

INTERACTION BETWEEN WORD ORDER AND ARGUMENT MARKING IN SEVEN UNRELATED LANGUAGES

Toni Pedrós Caballero

MA thesis Linguistics

Universiteit Leiden




INTERACTION BETWEEN WORD ORDER AND ARGUMENT MARKING IN SEVEN UNRELATED LANGUAGES:

Tuvan, Lao, Figuig Berber, Itonama, Savosavo, Madurese and Santali

Thesis for the MA Linguistics, specialisation Language Diversity of Africa, Asia and Native America

Universiteit Leiden, academic year 2013-2014

Finished in May 2014

Toni Pedrós Caballero

Student number: 1421638

ACKNOWLEDGEMENTS

First of all, I want to thank my thesis supervisor, Mily Crevels, for her efforts during the thesis preparation period, during which she devoted a lot of time to the revision of all the chapters of the thesis and was willing to meet me as much as necessary. I have to thank her as well for her remarks when I was preparing the essay for the second part of the course on the languages of South America, devoted to the languages of the Andes, thanks to which I learnt to gloss properly.

This thesis is based on the knowledge acquired in the Master for which the thesis is the final step. I would like to add that I consider that the volume of knowledge acquired in the Master is immense. Therefore, I want to thank as well my other main teachers in the Master: Maarten Mous, who is the second reader of this thesis, Marian Klamer and Eithne Carlin. I also want to thank whoever has designed this Master, which fitted so well my field of interest when I was looking for a Master to choose around Europe. I thank as well the informant of the first semester course on linguistic fieldwork, Andi Ahmad Yanni, for his efforts to answer all the grammatical questions we were posing him, which are so difficult to answer for a native speaker of any language.

I also thank my classmates who have had most contact with me for their warm company: Barbara Westerveld, Meritxell Fernández, Dunja Wackers, Koert Goossens, Sujin Lee, Marc Sala, and I have to mention specially David Valls: our Saturday evenings out for dinner and drinks made my stay in Leiden during the first semester much more pleasant and less lonely.

Outside the academic world, I want to thank overall my lifemate, Manoli Fuentes; without her encouragement and support, maybe I would have never had the courage to embark on the adventure of leaving my ordinary life for some months to go to the Netherlands to do a Master. I also want to thank my mother for her moral and practical support before and during the Master period.

I also have to thank my boss, Imma Cerdà, for reporting favourably the leave applications that allowed me to get some work leaves for the time needed to do the Master.

Finally, I would like to thank the people of Leiden and surroundings in general for their friendliness, which prevented me from getting homesick despite being far from home.

ABSTRACT

The origin of the research idea of this thesis comes from a feature discovered through fieldwork on Bugis, a Western Malayo-Polynesian language. In this language, two different word orders are possible: in the basic order, different affixes occur on the verb and the arguments, while the non-basic order receives no marking. The goal of this thesis is to examine some unrelated languages in order to find out if, with different word orders, there is any difference in marking. Seven languages belonging to different families and areas have been chosen for this purpose: Tuvan (South Siberian Turkic), Lao (Tai, Tai-Kadai), Figuig Berber (Berber, Afroasiatic), Itonama (Amazonian isolate), Savosavo (Papuan), Madurese (Western Malayo-Polynesian, Austronesian) and Santali (Munda, Austroasiatic). The obtained results indicate that most languages show some different marking with different word orders, and a few languages that do not, have a rigid word order that does not allow changes. The changes in marking are not as obvious as in Bugis except in Madurese, which shows a high interaction between marking and word order. Moreover, the fact that clauses with the basic word order receive more marking occurs also in Figuig Berber. Further research with many more languages would be needed in order to discover if this feature may be common cross-linguistically, given that the low number of the studied languages does not allow cross-linguistic generalisations.

