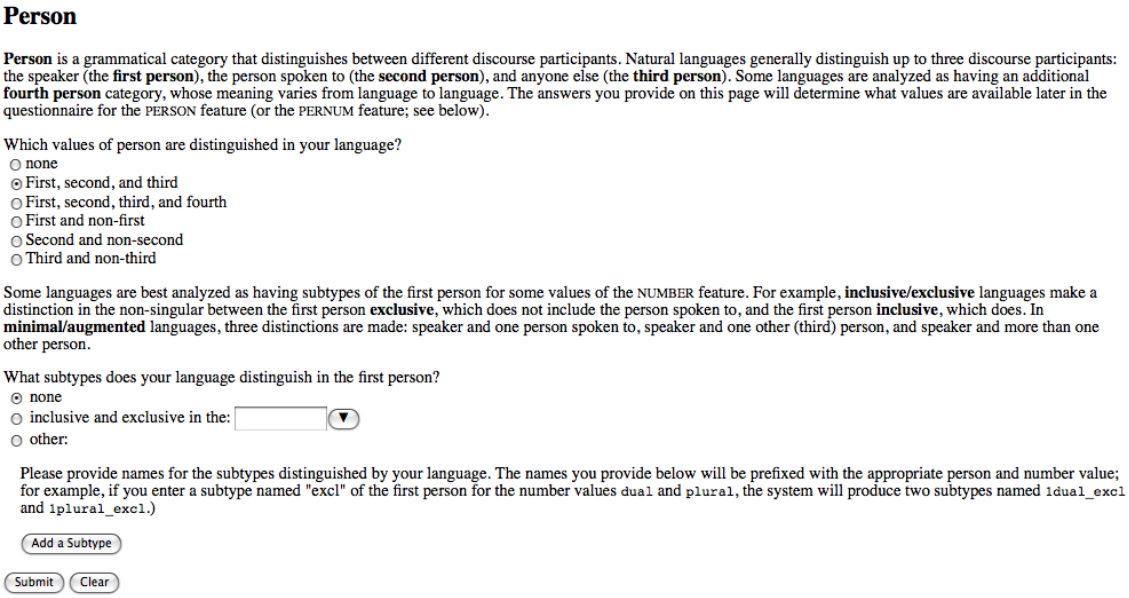

Figure 4.1.: The Person section of the customization system.

however, the customization system does not guarantee to provide full coverage for all phenomena.

Importantly, a user does not have to provide the full analysis at once. Even incomplete answers can be submitted: the system stores them and outputs as a 'choices' file. Choices can be saved and uploaded later. If needed, the choices can be changed manually in a text editor, but this requires a certain level of proficiency with the system. The ability of continuous work with the same choices is very valuable since filling out the questionnaire often follows changes in the analysis or test data.

4.2.2. Validation and creation of a grammar

In order to create a grammar, the user's answers have to be validated by the system. The customization system communicates its feedback on the user's analysis by means of an asterisk and a question mark. The asterisk signals contradictory or missing information: for example, repetitive names of features, types and slots, an inappropriate specification of a feature or an empty value for a compulsory feature. A question mark is called to flag ambiguous information, though it appears relatively infrequently.

The customization script takes the choices and produces a grammar downloadable as an archive. The grammar archive consists of a set of files, including *matrix.tdl* with the basic types that are identical to all grammars; *<name_of_the_language>.tdl* with language specific types and constraints defined by the user'; *lexicon.tdl* with lexical entries, types

and slots from Lexicon; *rules.tdl* with phrase-structure rules derived from the user's input; *irules.tdl* and *lrules.tdl* with lexical rule instance entries. Finally, *script* is the source for running a grammar in LKB, *choices* are the submitted answers for the customization system in a text format, and *test_sentences* are the sentences provided by the user.

4.3. Interaction between inputs and constraints

Since the LinGO Grammar Matrix produces lexical and constrained grammars, it requires a user to specify the linguistic properties and values through a system of lexical rules that have to be described in the Lexicon section of the customization system. In this section, I describe the Lexicon interface and discuss the strategies for choosing the input and constraints which proved to be efficient in constraining the rich Georgian inflectional morphology. The real-life application of these techniques to the Georgian data is demonstrated in Figure 4.8 and Section 4.5.

4.3.1. Lexicon

The Lexicon section asks a user to introduce noun and verb types, noun and verb inflection, as well as auxiliary verbs, determiners, and postpositions. A lexical type is defined by its own name, stems that are specified within it and their features. There are no restrictions on the number of stems within a type and their features, as long as they have been previously defined in the other sections of the system (e.g. person, case, number). For a verb stem, argument structure has to be defined for the whole type.

The inflectional part of the Lexicon is regulated by a system of slots and interaction between them. The most intuitive yet incomplete way of thinking about a slot is to consider it a morphological position: in the system slots are defined in terms of their inputs and constraints, and their relation might not be very feasible for a user. A slot in Lexicon is a constrained representation of a position: the position can be realized through multiple slots if the slots have different relations with other.

The definition of a slot starts with providing its name and deciding whether its presence is obligatory before or after a particular input. An input is another slot or a lexical type, and a user might choose as many inputs as possible, but the system allows to choose between 'before' and 'after' only once and to provide only one set of these many inputs. For example, prefix object markers can appear right after preverbs or before subject markers, pre-radical vowels, and roots. A user should first choose the direction of the specification: it could be either 'before', and then he should list the slots that follow prefix object markers, or "after", and then he chooses the preceding preverb slots.

Next, a user has to provide at least one morpheme with its own name and features. It is not possible to specify features for the whole slot, so all features have to be assigned for each morpheme. However, a slot does not require all morphemes to have the same features and values, so, for example, all possible case markers can be listed in one slot, and each of them will determine a specific case. Alternatively, the slot can include the morphemes with different values but with the same grammatical function.

When describing a slot in the Lexicon section, a user can choose whether the presence of this slot forbids or forces another slot or lexical type. The options of the lexical types for a verbal inflection slot include any verb, any auxiliary verb, any transitive or intransitive verb, and particular verb types and slots. At the moment of writing, the morphotactic system did not support the specification of 'forbids' or 'requires' due to a massive bug. In general, a user can choose multiple options for 'require' constraint by checking the boxes in a pop-up menu, which means that at least one of the slots or types is required. If a slot requires two or more other slots and types, one constraint per each requirement should be added. For 'forbid' one should add one constraint per each input.

The system disambiguates between slots on the basis of their inputs and constraints, and on the same grounds one can determine a set of constraints as an ultimate position. For example, there are two slots for case markers - one for vocal, and another for consonant stems. The slots accept different lexical types (let us assume that the noun types are divided in vocalic and consonant stems), the morphemes within them have different values, however, the similar case feature values. Thus, these slots occupy one position in the noun structure, although differ by the type of the input. This is a relatively straightforward example, whereas the implementation of complex morphologies requires building input and constraint relations between the slots rather than between the slots and the lexical types. The strategies for the input and constraint choices are discussed below.

4.3.2. Input Specification

Figure 4.2 illustrates the lexically motivated approach in determining the inputs for the morpheme slots. It is linguistically sound to say that a preverb *da* appears before the verb stem *c'er*; furthermore, it is extremely relevant that this particular preverb is related to this stem, since preverbs belong to the class of the lexically specific morphemes in Georgian. However, in this case only a verb form *dac'er* will be accepted: the relations of *da* with other morphemes, such as *g* and *a* in the example, are left underspecified. This is valid for every morpheme slot regardless of the occupied position with only one exception: the slot that is never separated from the stem by another morpheme must take the stem as its input. For the Georgian data, it cannot be generalized for the slots occurring immediately after the stem, since there is no verb stem in the test suite that appears with suffixes exclusively and does not have any corresponding verb forms with prefix inflection.[3]

Figure 4.3 demonstrates the input specification for a simple grammar with one stem and seven morpheme slots - four prefixes (A, B, C, D) and three suffixes (E, F, G). The arrows are oriented to the slots that are taken as inputs. The first technique (i) offers a consequent application of the lexical rules: slot A takes slot B as an input, slot B appears before slot C and so on until slot D - the closest to the stem - which takes the stem as its input. This order is established through choosing 'before' or 'after' and consequent input for each slot in the Lexicon. The first suffix E takes the furthest slot from the stem as

[3]The constructions with the suffix inflection only are relatively infrequent in the data set and constitute 11.8% out of the grammatical verb forms.

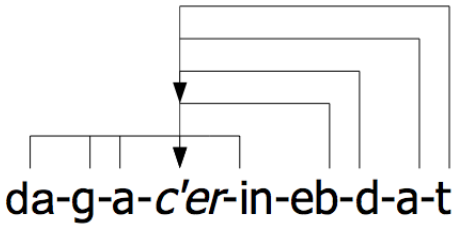

Figure 4.2.: Insufficient input techniques

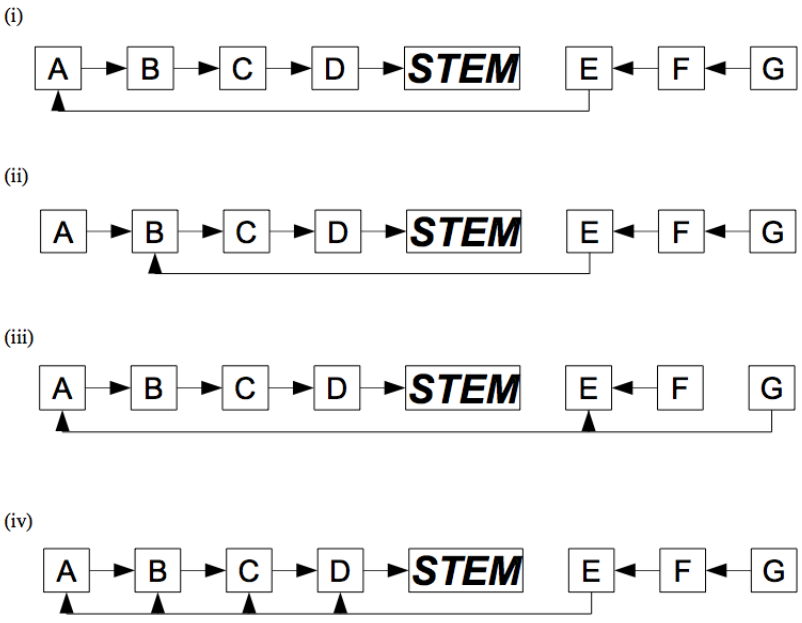

Figure 4.3.: The tested input models.

```
slot_a-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
slot_e-rule-dtr & slot_f-rule-dtr & slot_g-rule-dtr &
    [ DTR slot_a-rule-dtr ].
slot_b-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
slot_a-rule-dtr & slot_e-rule-dtr & slot_f-rule-dtr & slot_g-rule-dtr
&
    [ DTR slot_b-rule-dtr ].
slot_d-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
slot_a-rule-dtr & slot_b-rule-dtr & slot_c-rule-dtr & slot_e-rule-dtr
& slot_f-rule-dtr & slot_g-rule-dtr &
    [ DTR slot_d-rule-dtr ].
```

Figure 4.4.: Lexical rules resulting from input specification in Lexicon: inheritance.

an input, and all consequent suffix slots, F and G, follow E - they appear after E. This results into following lexical rules in the customized *tdl* grammar.

Since all the slots in our dummy grammar are optional, the inheritance relations in the lexical supertype rules guarantee that the information on the input of one slot is passed to the next slot for which the former could be related as an input, as illustrated in the lexical rules in Figure 4.4. The morpheme-a-lex-rule-super inherits from slot-e-rule-dtr, slot-f-rule-dtr and slot-g-rule-dtr : thus morphemes from slot-e, slot-f and slot-g can take a verb bearing a slot-a morpheme as an input.

The right-hand side of the lexical rules indicates the slots, or other rules in the grammar, for which it serves as an input. On the inheritance grounds, slots F and G are listed in the specification for the slot A and further inherited by B. The relation between D and such slots, as A, B, E or G is not explicitly established, but also inherited through the chain of the inputs.

The first input model allows the grammar to analyze the strings with every possible combination of the covert morphemes,so it parses the forms with the maximally possible number of the covert slots. The input models (ii) and (iii) in Figure 4.3 demonstrate the disjunction operation: in (ii) slot E takes B as its input which results in the disjunction of the lexical rule daughters of slot A and E , as illustrated in the slot B lexical rule below. The following lexical rules, such as the slot C rule, will choose for processing the slot A or slot E with the other lexical rules that they inherit (`slot_e-or-slot_a-rule-dtr`). This input model will license two parsing patterns: one for slot A and a stem with optional prefixes, and another for slot B occurring with the stem and possibly other prefixes C and D and suffixes. Slots A, on one side, and E, F, G, on another side, cannot appear in one form (Figure 4.5).

The input model (ii) manipulates the choice for 'before' and 'after' in the customization

```
slot_b-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
slot_e-or-slot_a-rule-dtr & slot_f-rule-dtr & slot_g-rule-dtr &
    [ DTR slot_b-rule-dtr ].
slot_c-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
slot_b-rule-dtr & slot_e-or-slot_a-rule-dtr & slot_f-rule-dtr & slot_g-rule-dtr
&
    [ DTR slot_c-rule-dtr ].
slot_a-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
    [ DTR slot_e-or-slot_a-rule-dtr ].
```

Figure 4.5.: Lexical rules resulting from input specification in Lexicon: disjunction.

system: slot E appears after slot B, and slot A before slot B. The input model (iii) tests the application of the 'before' arrangement only: slot F appears before slot E, and slot G appears before slot E, which also leads to disjunction, as in (ii). If inputs are specified in this way, the forms with either F or G can be parsed, and F and G cannot co-occur.

In (iv) slot E takes not only slot A as its input, but slots B, C, and D as well. However, these multiple inputs seem to be redundant due to the the inheritance of the lexical rules as demonstrated in (i). The input model (iv) does not change the performance as compared to the model (i).

4.3.3. Interaction between constraints

Lexical rules in a grammar obtained through the customization system can be constrained in terms of their relations with other rules and/or the role in creating inflected (lexeme) or non-inflected (word) forms. In this section, I describe how the morphotactic constraints 'require' and 'forbid' work, the effect of obligatory slots, and the feature interaction. I use the same simple test grammar as in the section above.

The 'require' and 'forbid' constraints are operated by means of flags (as opposed to the previous version with 'track' feature (O'Hara, 2008)). To see how the 'forbid' constraint works, assume the input model (i) (Figure 4.1) is adopted, and slot B forbids slot E. The resulting lexical rule for B is shown in Figure 4.6.

The lexical rule for the slot B supertype inherits the lexical rules for slot E, which would guarantee that the forms with slots F and G will be parsed even if E is blocked or omitted. With a rule specified this way, the forms where B and E do not co-occur will be parsed successfully. When one slot requires another slot, the patterns for inflectional flags application are the same: if slot B requires slot F, a form with B and without F is not parsed. If a slot forbids a lexical type, the constraint is irrelevant to the immediacy of the application: the requirement will be passed to all other slots that may appear between a forbidding slot and a lexical type. Also, the Lexicon page allows a user to

```
slot_b-lex-rule-super := add-only-no-ccont-rule & infl-lex-rule &
slot_a-rule-dtr & slot_e-rule-dtr & slot_f-rule-dtr & slot_g-rule-dtr
&
    [ DTR slot_b-rule-dtr,
      INFLECTED.SLOT_E-FLAG + ].
```

Figure 4.6.: Lexical rule resulting from 'forbid' constraint application: slot B forbids slot E

```
transitive-verb-lex := verb-lex & transitive-lex-item & slot_a-rule-dtr
& slot_b-rule-dtr & slot_c-rule-dtr & slot_d-rule-dtr & slot_e-rule-dtr
& slot_f-rule-dtr & slot_g-rule-dtr &
    [SYNSEM.LOCAL.CAT.VAL.COMPS < #comps >,
     ARG-ST < [ LOCAL.CAT.HEAD noun ],
       #comps &
       [LOCAL.CAT [ VAL [ SPR < >,
         COMPS < > ],
         HEAD noun ] ] >,
     INFLECTED.SLOT_G-FLAG - ] .
```

Figure 4.7.: Lexical rule resulting from 'obligatorily' specification: a transitive verb form is parsed only with slot G.

select multiple slots and types for 'require' constraint, which means that *at least one* of them has to appear with a slot that applies this constraint. To 'forbid' a slot or a type, a user has to provide one constraint per each item that is to be forbidden.

When defining a slot, a user can choose whether the slot appears obligatorily with respect to an input. The specification on the lexical rule is similar to the 'require' constraint. However, it extends the inflected parameters to the whole lexical type rule that is inherited by this slot. If slot G is specified as obligatory after slot F, it does not require slot F to appear every time it appears itself, but the verb form will not be parsed unless slot G is present (Figure 4.7).

Selection of the feature values is the most practical and necessary step in constraining the grammar. The specific choices for the Georgian polypersonal agreement are discussed in Chapter 5, where the interaction of multiple feature values specified on the same target is discussed: it is a particularly relevant issue for Georgian morphology since tense, object, and subject markers can be very ambiguous and allow multiple meanings for one morpheme value.

If slot A specifies a tense on a verb with a value of a whole tense hierarchy (e.g. *sr1*,

which includes present, future, conditional, imperfect and present and future conjunctive), and slot F has a tense feature with a value that is out of the hierarchy *sr1* (e.g. aorist), the form with both A and F cannot be parsed due to conflicting feature values. However, if slot G restricts the tense value to imperfect only, the verb form is analyzed successfully. Thus refining feature values on the same form are accepted by the system, unlike the conflicting feature values which, quite predictably, are not processed.

4.3.4. Summary

In this section the steps of morphological description in the customization system are shown on the example of a dummy grammar. Through this example, the role of inputs and constraints that are necessary for the morphemes, is emphasized. The specified inputs and constraints produce the inheritance (`and` operator) or disjunction (`or` operator) relations in the lexical types, thus supporting the morphemes' appearance in the verb. The constraints differ in creating inflected and non-inflected forms. It has to be noted that the ways the inputs and constraints are applied is not straightforward in the customization interface, so the tests on the dummy grammars are necessary; section 4.5 illustrates how they work on the Georgian material.

4.4. The Georgian grammar in the customization system

Section 4.2 described the customization system and its components, this section and the following show the solutions for the Georgian grammar implemented in the sections of the customization system.

Word order Georgian is a generally free word order language: the order is determined by the pragmatic and informational aspects, as well as the types of the arguments. The research literature often refers to a tendency towards a verb final position. In the test data verb final constructions are actually less frequent than the others, so in the base grammar I select the free (pragmatically determined) word order.

Number In the Number section a user defines his own hierarchy for number, and I define two types - singular and plural.

Person In this section, it is possible to use pre-defined combinations: a set for first, second, and third persons is chosen for Georgian. Another combination, third and non-third, could be an alternative: in the suffix positions, there is generally one marker for first and second person, one for third singular and one for third plural. However, this combination would not be valid for the markers in the prefix positions. This section asks about the subtypes of the first person, such as inclusive/exclusive. There are no such subtypes in Georgian.

Gender Georgian does not have a category of gender, so this section is left empty.

Case In the Case section a user might choose among core case marking patterns and, if necessary, add more cases. Structurally, the most suitable option for the Georgian case marking system is 'split conditioned on features of the verb': case marking in Georgian is, in general, a matter of tense-aspect variations across the verb class paradigms. The choices are the subject for intransitive verb (S), the subject of a transitive verb (A) and the object of a transitive verb (O). The 'split conditioned on features of the verb' pattern requires choosing from four cases : for A, for O, for S and O, and for S and A. Only one case can be used for each combination, and it cannot be repeated in other combinations. This option is therefore problematic for Georgian: the most frequent cases of subjects, objects, and indirect objects are the nominative and dative, and the ergative appears in one case marking pattern. In this perspective, the Georgian choices for this option should be:

S and A: nominative, ergative, dative
O: nominative, dative
A: nominative, ergative, dative
S and O: nominative, ergative, dative

Due to the aforementioned reasons, these answers are not acceptable in the system. When filling up this part, I have decided on the following cases:

S and A: nominative
O: dative
A: ergative
S and O: genitive

Specifying these 'dummy' cases is important for supporting argument structure feature, which is mandatory for the verb types. At this level the real proximity of the case choices remains irrelevant since the argument structure values include the extension for underspecified case. The choice for this case marking pattern is nevertheless important for further manual specification of the real cases.

Two additional cases are included in the case system in this section: instrumental and adverbial[4].

Direct-inverse The section is not filled in the main version and is left for the animacy hierarchy.

Tense and Aspect A user might select among common hierarchy elements or create his own hierarchy. The hierarchy for the Georgian grammar fragment is in line with a traditional view on the Georgian tense-aspect system: there are three types - series, sr1, sr2, and sr3, with a 'tense' supertype. The series are supertypes for the tenses. There are two additional types - pres_and_fut and fut_and_aor. They are defined to handle the tense features for two lexically predefined morphemes, the appearance of which signals

[4]Adverbial had to be called 'cadverbial' since 'adv' is a name of a type in the coordination section

these tenses: pres_and_fut is needed for a stem formant, and fut_and_aor is needed for a preverb. For example, the aorist screeve has two supertypes - sr2 and fut_and_aor. The latter's supertype is tense.

Other features Animacy is added as a semantic feature, with two values, anim(ate) and inanim(ate). I use it to mark the noun types and subject and object markers.