CONTENTS

LIST OF MAP AND TABLES

ABBREVIATIONS

1, 2, 3	1st, 2nd, 3rd person
A	Adjective
A	Aorist
A, A	Agent
ABL	Ablative
ACC	Accusative
ACT	Active voice
ADJ	Adjectiviser
AN	Animate
APPL	Applicative
ART	Article
ATT	Attributive marker
AV	Actor voice
B	Bare
C	Consonant
CAUS	Causative
CERT	Certainty marker
CL	Classifier
CNT	Continuative
CONJ	Conjunction
CONV	Converb
DA	Definite article
DAT	Dative
DEF	Definite
dem	Demonstrative
DERNOM	Derived nominal
DET	Determiner
DIST	Distal
DO	Direct object
DU	Dual
DV	Dummy vowel
EA	État d'annexion
EL	État libre
EMPH	Emphasis
EQU	Equative
excl	Exclusive
F	Feminine
FA	Familiar
FAC.NEWS	Factive, proposition is news
FIN	Finiteness marker
FOC	Focus marker
G, GEN	Genitive
IMP	Imperative
INAN	Inanimate
INC	Inchoative
incl	Inclusive
IND	Indicative
INST	Instrumental
INT	Interrogative
INV	Inverse
IO	Indirect object
IPFV	Imperfective
IRR	Irrealis
LOC	Locative
M	Masculine
MID	Middle voice
n.a.	Not applicable
N	Noun
NEG	Negation
NEUT	Neutral
NMLS	Nominaliser
NOM	Nominative
NP	Noun phrase
O, O	Object
OV	Object voice
P, P	Patient
PAST	Past
PE	Plural exclusive
PI	Plural inclusive
PF	Perfective
PL	Plural
Po	Postpositions
POSS	Possessive
PRF	Perfect
PRO	Pronoun
PROH	Prohibitive
PRS	Present
PRSINDF	Present indefinite

PRSPRT	Present participle
PSTINDF	Past indefinite
PSTPRT	Present participle
PTC	Participle
RECIP	Reciprocal
RED	Reduplication
REL	Relativiser
RELCL	Relative clitic
REP	Repetitive
R	Recipient
S, s	Subject
SG	Singular
STAT	Stative
SUB	Subordinate
T	Theme
TAM	Tense/aspect/mood
TPC	Topic
V	Verb
V	Vowel
VNT	Ventive

Map 1: World map with location of the studied languages

CHAPTER ONE
INTRODUCTION

1.1. ORIGIN OF THE RESEARCH IDEA

During the first semester of the academic year 2013-2014 (from the beginning of September until mid-December), I did a course on linguistic fieldwork at Leiden University together with five other students during the first part of the semester (until mid-October) and with four during the second part. The course consisted in fieldwork with a native speaker of Bugis, a Malayo-Polynesian language spoken in the south of Sulawesi Island (Indonesia). When we started eliciting the first basic clauses, we were all puzzled by a linguistic feature that was totally unfamiliar for us: transitive clauses could be built with two different word orders, AVP and VPA[1], the second one being clearly the most common. The VPA order had the feature of being adorned with some morphemes (an A agreement prefix and a P agreement suffix on the verb, and a kind of article on A and P), while with AVP order there was none of these morphemes. The same applied for intransitive sentences: both possible orders, SV ad VS, were possible, VS was clearly the most common and triggered morphemes on the V (a suffix agreeing with the S) and the S (the kind or article mentioned above), while none of these morphemes was present with the SV order. This pattern was something that was appearing very clearly and constantly, so that it was not the result of an elicitation, but something that came up quite obviously.