Sentential negation There are two types of negation offered in the section: inflectional negation that will make a boolean feature 'negation', or adverbial negation. The adverb type suits Georgian. I define it as an independent modifier of a verb (V) that appears before the verb, and provide spelling (*ar*). Another possibly appropriate object of negation could be a verb phrase (VP). The difference between V or to VP for negation can be tested, but it is not relevant for my purposes.

Coordination and Matrix Yes/No Questions The coordination and interrogative questions are out of the main topic of this research, so these sections are unspecified.

Argument Optionality The section includes two subsections, Subject Dropping and Object Dropping. In Subject Dropping the information about the conditions of dropping and the interaction between dropped or present Subject and subject markers has to be specified. Argument dropping is not lexically specific in Georgian and does not affect argument marking. Thus, subjects are defined to drop with any verb and in all contexts[5]. The next questions ask to provide more information on the appearance of the subject marker if the subject is covert and overt. Among possible options "required", "optional" and "not permitted", "optional" is chosen. Even though a subject marker is grammatically required on the verb, it is frequently not realized due to the blocking processes and syntactic changes. If the "required" option is selected, every potentially blocked marker has to include an empty pair, which likely leads to higher ambiguity and overgeneration.

The Object Dropping subsection asks only for information on the contextual and lexical determinacy of object dropping, and on marking with an overt object. The choice for "optional" is also valid when describing marking with an overt object, as well as "always allowed" for the lexical constraints.

Test Sentences 486 sentences are listed in the test sentences of the main grammar; the ungrammatical sentences are marked by an asterisk and are grouped after all grammatical sentences.

The solutions for my Georgian grammar in the Lexicon are described in two parts. In 4.5 the noun and personal pronoun implementation is described: it aims to demonstrate the representation in the Matrix of the basic morphological processes in the Georgian

[5]The section suggests an alternative for predetermined dropping: a user then should choose 'some verbs' and 'some contexts' that allow subject dropping.

pronoun	person	number	case	animacy
me	first	sing	nom, erg, dat	anim
chem	first	sing	gen	anim
chven	first	plur	nom, erg, dat, gen	anim
shen	second	sing	nom, erg, dat, gen	anim
tkven	second	plur	nom, erg, dat, gen	anim
is	third	sing	nom	-
igi	third	sing	nom	-
isini	third	plur	nom	-
man	third	sing	erg	-
mas	third	sing	dat	-
mat	third	plur	erg, dat, gen	-
mis	third	sing	gen	-

Table 4.1.: Personal pronouns in the Lexicon

grammar. The implementation of the verbal agreement is discussed separately in Chapter 5.

4.5. Georgian noun types and morphology in the customization system

This section illustrates the observations on the Lexicon structure (Section 4.2.2) and the constraining mechanisms in the Lexicon and output grammars (Section 4.3) on the example of the Georgian pronoun (Section 4.5.1) and noun (Section 4.5.2-3) system.

4.5.1. Personal pronouns

The first group of noun stems are personal pronouns which are used in a clause to contrast or emphasize subjects and objects of an action. Table 4.1 summarizes the information included in the personal pronoun description. The features specified on the lexical entries are person, number, animacy, and case. Georgian personal pronouns are generally not declined and do not appear in instrumental and adverbial cases. However, first person singular pronoun has two forms for case marking, and a demonstrative pronoun *is* or *igi* used to define the third person singular and plural, also declines. Lexical entries for first and second person pronouns are specified as animated, whereas for third person pronouns do not include the animacy feature. All personal pronouns have a predicate *_pron_n_rel*: it aims to distinguish them from the noun types.

4.5.2. Nouns

Two parameters - type of inflection and animacy value - are used in noun classification. There are two main case marking patterns in Georgian for vocalic and consonant noun stems. Division on the animacy level is relevant given my hypothesis on the role of animacy hierarchy in subject-object marking on a verb. Proper nouns such as personal names also undergo animate/inanimate classification[6]. For example, stems *ekim* (doctor), *model* (model) and *mxat'var* (painter) constitute the group of animated nouns with consonant stem; stems *deda* (mother), *givi* (Givi), *dzma* (brother) are in the group of animated nouns with vocalic stem. Stems *leks* (poem) and *ena* (language) are included respectively in the groups of inanimate with consonant stem and inanimate with vocalic stem. The features specified for each noun type are animacy, with either 'animate' or 'inanimate' values, and person, with a value 'third'. Overall, there are four noun types with 24 vocalic stems, 12 classified as animate and 12 as inanimate, and 36 consonant stems, 25 out of them are animate and 11 - inanimate.

4.5.3. Noun inflection

The structure of the Georgian noun is illustrated by the following construction *bichebistvis* 'for boys':

stem	plurality marker	case marker	extended case	postposition
bich	*eb*	*is*	-	*tvis*
boy	Pl.	Gen		for

The plurality marker appears first after a stem: it is the same for both consonant and vocalic stems, so it requires only one slot and takes all four noun types as its possible input. The plurality marker always precedes other noun markers, so the consequent slot, case marker, will take it as its input. Postpositions in Georgian govern particular cases, and the governing relations can be expressed through the 'require' constraint. Since constraints can be applied to slots only, the specific case slots are created according to the type of stem, vocalic or consonant, and the postpositions that they must co-occur with. For example, postposition *tvis* 'for', analyzed as a morpheme in the noun inflection, requires a genitive form that can be optionally extended with the *a* extended case morpheme[7]. I create two slots for Genitive case, one of them takes a slot with the plurality marker as its input and specifies the 'require' constraint on the consonant animate and inanimate noun types. The second Genitive slot has the same structure applied to the vocalic stems. A slot for the postposition *tvis* takes a slot for extended

[6] There is only one proper noun, a geographical nomination (parizh) that is assigned to an isolated noun type.

[7] Extended case, marked by a morpheme *a*, appears after Dative, Genetive, Instrumental, and Adverbial before some postpositions and in specific syntactic conditions, such as conjunctive constructions after *da* 'and', before the short form of the verb *aris* 'be', and in double declension constructions.

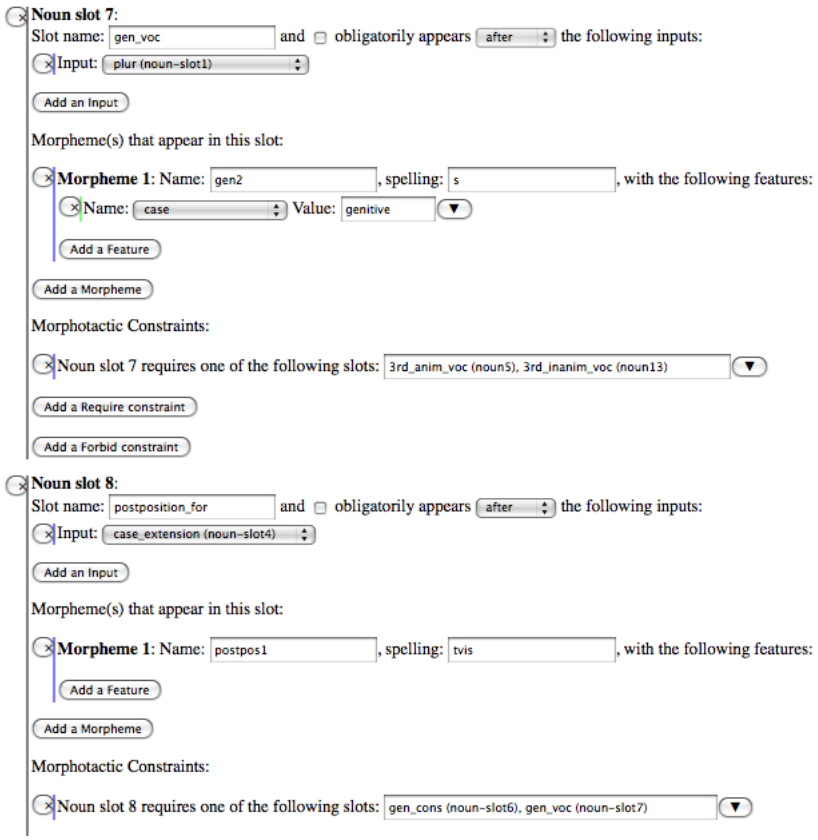

Figure 4.8.: Noun inflection slots in the Lexicon.

case as its input and sets the 'require' constraint for two Genitive slots. Figure 4.8 shows how this information is carried in Lexicon.

The rest of postpositions from the test data require a Dative zero morpheme: it is modelled as a separate slot with an empty spelling box and the Dative case feature value. This slot is compatible with both types of stems, so it does not apply any constraints. Other cases (Nominative, Ergative, Instrumental, and Adverbial) are represented in two slots for consonant and vocalic declension paradigms, with four morphemes in each, where one morpheme corresponds to one case. Variation within case morphemes in the paradigms is limited to the changes in Dative if postpositions *shi* 'in' and *ze* 'on' appear after the morpheme. In this case Dative will be marked by an empty morpheme (the so called bare stem Dative form). To handle such cases of zero-marking, I add another morpheme with the Dative case feature value and an empty spelling box.

4.6. Summary

The LinGO Grammar Matrix view the linguistic phenomena in two dimensions: the phenomena considered as generally applicable are described in the core matrix, and the phenomena studied as common, but not universal, or showing variations cross-linguistically, are implemented through the set of specific libraries. A user-linguist defines the phenomena relevant for a particular language in the sections of the web-based questionnaire. All parts of the LinGO Grammar Matrix resources are built on the base of the type descriptions and interactions, specific for HPSG, and the semantic relations through the predicate structure of MRS. Agreement is supported by the feature hierarchies and unification properties in HPSG; two particular agreement types, semantic and syntactic, are distinguished. The customization system maintains agreement between verbs and their arguments, and nouns and determiners. This chapter paid special attention to the correspondence of the slot and constraint chosen in Lexicon and resulting lexical rule. I described the choices for Georgian that provide a general overview of the grammar and the noun inflection rules. The following chapter is devoted to the discussion on verbal agreement, a proposal for its analysis, and the implementation.

5. Verbal agreement

Chapter 5 discusses a proposed analysis of the Georgian verbal agreement and its implementation through the customization system. I start with the description of the verb types (Section 5.1) and verbal inflection (Section 5.2), with a special focus on the pronominal markers distribution. The paradigmatic nature of the markers is underlined by the varied temporal meanings and expressed by the majority of the argument markers - the markers in the suffix positions and the combinations of the certain prefix markers with pre-radical vowels. The particular solutions in the implementation of the interaction between the argument markers (such as splitting the slots, adding empty morphemes, representing screeve values) are demonstrated in Section 5.3. In 5.4 I discuss the analyses for the indirect objects and animacy hierarchy.

5.1. Verb types

There are 27 verb types defined in Lexicon and 36 verb stems assigned to them. The key parameters for grouping verb stems into classes are their argument structure, verb class, and the types of lexically predefined morphemes (preverb, version vowel, stem formant) that a stem takes.