The final essay for the course was a grammar sketch of Bugis, in which we had to include a chapter on a special topic; mine was devoted to the above described feature, in Chapter 4, "The pattern of Bugis sentences: word order, verbal morphemes and the article *=e*", of my own grammar sketch[2]. I used 10 pages to explain this system, what means that it is quite complex if we take into account transitive and intransitive

[1] Throughout the whole thesis, I will use for the clause arguments the abbreviations S (subject), A (agent), P (patient), R (recipient) and T (theme). In languages where a subject category is clearly identified and the sources use the terms *subject* and *object* (most of those studied in the thesis), I use also these terms to refer to these grammatical categories. The use of S, A, P, R and T is crucial in this thesis in order to identify the same referents in all discussed languages. They are not intended to refer to grammatical categories, but to the semantic roles identified by Comrie (1989: 58-59). S is the only argument of an intransitive clause; A and P are the most agent-like and patient-like arguments of a transitive clause respectively, T and R are the patient-like and the beneficiary or goal of a ditransitive clause respectively.

[2] I have included my sketch in the references list and it is available online at: http://tinyurl.com/odqsyoe

sentences, and A and P noun phrases and pronouns. I will show here a brief token of this system. The two ways of building a sentence are shown in (1) (intransitives) and (2) (transitives)[3]:

(1) a. anana kərá
child cry
'A/the child is crying'. (Pedrós 2014, 36)

(1) b. [makːaˈlolo̩i ceˈbaːe]
makkalolo-i ceba=e
creep-3S monkey=ART
'A/the monkey is creeping.' (Pedrós 2014, 37)

(2) a. asu manre bale
dog eat fish
'A/the dog eats a/the fish.' (Pedrós 2014, 36)

(2) b. [ˈnunu̩i manuˈe asuˈe]
na-munu-i manu=e asu=e
3A-kill-3P chicken=ART dog=ART
'A/the dog kills a/the chicken.' (Pedrós 2014, 37)

In (1a) and (2a), the constituent order is SV/AVP, while in (1b) and (2b), the order is VS/VPA and morphemes are added to the verb and the enclitic article *=e* is attached to the nominal constituents, none of which occurs in (1a) and (2a). This shows clearly that a different word order implies different marking on verb and arguments.

The question arises if this feature (change of marking in relation to change of order) exists in more languages. This question constitutes the point of departure for this thesis: when languages have different constituent orders, is there any relation with argument marking or any kind of marking? In principle, I did not know of any other language with this feature nor was I familiar with any literature on the subject. If more languages were examined in search of such a relation, the result could be that it does not appear and that this feature of Bugis is quite rare. If this were the case, the absence of this relation in the examined languages would be the result of the research, but this absence would not be the only research goal, because the different ways or argument marking and different word orders in the selected languages would be examined and compared, and the result of this detailed examination and the comparison could also be interesting.

[3] Where no phonetic transcription is given, all characters are the same as used in the IPA system. All words are stressed on the penultimate syllable except where an acute accent is indicating the stress on the last syllable.

Yet I can advance that some relation has been found: in some of the investigated languages, there appear certain markers when the order of constituents changes, either on the arguments or in another way, but there are some different markers with different word orders, although not as obvious as in Bugis. This is even more interesting due to the fact that I give an insight into features to which the authors of the investigated grammars do not seem to have paid much attention, given that some of these features are merely mentioned en passant.

1.2. THE CHOICE OF THE SAMPLE OF LANGUAGES

The choice of languages has been done with the criteria of choosing languages that do not belong to the same family or to the same area. Regarding the areal bias, a language has been considered to belong to the same area as another one when there are typological traits extending to a whole area, or when languages in one area are known to have influenced one another, even though they lack clear common typological traits.

Firstly, I chose four languages that were representatives of the four traditional morphological types: Tuvan, a South Siberian Turkic language, as an agglutinative language; Lao, a Tai language spoken in Thailand and Laos, being the only official language in the latter country, as an isolating language; Figuig Berber, a Berber language spoken in eastern Morocco, as a fusional language, and Itonama, an isolate language spoken in the Bolivian Amazonian lowlands, as a polysynthetic language. I will not start to theorise here about the practicality of this traditional morphological classification, but will limit myself to say that, in my opinion, these languages are good representatives of these types, or at least show a high degree of features that can be rightly labelled with these types.