Transitivity is the only feature specified on the verb types, and it is closely related to the verb classes: as described in Chapter 3.1, there are four verb classes in Georgian. Classes 1 and 4 are generally transitive, and Classes 2 and 3 are largely intransitive. There is an additional division into relative and absolute verbs in Classes 2 and 3, as discussed in 5.4. Preverbs, pre-radical vowels, and post-stem formants constitute a class of lexically predefined morphemes: they carry grammatical, syntactic, and semantic functions, but their choice is normally determined by the verb itself. Preverbs originate from prepositions and still sometimes realize directional meaning; in terms of grammatical meaning, they are required to express certain tenses. This is also relevant for stem formants, the appearance of which also signals tense-aspect properties. The choice of the pre-radical vowel is initially motivated by the verb stem, however, the change of the pre-radical vowel can mark an appearance of an indirect object, a new tense form, a transformation from a transitive verb to intransitive, a reference to a surface that is involved in the action, and many other meanings.

The classification in the verb types according to the choice of the lexically predefined morphemes proved to be valid for a more linguist-friendly implementation process rather than necessary for the lexical rules.

5.2. Verb inflection

The section starts with an overview of the argument markers in the suffix and prefix positions through the revised verb slot structures. Assuming these structures, I propose an analysis based on the pronominal function of the markers, paying special attention to the slot division and tense-aspect characteristics. After this, proposals for indirect object and animacy hierarchy are discussed.

5.2.1. Revised verb slot structure

Tables 5.1 and 5.2 introduce two verb slot structures that modify the conventional structure (introduced in Chapter 3.3) on the basis of my observations from the test suite. The first columns in the tables are identical to each other and, in terms of argument markers, compatible with the initial verb structure. What motivated me to extend the initial verb slot structure is the insufficiency of the division of the affixes that are responsible for agreement into suffix and prefix pronoun markers only. Defining a slot as 'Pronoun 1' or 'Pronoun 2' does not fully satisfy the task of building an agreement system: they are underspecified in terms of the person, argument and its case features. Also, two slots are not capable of capturing the morphophonetic and syntactic interactions between the subject-object markers (most commonly, blocking) and determining their semantic roles.

The slots Thematic Suffix, Existential Marker and Screeve Marker are eliminated in the revised verb slot structure due to the lack of information that these definitions carry. Instead, slots for a passive marker, stem formants, and thematic suffix are suggested. The scope of each of them is discussed in the section on tense-aspect hierarchies (Section 5.2.2.2).

Table 5.1 divides the Pronoun 1 and Pronoun 2 slots by the argument role (second column) that a marker refers to: in the prefix position there are the indirect object, direct object, and subject; in the suffix position they are split into the direct object and subject slots. In the classification of the prefix slot by the argument role I am generally following the sets for subject, object and indirect object marking (Tables 3.4-6 in Section 3.2.1). The values from the sets are specified considering the data from the test suite and the function of the pre-radical vowel. When determining the suffix slots, I refer to the tense-specificity of the markers, as discussed below.

Table 5.2 extends the Pronoun 1 and Pronoun 2 slots with respect to the pronoun that the marker refers to regardless the argument role. Three person slots for the first, second and third values are present in the prefix position and the suffix position, and in the latter there is a forth slot for the markers of the first and second person values.

The majority of the agreement markers represent more than one grammatical function in terms of tense, person, number and case marking. The combinations of properties derived for each morpheme value are listed in Appendix A.

First, let's consider the slots in the prefix position. There are 12 values of the argument markers: *e, g, ge, gi, gv, gve, gvi, m, me, mi, u, v*; 6 of them appear as direct object markers: *g, gv, m, s, u, v*; all 12 can function as subject markers. In general, the prefix argument markers have case, person, and number features specified on the noun phrase.

Morpheme	Argument Role	Values
Preverb		*a, cha, da, ga, mi, mo, she*
Pronoun1	Indirect object marker1	*g, m, u*
	Direct object marker1	*g, gv, m, s, u, v*
	Subject marker1	*e, g, ge, gi, gv, gve, gvi, m, me, mi, u, v*
Pre-radical vowel		*a, e, i, u*
Stem		
Passive marker		*d*
Stem formant		*am, av, eb, i, ob*
Thematic suffix		*d, od, i, il, ul*
Pronoun2	Direct object marker2	*a, e, es, i, nen, s, var, xar*
	Subject marker2	*a, an, e, en, es, i, ian, iqavi, iqo, iqvnen, nen, o, on, os, s, var, xar*
Plurality marker		*t*

Table 5.1.: The revisted structure of Georgian verb: argument role

Morpheme	Person	Values
Preverb		*a, cha, da, ga, mi, mo, she*
Pronoun1	1	*gv, gve, gvi, m, v, me, mi*
	2	*g, ge, gi*
	3	*s, u, e*
Pre-radical vowel		*a, e, i, u*
Stem		
Passive marker		*d*
Stem formant		*am, av, eb, i, ob*
Thematic suffix		*d, od, i, il, ul*
Pronoun2	1	*var*
	2	*xar*
	1/2	*e, i, iqavi, o*
	3	*a, an, en, es, ian, iqo, iqvnen, nen, o, on, os, s*
Plurality marker		*t*

Table 5.2.: The revised structure of Georgian verb: person

The initial object and subject markers in the prefix position *m, g, gv, s,* and *v* do not express a tense role: they obtain a tense marking meaning once they are merged with the following pre-radical vowel. The opposition of the present perfect subject markers *mi, gvi, gi (-t), u (-t)* and the pluperfect subject markers *me, gve, ge (-t), u (-t)* shows that the change of the version vowel in combination with the tense-neutral object markers indicates a certain tense.

Next, in the suffix position, there are 17 marker values expressing 70 different grammatical roles, with 15 associated with object and 55 with subject argument roles. Grammatical role here refers to a combination of the tense, person, number and case values for one morpheme. An approach to the grammatical role as a set of features expressed in one value is grounded on the absence of exclusive markers for these features, as discussed on the example of tense-aspect characteristics below. The average rate of the grammatical roles per a morpheme value across subject and object markers is 4.1, for object markers only 2.5, and for subject markers 3. The ambiguous grammatical load for a single value in the suffix markers is due to the tense features that nearly all of the suffix markers express, however, different person values and case marking patterns add variability to these slots. As an example, consider a marker *o*: it refers to the first or second person subject in optative in transitive and intransitive constructions and to the third person singular subject in aorist of intransitive verbs.

5.2.2. Proposed Analysis: Pronominal affixation

Let's recall two properties of the verbal agreement that were introduced in Chapter 3: on one hand, there are markers that are morphologically present in the verb form, but due to the morphophonological and number constraints are blocked, in other words, represented by empty morphemes. On the other hand, there are reflexive *tav* constructions that state an external agreement pattern with a possible verb-internal alternative expressed by a pre-radical vowel. The observations on the distribution of grammatical roles per morpheme in the section above add another concern that should be addressed in my analysis - ambiguity in determinacy of subject or object roles in all argument markers and temporal features in the markers in the suffix position. A clear distinction of subject and object roles and the possible tense specificity are necessary for agreement modelling.

The first two agreement properties have been considered generally compatible with 'pronominal argument' analysis of the Georgian verb (Boeder, 2002). With this in mind, I suggest to extend this hypothesis to resolve the markers ambiguity, which would require treating the argument markers as pronominal affixes. Although the division between pronominal and agreement affixation is quite arguable, especially in polysynthetic languages, such as Georgian, and is considered fully language-specific (more on this in Evans and Sasse (2002), Baker (1996), Jelinek (1984)), Corbett (2006)derives several tentative instances of pronominal affixes in their opposition to agreement markers and free pronouns. Each of them is applied to Georgian and described below.

First, pronominal affixes are said to index the case roles for two arguments, unlike an agreement marker, which refers to only one. If we recall that Georgian case marking depends on the tense and the verb class, two case can be assigned only when three

parameters are known: a verb class accepting a marker, a tense of a verb form, and a case marking pattern required for the verb class in the particular tense. These parameters belong to the paradigmatic characteristics of the polypersonalism and are very frequently available for analysis. For example, knowing that *o* suffix subject marker appears in the optative Class 1 verb forms and that in this tense Class 1 verbs take A case marking (a subject is marked by Ergative, and object by Nominative), allows specifying two cases.

Next, pronominal affixation requires higher degree of referentiality and anaphoric properties that denote an affix's potential to appear independently as a reference to a free pronoun. Presumably these features are not available in Georgian. Descriptive content - the ability of an affix to express not only a grammatical or functional meaning, but lexical properties as well - might be limited the characteristic of Georgian as a pro-drop language and a consequent function of pronominal affixes. However, the argument affixes have a higher balance of information, which means that they expose more features than noun phrases that they refer to. First of al it refers to the case marking. The absence of declension in the first and second person pronoun, the limited case marking in the third person pronouns, and zero case marking in the nouns with some postpositions illustrate this imbalance: the argument affixes do store the case values for the arguments. The final criterion is multirepresentation, co-appearance of the elements with the same properties, and it is also constrained by argument dropping in Georgian, making multirepresentation possible, but not required, as in canonical agreement.

Summing up the pronominal properties of the argument affixes for Georgian, I suggest the pronominal argument affixes to include such features, as the person value and the cases values for two arguments if information on tense is available and therefore stored in the same affix. Having two case roles specified is a crucial technique for expressing agreement features in the blocked morphemes: in these cases, a feature is realized even in the absence of a morphosyntactic representation due to the broad controlling function of the marker. Another benefit of this approach is that the person reference is, in general, not ambiguous. There are four pronominal affixes marking both the first and the second persons: they store two feature values and are disambiguated by the presence of the first person prefix marker *v*. The cases, in which one marker denotes an argument of one person, but in the different class-tense combinations demands different case marking are extremely rare (consider two object markers *var* and *xar* in Table A.1). The support of the external agreement requires the properly working morphotactic module in the customization system, so that a slot can forbid or demand a *tav* pronoun. Unfortunately, this module was not available for the evaluation of this analysis, however, it is expected to work straightforwardly once the technical issues are solved.

5.2.2.1. Slot division

Slots are defined on the basis of similar input specification, and the division into morphemes within a slot is then motivated by applicable features and their values. The set of the possible features includes person, number, animacy, case information specified on the arguments, and tense values, specified on the verb. Let's consider an example of a suffix object marker *a* that expresses multiple grammatical functions, as shown in Table

value	arg	tense	person	numb	transitivity	case
a	obj	pres_perf	3	no	trans	nom
a	obj	fut	3	no	trans	dat
a	obj	imp	3	no	trans	nom
a	obj	pluperf	3	no	trans	nom
a	obj	aor	3	no	trans	nom

Table 5.3.: Object suffix marker *a*.

5.3.

Object marker *a* corresponds to the third person object with no number specified, therefore all slots with object marker *a* require an object plurality marker *t*. The object marker *a* appears in present perfect, future, imperfect, pluperfect and aorist tenses, and, in order to realize a particular tense meaning, each *a* marker requires a special preceding suffix. So, an object marker for present perfect appears only after a present perfect suffix *i*, an object marker for future co-occurs only with a post-stem formant (in my data - *eb*), an object marker for imperfect follows an imperfect-conditional-conjunctive marker *d*. Markers in pluperfect and aorist do not require any special suffixes in the verb form. These restrictions would give us already four slots for the object marker *a*. Creating these slots, however, is not the most favorable solution: the variability of the subject and object markers in terms of the marked grammatical functions with the similar phonetic representations is so high that the grammar should preferably stay as compact as possible. Thus the markers will be arranged with the corresponding object markers with the similar morphotactic requirements. So, an imperfect object marker *a* is joined with an imperfect object marker *i* for first and second persons: this slot then will have two morphemes with spellings *i* and *a*, and takes a slot with imperfect-conditional-conjunctive markers *d/od* as an input.

There are some specific steps in the slot and morpheme distribution undertaken in the implementation of the analysis. Third person singular and plural suffix markers cannot co-occur with the plurality marker *t*. To avoid their co-appearance, I specify the separate inputs for the respective object and subject slots, one one hand, and the plurality marker *t*, on the other hand. The mutually exclusive inputs help to handle the presence of mutually exclusive slots. Defining the inputs that outrule the presence of two slots is a very important technique in the absence of the 'forbid' constraint: it was applied to the morphophonological and syntactic rules (Section 3.2.1), meaning that, for instance, *v* marker will take completely different inputs than *g* marker, so that they will not appear in one verb form.

A slot with a plurality marker *t* consists of two morphemes: an empty one corresponds to singular, and a morpheme with *t* indicates plural. The first and second person in the argument suffix markers are disambiguated by means of the argument prefix marker *v*. Its person value is specified as first, and the first and second person markers bear two

values - first and second. Whenever it appears in a verb form and then is followed by the first and second person marker, the value 'first' will be chosen for the person of the argument.

5.2.2.2. Tense-aspect values

As a result of the test data analysis, no morpheme is viewed as a sole tense marker: tense is considered as a morphologically scattered, rather than concentrated, feature. The appropriate tense marking is achieved through the compatibility of the feature values in the lists of different slots.

Preverbs are said to appear in future, conditional, future conjunctive, aorist, optative, present perfect, pluperfect. Each preverb occupies a separate slot anf takes appropriate verb types. I list the tense-aspect values all in the feature values for tense, assuming that once it is combined in a form with another tense marker, the tense will be restricted to one meaning. In the following example, subject marker *s* marks present and future tense values. Thus, when a form *dac'ers* appear, present tense meaning will be ruled out since it is not compatible with *da* tense values.

(17) *dac'ers*
da-c'er-s
prv-write-subj3sg
'(He) writes (it).'

(18) *dac'era*
da-c'er-a
prv-write-subj3sg
'(He) wrote (it).'

Stem formants are lexically predefined suffixes that appear in some classes in present, future, conjunctive and perfect verb forms: the tenses are listed as in preverb slots.

Thematic suffixes slots comprise the morphemes with stable temporal-aspectual meaning regardless of the verb class, lexical properties and argument structure. The markers *d* and *od* have to be disambiguated: they mark imperfect, conditional, and conjunctive tenses from Series 1, or perfect tenses from Series 3. They can be split into two slots, so that certain suffix argument markers will take imperfect, conditional and conjunctive marker as their input, and the other will take the perfect one. Another perfect marker *i* has a limited number of the markers that can follow it. Participle markers *il* and *ul* do not explicitly mark a tense, but they are necessary for perfect intransitive Class 2 verbs.

Stem suppletion is an interesting change related to tense-aspect variations and occasionally transitivity. In the present perfect tense the absolute, or intransitive passives, and relative, or intransitive with a compulsory indirect object, verbs require different stems besides different subject/indirect object markers: absolute verbs take their regular stems (e.g. *mal*), and relative verbs take stems from the corresponding verbal nouns (e.g. *malv*). To implement this, two stems are defined as different lexical types, and an appropriate stem will be required by the matching marker in the verbal inflection.

5.3. Indirect objects and animacy hierarchy in the Georgian grammar

Two components of my analysis require an additional implementation step. One of them regards the language data: indirect objects are not supported in the customization system. Another one relates to my hypothesis that the position on the animacy hierarchy can effect subject-object transformations.

5.3.1. Indirect objects

I implement a type of ditransitive verbs that is of the special typological interest: Class 2 intransitive verbs with compulsory indirect objects.

Intransitive verbs with compulsory indirect object

There are two groups of Class 2 verbs: intransitive, or absolute, and intransitive that require an indirect object[1], called relative. Frequently relative and absolute verb forms differ in a change of one morpheme, e.g. a pre-radical vowel and/or a stem formant.

A solution for the relative verbs constructions is to treat them as transitive verbs in the customization system: the difference from the corresponding transitive verbs is specified in the lexical type's predicate and in the slot requirements. As an example, consider the verb 'hide':

(19)

bavshvi	*imaleba*
bavshv-i	i-mal-eb-a
child-nom	ver-hide-tf-subj3sg

'The child is hiding himself.'

The verb *imaleba* is a Class 2 absolute reflexive verb, the corresponding relative form requires a change of the pre-radical vowel:

(20)

bavshvi	*tavis*	*dzmas*	*emaleba*
bavshv-i	tav-is	dzma-s	e-mal-eb-a
child-nom	himself-dat	brother-dat	ver-hide.from-tf-subj3sg

'The child is hiding himslef from his brother.'

Absolute verbs in the perfective tenses keep the stem and require a participle suffix, whereas relative verbs change a stem for one from the verbal noun and require a perfective suffix:

(21)

dagmalvivar
da-g-malv-i-var
prv-io2-hide.from-ts.perf-subj1

'I hid from you.'

[1]It has to be noted that the scope of the indirect object semantics in Georgian is quite broad; for the sake of convenience the ambiguous prepositions translations are given with a slash sign.

(22) *movjrilvar*
mo-v-mal-il-var
prv-subj1-cut-ts.partic-subj1
'I have been cut.'

To handle these cases, first, I introduce three verb stems: *mal* with a predicate value '_hide_v_rel' for Class 2 absolute verb form, *mal* and *malv* with same predicate value '_hidefrom_v_rel' for Class 2 relative. The argument structure for *mal* absolute is described as intransitive and for *mal* and *malv* relative as transitive. Second, a slot with a pre-radical vowel *i* specifies an obligatory input of the verb type with *mal* absolute, and a slot with a pre-radical vowel *e* with specified indirect object marking as obligatory for *mal* relative. The appropriate verb formation in the perfective tenses is also supported by making obligatory inputs: a slot with *i* perfective suffix requires a verb type with a *malv* stem, and a participle suffix *il* requires a corresponding absolute verb stem. Lastly, indirect object markers used in relative forms have person and case features specified on the object NP.

5.3.2. Animacy hierarchy

An animacy hierarchy is frequently viewed as an agreement constraint in many languages (for more examples, see (Silverstein, 1976)). Tuite (1988) and Aronson (1994) argue for the development of an animacy hierarchy in the Georgian polypersonal agreement with a particular focus on the indirect object. They suggest a general tendency towards having an animate or personal indirect object.

The hierarchy suggested by Aronson regards the indirect objects that are marked on the verb. In Class 1 verbs an indirect object is lower in animacy than a subject and higher than a direct object, in Class 2 and in the majority of Class 4 verbs an indirect object is higher in the animacy hierarchy than the subject (in Nominative), and could be, as well as a direct object, the grammatical subject.

Relevance The animacy hierarchy is helpful due to two main reasons. First, the status on the animacy hierarchy can help to determine when a semantic subject in 3singular governs a 3plural marker on the verb. In the example below with a Class 2 relative verb example, 'eyes' is the subject, and 'old people' is the indirect object; the subject is inanimate and the indirect object is animate. What happens here is that the verb marks the subject as a singular, and the indirect object as plural.

(23)

gaunatdat	*tvalebi*	*moxucebs*
ga-u-nat-d-a-t	tval-eb-i	moxuc-eb-s
prv-ver-shine-sf-subj3sg-pl	eye-pl-nom	old.people-pl-dat

'The old people's eyes lit up. (The eyes lit up for the old people)'

In a construction with an equivalent Class 2 absolute verb a plural subject triggers a singular subject marker.

(24)

otaxebi	*ganatda*
otax-eb-i	ga-nat-d-a
room-pl-nom	prv-shine-sf-subj3sg

'The rooms lit up.'

The shift from a direct to indirect syntax is also valid for the verbs of other classes. The examples below are with a Class 1 verb:

(25)

amxanagebi	*axareben*	*ertmanets*
amxanag-eb-i	a-xareb-en	ertmanet-s
comrade-pl-nom	ver-make.happy-subj3pl	each.other-dat

'The friends are making each other happy.'

(26)

amxanagebs	*axarebt*	*ertmanetis*	*ambebi*
amxanag-eb-s	a-xareb-t	ertmanet-is	ambeb-i
comrade-pl-dat	ver-make.happy-obj3pl	each.other-gen	news-nom

'The friends are happy about each other's news.'

Second, the animacy hierarchy constraints help to predict the cases when an argument will not compete to mark on the verb. Aronson (1994) argues that Georgian verb tends to mark one animate argument on a form. Given that only two arguments, a subject and an object, can be marked on the verb, implementation of the animacy hierarchy will help to constrain the argument marking patterns.

Implementation The animacy hierarchy, roughly structured as 1, 2 pers > 3 anim > inanim, is implemented in two steps in the Direct-Inverse section of the customization system. First, I define the animacy on a scale for the verb classes following the patterns derived from (Aronson, 1994):

Class 1:
subj > indobj > dirobj
Class 2:
indobj > subj
Class 4:
dirobj > indobj > subj

Second, the scale structures are merged with an appropriate case marking for a class-series:

Class 1-1: nom > dat > dat
Class 1-2: erg > dat > nom
Class 1-3: dat > obl > nom
Class 2-1/2/3: dat > nom
Class 4-1/2/3: nom > obl > dat

5.4. Summary

This chapter has described my analysis of the argument marking on the verb in Georgian: it relies on the pronominal affixation that controls a person feature value for one argument and case feature values for two arguments: in order to support assignment of two case roles, a marker has to store the information on the verb class and the tense. Input specification is an important parameter for slot division, and the ability to specify feature values as lists allows to implement multiple tense values. Finally, I present the proposals and implementations for indirect objects and the animacy hierarchy. The next chapter provides the results of my implementation.

6. Evaluation

The evaluation of the proposed analysis has gone in line with the implementation: any significant change in the choices for Lexicon has to be tested by an immediate creation of a grammar and testing it in LKB. Taking multiple intermediate steps helps to detect validation and compatibility issues early, as discussed in Section 6.4. At least 77 intermediate Georgian grammars and 24 'dummy' training grammars had been created before the final one. The results for the final grammar, composed of 91 lexical rule, are overviewed in this chapter: besides providing the conventional metrics on the grammatical coverage (Section 6.1) and overgeneration (Section 6.2), the emergence of ambiguous readings is introduced on one example (Section 6.3).