After this first choice, three more languages were added, in this case disregarding their morphological type, or I should rather say *not knowing* their morphological type, because they were totally unknown to me until I started studying their morphology. The choice of these three languages was done taking into account only the two first criteria mentioned above: avoiding the study of languages belonging to the same family or area. These languages are Savosavo, a Papuan language spoken in the Solomon Islands; Madurese, a Malayo-Polynesian language spoken in the island of Madura (Western Indonesia), and Santali, a Munda language spoken in Northern India. Madurese is a Western Malayo-Polynesian language, as Bugis also is (Adelaar 2005b: 10). The study of this language can show if the Bugis features can have any

relation with other languages of the same family. Although some wide-range typological studies put together Indochinese with Austronesian languages in the same area (e.g. Dryer 1992: 83), Indochinese languages are known to share many typological traits that justify the proposal of an Indochinese linguistic area (Enfield 2005) and that are absent in Austronesian languages. Regarding Savosavo, it could be argued that it is under influence of the neighbouring Oceanic languages, but, on the one hand, Papuan languages are treated independently in areal typological studies[4], although being surrounded by Malayo-Polynesian languages; on the other hand, Oceanic and Western Malayo-Polynesian are at opposite extremes of the Malayo-Polynesian internal subbranching. According to this, I consider that the choice of Lao, Madurese and Savosavo does comply with the intention of not choosing languages that belong to the same area.

In any case, no language has been chosen because some of the results were already expected, i.e. I had no idea if there was any relation between argument marking and word order in any of the chosen languages. This has been done on purpose, so that the choice of languages were a small sample of the languages of the world, in the sense that, if the above described feature of Bugis were actually a rareness of this language, probably no relation would be observed in a random sample. The only criteria for the choice of languages are those exposed above.

1.3. TYPOLOGICAL APPROACH

Since the publication of Greenberg's universals of grammar (1966), which propose some word order correlations, several studies have been published in which word order features of a high number of languages are compared. Some of them have proposed a new way of establishing correlations (e.g. Branching Direction Theory, Dryer 1992) and some have challenged the previous proposed correlations (Dunn et al. 2011). In these comparisons carried out with a high number of languages (e.g. 625 in Dryer 1992), the features to compare are assigned to each language and a possible correlation among the features is researched. In this way, each language receives a sort of *label* (e.g. SVO, DetN, GN, Po, etc.). This is something that can be best observed in

[4] Dryer 1992 (84-85) puts in the same area Papuan and Australian languages, but gives a long explanation in order to justify it and clarifies that, with this grouping, no claim is made regarding any influence between them. Dyer says that he puts them together rather for practical reasons.

the WALS (World Atlas of Linguistic Structures) online[5], where, among other things, many known features of a given language are shown. The problem with this typological approach is that, with every label, a lot of nuances and features that interact with this label are disregarded. Continuing with the example of the WALS and its list of features for Bugis, one of the features is word order, and Bugis is said to have AVP word order. Nevertheless, I have argued above and in my grammar sketch (Pedrós 2014: 36) that Bugis has a dominant VPA and a secondary AVP word order. WALS gives as a source Sirk 1983[6]. Sirk's grammar is based on ancient texts written with the Lontara alphabet[7], most of which were written in the 19th century and the oldest texts in the 17th century (Sirk 1979: 22), so the problem with this source is that it is representative neither of the modern nor of the spoken language. Besides this, Sirk (1979: 137) says: "Le premier type de proposition simple, le plus employé, est la construction avec prédicat préposé […]. Le prédicat se trouve donc ici en début de phrase […]. Le sujet suit le prédicat"[8]. After this statement, there are both verb-initial and subject-initial examples, but the explanation is quite clear: the predicate is at the beginning of the sentence. Therefore, not only the source is not very reliable as a token of the present spoken language, but it is also misinterpreted or only superficially consulted.