6.1. Grammatical coverage

The coverage results are shown in Figure 6.1. The 'i-length in [0-5]' in the first column on the left indicates that the sentences in the test suite consist of less than five words. The second column refers to the total number of test items, or sentences (404), and the third column shows the number of grammatical, or positive sentences (322). The 'word string' column shows the average length of the sentence: it is 1.57 in the test suite. The relatively short average length is due to my goal: to handle and disambiguate the agreement marking within one verb form and, when possible, test marking in absence of one argument in transitive constructions. The 'lexical items' column shows the average of the lexical entries per sentence (7.71), and the comparison between this and the previous columns provides the information on the high lexical ambiguity. The next column 'distinct analyses' provides the mean number of parses per sentence (4.89) and helps

Aggregate	total items #	positive items #	word string Ø	lexical items Ø	distinct analyses Ø	total results #	overall coverage %
i-length in [0 .. 5]	404	322	1.57	7.71	4.89	321	99.7
Total	**404**	**322**	**1.57**	**7.71**	**4.89**	**321**	**99.7**

(generated by [incr tsdb()] at 2-aug-10 (07:51))

Figure 6.1.: Grammatical coverage of the Georgian grammar fragment.

Aggregate	total items #	negative items #	word string Ø	lexical items Ø	distinct analyses Ø	total results #	overall coverage %
i-length in [0 .. 5]	404	82	1.90	8.11	5.15	20	24.4
Total	**404**	**82**	**1.90**	**8.11**	**5.15**	**20**	**24.4**

(generated by [incr tsdb()] at 2-aug-10 (07:41))

Figure 6.2.: Overgeneration in the Georgian grammar fragment.

to conclude about the syntactic ambiguity of the implemented grammar. The last two columns refer to the number and percentage of the parsed sentences in the test suite.

The analysis has reached 99.7% grammatical coverage: all grammatical sentences except for one are parsed. The sentence *shinaarsebi daixatwon* failed to parse due to a bug in the customization system that aims to avoid a cycle between the slots defining *da* and *on* in the verb form. The lexical ambiguity is somewhat higher than expected, although presumably normal for such an ambiguity exposure as observed in the Georgian verb.

6.2. Overgeneration

The overgeneration metrics are provided in Figure 6.2: the grammar allows to parse 20 out of 82 ungrammatical sentences reaching 24.4% overgeneration.

6.3. Ambiguity

Despite reaching almost full grammatical coverage, the grammar created from the application of the pronominal agreement analysis shows a significant overgeneration rate, which is quite predictable given a large number of slots defined in morphology and the absence of compatibility constraints on them, as well as some engineering decisions (e.g. empty morphemes, free word order). As an example, consider a sentence *me miqvars c'igni*[1] 'I like the book'. Although the sentence allows only one reading with *me* being the subject in Dative and marked on the verb *miqvars* with a marker *m*, and *c'igni* being the object in Nominative marked by *s*, three parses are returned (Figure 6.3). The reason for this is that there is a lexical rule for object marker *m* that specifies Dative on the object, and the subject marker *s* that requires Nominative on the subject and Dative on the object. The application of these feature values becomes possible through inheritance relations and due to the properties of the pronoun *me* that does not decline. The chart in Figure 6.4 shows in details the relations between the lexical rules that contribute to

[1] LKB does not support the apostroph, so in the implementation the sign has been substituted for 'w' as the only letter not used in transliteration.

the ambiguous readings. Figure 6.5 illustrates the MRS representation of the sentence that provides all the necessary information that has been intended to make present, such as the tense, the animacy, the person and number markers, the predicate relations, and others.

Figure 6.3.: *me miqvars c'igni*: Ambiguity in the grammar.

6.4. Discussion

The suggested analysis has been successfully implemented through the customization system of the LinGO Grammar Matrix. The paradigmatic changes are expressed through

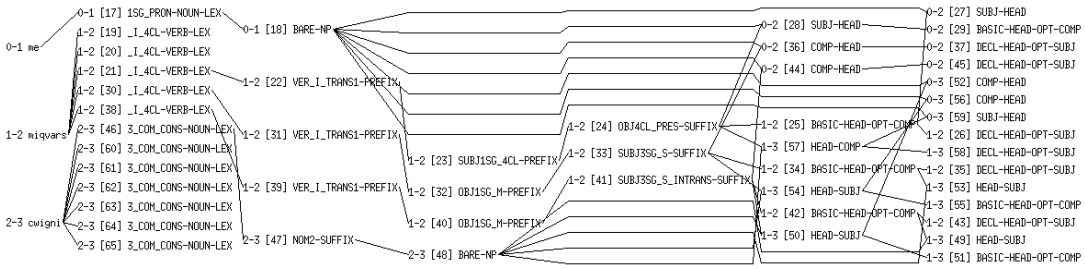

Figure 6.4.: *me miqvars c'igni*: lexical rules patterns.

```
[ LTOP: h1
  INDEX: e2 [ e SF: PROP-OR-QUES E.TENSE: PRES E.ASPECT: ASPECT E.MOOD: MOOD SORT: SEMSORT ]
  RELS: <
        [ "_pron_n_rel"
          LBL: h3
          ARG0: x4 [ x SORT: SEMSORT SPECI: BOOL COG-ST: COG-ST PNG.PER: 1ST PNG.NUM: SINGULAR PNG.ANIMACY: ANIM ] ]
        [ "exist_q_rel"
          LBL: h5
          ARG0: x4
          RSTR: h6
          BODY: h7 ]
        [ "_love_v_rel"
          LBL: h1
          ARG0: e2
          ARG1: x4
          ARG2: x8 [ x SORT: SEMSORT PNG.PER: 3RD PNG.NUM: NUMBER PNG.ANIMACY: INANIM COG-ST: COG-ST SPECI: BOOL ] ]
        [ "_book_n_rel"
          LBL: h9
          ARG0: x8 ]
        [ "exist_q_rel"
          LBL: h10
          ARG0: x8
          RSTR: h11
          BODY: h12 ] >
  HCONS: < h6 qeq h3 h11 qeq h9 > ]
```

Figure 6.5.: *me miqvars c'igni*: an MRS representation.

the tense-aspect markers on subject-object slots when the temporal and modal characteristics become explicit. Their relevance is demonstrated in the evaluation results. The animacy hierarchy has been also implemented through the Direct-Inverse library of the customization system. Nevertheless, its necessity could be negotiated: a grammar that is highly constrained in terms of semantics and that comprises a granular morphology proves to be sufficient for the agreement support. In addition, the Direct-Inverse hierarchies significantly increase the number of lexical rules leading to the slower grammars.

The evaluation showed that the implementation of the proposed analysis in the working version of the customization system has created a grammar with a high recall but lower precision. The application of morphotactics constraints will help to reduce lexical ambiguity, and evaluating on the basis of phenomena-specific test suites will provide more informative metrics.

The presented study is accomplished within the system that has been under permanent and intense development, and this has definitely influenced the implementation. My initial approach to the analysis of the Georgian polypersonal agreement in the customization system had to be denied due to the failure of the necessary components in the current version of the Matrix (Section 7.4). Two types of bugs have been detected and reported to the developers: validation bugs and LKB compatibility bugs. The validation bugs refer to the issues in validating the choices files and customization system

output. For example, at a certain step the customization system accepted user's inputs containing slash ('/') and conjunction ('&') signs, but would not create a grammar. There were also multiple cases when the system returned the message rejecting to create a grammar, and when a cause of a problem could not be detected by a user. In these cases, the choices that caused the system failure were forwarded to the developers. The LKB compatibility bugs refer to the bugs that appeared when successfully created grammars could not be loaded in LKB. A multiply occurring bug of this type is a 'recursion' bug: when running *script* file, the LKB browser reported that a certain rule fed itself, even though recursion could not be detected in the lexical rule's inputs. Another bug with a case name 'adv' confused with a type name for coordination was also detected only in the LKB running phase. Most of these bugs have been fixed, however, it seems that testing stage in the Matrix development can be improved in order to provide a more stable platform.

User's experience with the customization page interface is also an important component of the implementation. One interesting feature that has been experienced during the first steps of implementation is that the data-driven approach to the implementation is not completely evident. The first Georgian grammar included a significant number of rules that were discussed in the grammar sources but never appeared in the test data, therefore were completely redundant at that point. Though this could be a flaw of only current study, there is possibly a space for improvement in building a clear linkage between the phenomenon description and the data. The difficulty that a user might face in grammar engineering for hypothesis testing is quite obvious: a user first has to tailor the language knowledge to one particular aspect, and second, to describe the phenomenon in terms of its appearance in the test suite. In essence, it is a bidirectional issue since it involves creating a fully representative test suite.

The final point to be discussed is reproducibility of my analysis given the same platform and formalism. As it could be predicted from the complex morphology implementation, the deviations in the morpheme structure and definitions could be possibly of the lower rate than differences in the lexical types classification. However, rapid changes in the LinGO Grammar Matrix morphotactic system and the absence of any standard user guidelines might lead to a sharply different analysis from another user. For example, an addition of 'require' and 'forbid' constraints to morphemes within a slot will reduce significantly the final grammar.

7. Conclusion

In this section I present and discuss the results of the present research in terms of the evaluation of the analysis (Section 7.1) and the evaluation of the customization system (Section 7.2). In Section 7.3 I suggest the original design for the analysis of the polypersonal agreement that is left for a future investigation.

7.1. Conclusion on the analysis

The Georgian polypersonal agreement has been successfully implemented through the customization system; the developed grammar fragment reached an almost full grammatical coverage and an acceptable overgeneration rate. The implementation supported my hypothesis on the relevance of paradigmatic information, expressed through the temporal-aspectual-modal values in the argument markers in Georgian, especially in the suffix position. The specified temporal-aspectual features often eliminated the necessity for other rules (e.g. morphophonetic and syntactic).

My implementation approaches an agreement marker as a pronominal controller for one argument and a case controller for two arguments: case assignment is possible if a marker is constrained in its application to a specific verb class and screeve. The cumulative exponence (Matthews, 1993) of the grammatical properties proved to be valid for the Georgian polypersonal agreement.

An account for the animacy hierarchy as a direct-inverse scale and support for indirect objects have been developed. In particular, an implementation of indirect objects partially supports a claim that transitivity in Georgian comprises indirect objects as well as direct. Along with the observations from the data analysis on the appearance of indirect objects, this defines a possible next step of investigation of the Georgian verbal agreement. In a broad perspective, future work on the Georgian polypersonal agreement might include a number of phenomena that will test an advanced interaction between the matrix libraries. These phenomena regard certain dependencies between the number agreement and word order, information structure and person hierarchy, and discourse modifications and a change in pre-radical vowel.

Lastly, the implementation of the Georgian polypersonal agreement within a grammar fragment has started a new resource for a low-density language, contributing to the multilingual development of the natural language processing technology and offering a starting point for further development.

7.2. Conclusion on the customization system

The customization system is capable of supporting complex morphological structures. The support of the interaction between inflectional slots and types is limited to the input specification due to the technical failure of the morphotactic constraints in the current Matrix version (Section 7.4). A valuable point of the implemented morphology is its sustainability, based on a preference towards slot-to-slot rather than slot-to-type relations whenever possible. A presence of such morphology allows generation, however, the efficiency of a complex verbal inflection in generation should be investigated in future work.

In line with other studies of the customization system, my research shows that by using solely the customization system one may develop a highly competent and precise grammar. The possible improvements of the system might concern the following aspects:

morphotactics: Constraint specification for morphemes within a slot rather than for an entire slot can make the implementation more intuitive and can help to avoid slot doubling.

inheritance: The customization system is very flexible with input specification, which might lead to adding redundant types for the supertypes in the lexical rules. A system of the flags that indicate which inputs will be inherited from one chosen input would be helpful.

usability: Even though the customization system aims to be a platform for a user-linguist with no or very limited background in HPSG and MRS, many aspects of its performance are not straightforward (e.g. input specification, constraint application, typed feature structures). Development of a system documentation or a user manual will be an important step toward the use of the customization system for linguistic hypothesis testing, language documentation, and other purposes.

validation: Throughout this implementation a number of bugs have been identified, which suggests an improvement of validation sets.

developers-users: The customization system is an experimental platform, and it undergoes frequent and substantial changes. This sometimes results in problems with a grammar in validation and grammar creation. Supporting a working matrix version while another is under development, as well as other changes in communication between developers would be a great step forward.

7.3. Future work: Combinational analysis

My initial analysis for the Georgian polypersonal agreement relies on the morphotactic constraints in the customization system: by means of 'forbid' constraint it can be possible to realize, first, rule competition in the argument marking patterns (Section 3.2), and using 'require' constraint to support the combinations of argument markers. These

Verb class and series	Markers combination
Class 1, 2, 3; series 1/ 2	subject 1 + object 1
	object 1 + subject 2
	subject 1 + object 2
Class 1, 3; series 3	object_v 1 + subject 1 + object 2
Class 2 intrans, series 3	subject_v + subject 2
Class 2 trans, series 3	subject_v 1 + ind obj + subject 2
Class 4, series 1/2/3	object_v 1 + subject 1 + object 2

Table 7.1.: Subject-object markers combinations

combinations depend on the verb class and on the screeve. A detailed theoretical account of these combinations is presented in Appendix B, here I summarize those findings in Table 7.1: the morpheme and position notation is as introduced in Table 5.2 (1 stands for a prefix position, and 2 for a suffix one), except for omitting the term 'marker' and distinguishing the first person subject and object prefix marker (subject_v 1 and object_v 1 respectively). This morpheme usually serves disambiguation of a the subject or object suffix markers that express both the first and second person arguments.

A combination of subject and object markers is expressed by applying 'require' constraint, since it applies to a list of slots requiring at least one of them to appear. As an example, to support an argument markers combination in Class 4 verbs, an object marker *v* places two 'require' constraints - one on the appropriate subject slots and one on the possible object slots. This will guarantee that an object marker will co-occur with one subject marker and one object marker. The slots are divided on the same grounds as in the evaluated analysis adding the preferences for morphotactic constraints.

This account emphasizes the relational properties of the argument markers of simple rather than cumulative exponence. It is predicted to provide a higher precision grammar, however, due to a long-lasting failure of the morphotactic module of the customization system, it could not be evaluated: the current version of the Lexicon interface[1] allows a user to set the constraints, but are not cutomized in the grammar. The evaluation of this analysis will provide interesting results, especially in comparison with the implemented pronominal analysis.

[1]The version of Fri Jul 23 15:17:41 UTC 2010, as available on August, 19 2010.

Bibliography

Nino Amiridze. Georgian Reflexives in Subject Function in Special Contexts. In Stefan Mueller, editor, *Proceedings of HPSG05 Conference*, 2005.

Nino Amiridze. *Reflexivization Strategies in Georgian.* PhD thesis, Utrecht University, 2006.

Stephen Anderson. On the Notion of Subject in Ergative Languages. In *Subject and Object*, pages 1–24. New York: Academia Press, 1976.

Stephen Anderson. *A-Morphous Morphology.* Cambridge University Press, 1992.

Jemal Antidze and David Mishelashvili. Software Tools for Morphological and Syntactic Analysis of Natural Language Texts, 2006. URL `www.risc.uni-linz.ac.at/projects/intas/Timisoara/.../Antidze-paper.pdf`.

Howard I. Aronson. *Georgian: A Reading Grammar.* Slavica Publishers, 1990.

Howard I. Aronson. Datives and Indirect Objects in Georgian. In *NSL.7: Linguistic Studies in the Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*, pages 1–14. Chicago: Chicago Linguistic Society, 1994.

Mark Baker. *The Polysynthesis Parameter.* New York: Oxford University Press, 1996.

Emily Bender. Evaluating a crosslinguistic grammar resource: A case study of Wambaya. In *ACL'08: HLT*, 2008.

Emily Bender. Reweaving a Grammar for Wambaya: A Case Study in Grammar Engineering for Linguistic Hypothesis Testing. *Linguistic Issues in Language Technology*, 3(3):1–34, 2010.

Emily Bender, Dan Flickinger, and Stephen Oepen. The Grammar Matrix: An Open-Source Starter-Kit for the Rapid Development of Cross-Linguistically Consistent Broad-Coverage Precision Grammars. In *Procedings of the Workshop on Grammar Engineering and Evaluation at the 19th International Conference on Computational Linguistics.*, pages 8–14, 2002.

Emily Bender, Scott Drellishak, Antske Fokkens, Michael Wayne Goodman, Daniel P. Mills, Laurie Poulson, and Saffiyyah Saleem. Grammar Prototyping and Testing with the LinGO Grammar Matrix Customization System. In *Proceedings of the ACL 2010 System Demonstrations*, pages 1–6, 2010.

Barry J. Blake. *Relational Grammar*. London: Routledge, 1990.

Winfried Boeder. Syntax and morphology of polysynthesis in the Georgian verb. In Nicholas Evans and Hans-Jurgen Sasse, editors, *Problems of Polysynthesis*, pages 87–113. Academie Verlag, 2002.

Bob Carpenter. *The logic of Typed Feature Structures*. Cambridge University Press, 1992.

Marcello Cherchi. *Modern Georgian Morphosyntax: A grammatico-hierarchy-based analysis with special reference to 'indirect verbs' and 'passives of state'*. Wiesbaden: Harrassowitz Verlag, 1997.

Bernard Comrie. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge University Press, 1995.

Ann Copestake. *Implementing Typed Feature Structure Grammars*. CSLI Publications, Stanford, CA, 2002.

Ann Copestake, Dan Flickinger, Carl Pollard, and Ivan A. Sag. Minimal Recursion Semantics: An Introduction. *Research on Language and Computation*, 3(4):281–332, 2005.

Greville G. Corbett. *Agreement*. Cambridge University Press, 2006.

Scott Drellishak. Complex Case Phenomena in the Grammar Matrix. In *Proceedings of HPSG08 Conference*, 2008.

Scott Drellishak. *Widespread but not Universal: Improving the Typological Coverage of the Grammar Matrix*. PhD thesis, University of Washington, 2009.

Nicholas Evans and Hans-Jurgen Sasse. Introduction: problems of polysynthesis. In Nicholas Evans and Hans-Jurgen Sasse, editors, *Problems of Polysynthesis*, pages 1–13. Academie Verlag, 2002.

Antske Fokkens, Laurie Poulson, and Emily Bender. Inflectional Morphology in Turkish VP-coordination. In *HPSG'09*, 2009.

Adele Goldberg. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press, 1995.

Olga Gurevich. The status of the morpheme in Georgian Verbal Morphology. In *Proceedings of Berkeley Linguistic Society*, volume 29, pages 372–386, 2003.

Olga I Gurevich. *Constructional Morphology: The Georgian Version*. PhD thesis, University of California, Berkeley, 2006.

Moris Halle and Alec Marantz. Distributed Morphology and the Pieces of Inflection. In K. Hale and S.J. Keyser, editors, *The View from Building 20*. Cambridge: The MIT Press, 1994.

Alice Harris. *Georgian Syntax: A study in relational grammar.* Cambridge University Press, 1981.

Alice Harris. Georgian and the Unaccusative Hypothesis. *Language*, 58(2):290–306, 1982.

Brian G. Hewitt. *Georgian: A Structural Reference Grammar.* John Benjamins, 1995.

Dee Ann Holisky. *Aspect and Georgian Medial Verbs.* Delmar, N.Y: Caravan Books, 1981.

Eloise Jelinek. Empty categories, case, and configurationality. *Natural Language and Linguistic Theory*, 2:39–76, 1984.

Oleg Kapanadze. Applying Finite State Techniques and Ontological Semantics to Georgian. In Sergei Nirenburg, editor, *Language Engineering for Lesser-studied Languages*, pages 313–335, 2009.

Tracy Holloway King. SpecAgrP and Case: Evidence from Georgian. In *The Morphology-Syntax Connection. MIT Working Papers in Linguistics*, volume 22. 1994.

Hans-Ulrich Krieger and Ulrich Schäfer. TDL - A Type Description Language for HPSG. Technical report, Deutsches Forschungszentrum für Künstliche Intelligeny GmbH., 1994.

Rochelle Lieber. *Deconstructing Morphology: Word Formation in Syntactic Theory.* Chicago: University of Chicago Press, 1992.

P.H. Matthews. *Morphology.* Cambridge University Press, 1993.

Martha McGinnis. Projection and position: Evidence from Georgian. In *Proceedings of ConSole IV*, 1995.

Martha McGinnis. Two Kinds of Blocking. In *Morphosyntax in Generative Grammar: Procedings of the 1996 Seoul International Conference on Generative Grammar*, 1996.

Martha McGinnis. Case and Locality in L-Syntax: Evidence from Georgian. In *MIT Working Papers in Linguistics-32: The UPenn/MIT Roundtable on Argument Structure and Aspect*, 1998.

Damana Melikishvili. *System of Georgian verbs conjugation.* Tbilisi, 2001.

Paul Meurer. A Computational Grammar for Georgian. In *Logic, Language, and Computation, TBILLC 2007*, pages 1–16, 2009.

Stephan Oepen. [incr tsdb()] - Competence and Performance Laboratory. User Manual. Technical report, Computational Linguistics. Saarland University, 2001.

Kelly O'Hara. A Morphotactic Infrastructure for a Grammar Customization System. Master's thesis, University of Washington, 2008.

Carl Pollard and Ivan A. Sag. *Head-Driven Phrase Structure Grammar*. University of Chicago Press: Chicago, USA, 1994.

Safiyyah Saleem. Argument Optionality: A New Library for the Grammar Matrix Customization System. Master's thesis, University of Washington, 2010.

Akaki Shanidze. *Kartuli enis gramat'ik'is sapudzvlebi I. Morpologia. [Foundations of the grammar of the Georgian language I. Morphology].* Tbilisi: Tbilisi University gamomcebloba [publishing house]., 1953.

Michael Silverstein. Hierarchy of features and ergativity. In R. Dixon, editor, *Grammatical categories in Australian languages*, pages 112–171. Canberra: Australian Institute for Aboriginal Studies, 1976.

Stavros Skopeteas and Gisbert Fanselow. Focus in Georgian and the Expression of Contrast. *Lingua*, 2007.

Stavros Skopeteas, Caroline Fery, and Rusudan Asatiani. Word Order and Intonation in Georgian. *Lingua*, 119(1):102–127, 2009.

Susan Steele. Word order variation: a typological study. In Charles Ferguson Joseph H. Greenberg and Edith Moravcsik, editors, *Universals of Human Language IV: Syntax*, pages 585–623. Stanford: Stanford University Press, 1978.

Susan Steele. Many Plurals: Inflection, Informational Additivity, and Morphological Process. In *Many Morphologies*, pages 82–108. Cascadilla, 2002.

Gregory Stump. *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge University Press, 2001.

L. Thamarashvili. Morphological analysis of the mechanically translated Georgian language. *Edited trans. of Akademiya Nauk Gruzinskoi SSR, Tiflis. Institut Elektroniki, Avtomatiki i Telemekhaniki. Trudy*, 4(4):8–11, 1967.

Kevin Tuite. Case Attraction and Case Assignment. In *1st Eastern States Conference on Linguistics (ESCOL 84)*, pages 140–151, 1984.

Kevin Tuite. Indirect Transitives in Georgian. In *Proceedings of Berkeley Linguistic Society*, pages 296–309, 1987.

Kevin Tuite. *Kartvelian Morphosyntax. Number Agreement and Morphosyntactic Orientation in the South Caucasian Languages*. PhD thesis, Université de Montréal, 1988.

Kevin Tuite. Syntactic Subject in Georgian. In Howard I. Aronson, editor, *Non-Slavic Languages of the USSR*. Columbus, OH: Slavica Publishers, 1994.

Kevin Tuite. Paradigm Recruitment in Georgian. In Howard I. Aronson, editor, *NSL 8: Linguistic Studies In The Non-Slavic Languages of the Commonwealth of Independent States and the Baltic Republics*. Chicago Linguistics Society, 1996.

Kevin Tuite. Liminal Morphosyntax: Georgian Deponents and Their Kin. *Chicago Linguistic Society*, 39:774–788, 2007.

Kevin Tuite. Agentless transitive verbs in Georgian. In *Max Planck Institute for Evolutionary Anthropology, Leipzig, Department of Linguistics*, 2009.

Karina Vamling. *Complementation in Georgian*. PhD thesis, Lund University, 1989.

Hans Vogt. *Grammaire de la langue géorgienne*. Oslo: Instituttet for sammenlignende kulturforskning, 1971.

Warren Weaver. Translation. In Willian N. Locke and A. Donald Booth, editors, *Machine translation of languages: fourteen essays*, pages 15–23. Technology Press of the Massachusetts Institute of Technology, 1955.

A. Tables A.1-3: Subject and object markers in the Georgian verb

value	arg	tense	person	numb	transitivity	class	case
a	obj	pres_perf	3	no	trans		nom
a	obj	fut	3	no	trans		dat
a	obj	imp	3	no	trans		nom
a	obj	pluperf	3	no	trans		nom
a	obj	aor	3	no	trans	4	nom
e	obj	conj_fut	1, 2	no	trans		dat
e	obj	pluperf	1, 2	no	trans		nom
i	obj	imp	1, 2	no	trans		dat
i	obj	fut	1, 2	no	trans		dat
i	obj	pluperf	2	no	trans		nom
s	obj	pres_perf	3	sg	trans		nom
s	obj	pres	3	sg	trans		nom
s	obj	conj_fut	3	no	trans	4	nom
s	obj	conj_pres	3	no	trans	4	nom
s	obj	opt	3	no	trans	4	nom
var	obj	pres	1	no	trans	4	nom
var	obj	pres_perf	1	no	trans	1,3,4	nom
xar	obj	pres	2	no	trans	4	nom
xar	obj	pres_perf	2	no	trans	1,3,4	nom

Table A.1.: Object markers

value	arg	tense	person	numb	transitivity	class	case
a	subj	aor	3	sg	trans	1	erg
a	subj	imp	3	sg	trans	1	nom
a	subj	cond	3	sg	trans	1	nom
a	subj	pluperf	3	sg	intrans	2	nom
a	subj	pres_perf	3	sg	intrans	2	nom
a	subj	aor	3	sg	intrans	2	nom
a	subj	imp	3	sg	intrans	2, 3	nom
a	subj	cond	3	sg	intrans	2, 3	nom
a	subj	aor	3	sg	intrans	3	erg
an	subj	pres	3	pl	trans	1	nom
an	subj	pres	3	pl	intrans	3	nom
an	subj	pres_perf	3	pl	intrans	3	dat
an	subj	pres_perf	3	pl	intrans	2	nom
e	subj	aor	1, 2	no	trans	1	erg
e	subj	opt	1, 2	no	intrans	2	nom
e	subj	aor	1, 2	no	intrans	3	erg
e	subj	conj_pres	1, 2	no	intrans	3	nom
en	subj	pres	3	pl	trans	1	nom
en	subj	fut	3	pl	trans	1	nom
en	subj	pres	3	pl	intrans	3	nom
en	subj	fut	3	pl	intrans	3	nom
es	subj	aor	3	pl	trans	1	erg
es	subj	opt	3	sg	intrans	2	nom
es	subj	conj_pres	3	sg	intrans	3	nom
i	subj	cond	1, 2	no	trans	1	nom
i	subj	imp	1, 2	no	trans	1	nom
i	subj	imp	1, 2	no	intrans	3	nom
i	subj	cond	1, 2	no	intrans	3	nom

Table A.2.: Subject markers

value	arg	tense	person	numb	transitivity	vclass	case
i	subj	pluperf	1, 2	no	intrans	2	nom
i	subj	aor	1, 2	no	intrans	2	nom
ian	subj	fut	3	pl	intrans	2	nom
ian	subj	pres	3	pl	intrans	2	nom
iqavi	subj	pluperf	1, 2	no	intrans	2	nom
iqo	subj	pluperf	3	sg	intrans	2	nom
iqvnen	subj	pluperf	3	pl	intrans	2	nom
nen	subj	imp	3	pl	trans	1	nom
nen	subj	cond	3	pl	trans	1	nom
nen	subj	conj_fut	3	pl	trans	1	nom
nen	subj	conj_pres	3	pl	trans	1	nom
nen	subj	pluperf	3	pl	intrans	2	nom
nen	subj	aor	3	pl	intrans	2	nom
nen	subj	opt	3	pl	intrans	2	nom
o	subj	opt	1, 2	no	trans	1	erg
o	subj	opt	1, 2	no	intrans	2	erg
o	subj	aor	3	sg	intrans	2	erg
on	subj	opt	3	pl	trans	1	erg
on	subj	opt	3	pl	intrans	3	erg
os	subj	opt	3	sg	trans	1	erg
os	subj	opt	3	sg	intrans	3	erg
s	subj	pres	3	sg	trans	1	nom
s	subj	fut	3	sg	trans	1	nom
s	subj	pres	3	sg	intrans	3	nom
s	subj	fut	3	sg	intrans	3	nom
var	subj	pres_perf	1	no	intrans	2	nom
xar	subj	pres_perf	2	no	intrans	2	nom

Table A.3.: Subject markers (cont)

B. Subject-object markers combinations in verb classes and series

Class 1, 2 and 3 verbs in Series 1 (present-future-imperfect) and Series 2 (aorist-optative) allow the most diverse combinations of the subject and object markers. Apparently, the distribution between the subject and object markers across the suffix and prefix positions is due to the person feature value. First, after preverbs a subject marker is present. It is followed by the object markers in first or second person *m, gv, g (-t).* Subject and object markers in third person take the suffix position; the possible values here are: *s, a, o* for singular, and *en, nen, es, n* for plural.

The object-subject marking in the perfect series for Class 1, 2, and 3 differ. There is one paradigm for Classes 1 and 3, and two paradigms for intransitive and indirect transitive verbs of Class 2. Class 1 and 3 verbs in the perfect series allow one object prefix marker *v* that is followed by the subject markers. The structure of these subject markers represents an origin of the temporal and aspectual meaning in the prefixed markers. The base direct object markers (*m, g, gv*) are used to attach pre-radical vowels: *i* for present perfect, and *e* for pluperfect. One third person subject marker *u* is valid for both tenses. These combinations would create the subject markers with explicit tense specification: *mi*, *gvi, gi (-t), u (-t)* for the present perfect, and *me, gve, ge (-t), u (-t)* - for the pluperfect. These subject markers cannot be followed by additional pre-radical vowels. The object markers for Class 1 and 3 verbs in the perfect series are marked in the suffix positions: for the present perfect they are *var (-t), xar (-t),* and *s* or *a*; for the pluperfect the markers are *i, a/o (-t).*

Intransitive Class 2 verbs can have a prefixed subject marker *v* standing for a first person subject, and the other subject markers are moved to the suffix position. The subject markers in the present perfect are *var, xar, a, an*; in the pluperfect: *iqavi, iqo, iqvnen.* The transitive indirect verbs occupy the object prefix slot with indirect object markers (fully compatible with the direct object markers): *m, g, gv, h/s.* As with intransitive verbs, only the subject prefix marker *v* is allowed, and the perfect subject suffix markers are compatible with intransitive Class 2 verbs, whereas in pluperfect subject markers are *i, a,* and *nen.*

Two points should be noted regarding the subject or object suffix markers that are ambiguous between the first and second persons. First, they can be disambiguated by the subject or object first person prefix *v* as long as the latter is not blocked. Second, they have to be disambiguated from the preceding perfect marker *i.*

In Class 4, there are three stable subject markers: *m* for first singular, *gv* for first plural, and *g (-t)* for second singular (and plural); third person subject can be marked by a scope of the morphemes (*s, h, u*, etc). The subject markers can be preceded by

the object marker *v* only, which can be realized only with a third person object: *v* has to be deleted before the markers starting with *g*; as to *m*, a first person subject cannot appear with a first person object marker. In terms of the suffixed object markers, Class 4 verbs allow markers *var (-t)* for first person singular (plural), *xar(-t)* for second person singular (plural), *s* for third person in present, and *a* and *s* for imperfect/aorist and conjunctive/optative, respectively.