This shows that, when using many labels for many languages, there are chances that the research is based on a good number of mistakes. Another problem is that these labels can hide important facts. I will try to give a good example of this: when in a recent class at Leiden University, the percentage of languages that use only prefixes, only suffixes or both was presented, one of the students pointed out that, although Bantu languages use prefixes and suffixes, they use many prefixes and very few suffixes. Bantu languages had been included in the statistic presented in this class as being prefixing and suffixing, but their clear preference for prefixing remained hidden, although this is an important fact.

---

[5] Listed in the references under Dryer & Haspelmath 2013

[6] This is a translation in English from an original in Russian. My reference (Sirk 1979) is the translation in French of the same original.

[7] The Lontara alphabet was used for Bugis and Makassar, a neighbouring language of Bugis (Sirk 1979: 30).

[8] My translation: "The first type of simple clause, the most used, is the construction with a preposed predicate [...]. The predicate is placed thus at the beginning of the clause, [...]. The subject follows the predicate."

Work with feature labels on a high number of languages is therefore hiding lots of facts related to these features that can be crucial for achieving a better knowledge of how human cognition is reflected in language, if this has to be one of the goals of the research on universals. With this, I do not want to dismiss this typological approach with a high number of languages, given that it has rendered valuable results, but I would like to take a different approach in this thesis. This approach consists of examining in detail some features to which much attention has been paid in the literature: word order and argument marking; the interaction between word order and argument marking, however, has been little studied.

Obviously, when a feature is examined in depth instead of managing it as a label, a much smaller number of languages can be investigated —otherwise several volumes of an encyclopaedia would be needed for the purpose. This implies that typological generalisations cannot be obtained, because cross-linguistic features observed in a small number of languages cannot tell us something that occurs in most languages of the world. Yet the examination of an interaction to which little attention has been paid in a small number of languages can encourage future research if the obtained results are interesting. If, on the contrary, the results are of little interest, it could imply that there is not much point in further research in that direction. The possible interest of the results lies in the fact that, if some common patterns arise when unrelated languages are compared, they could also appear in more cases when more languages are examined. Therefore this could happen if further research on the subject is carried out.

Summing up, the idea of this thesis has originated in an *exotic* feature of Bugis that was discovered through fieldwork. This feature concerns key aspects of languages to which much attention has been paid in typological research: word order and argument marking. Nonetheless, little attention has been paid to the interaction of these two features, which is the goal of this thesis. Since the research of this feature in other languages cannot be done in the form of the above so called *labels*, a small number of languages is examined, which is intended to be a little *sample* of the world's languages. This constitutes an approach different from those mentioned above with a high number of languages. This thesis' approach has the advantage of examining the studied features in a more in-depth manner and the disadvantage of the impossibility of extracting typological generalisations from the results because of the small number of languages. Therefore, if there are interesting results, further research could be carried out in order to draw some typological generalisations. As stated above, certain relations between change of word order and marking (in some cases, argument

marking and, in other cases, another kind of marking) have been found in some languages. This is shown in detail in the respective section for each language.

1.4. LAYOUT OF THE THESIS

After this introductory chapter, chapter 2 is devoted to present general facts of all the studied languages (genetic affiliation, number of speakers, geographical scope, etc.).

Chapter 3 has a section for each language, in which the way of marking arguments, the word order and their possible interaction are examined in order to find if there is something happening in the rest of the clause when the arguments are put in an order different from the most common one.

The orthography and glosses of the examples are those used by the authors of each reference grammar, and at the beginning of the section of each language, the orthography used in the examples is explained shortly so as to make possible for the reader to know well all the phonological occurrences. When more than one grammar or shorter paper has been used and the orthography or the glosses are different, I have chosen those that I consider clearer for the reader.

After having examined the same features in all the studied languages, the results are compared in chapter 4, and, finally, in chapter 5, some conclusions will be drawn.

CHAPTER TWO
SOME FACTS ABOUT THE STUDIED LANGUAGES

In this chapter, some facts as number of speakers, geographical scope, genetic affiliation and other data of the studied languages are described.

2.1. TUVAN

Tuvan is a Siberian Turkic language spoken by "slightly more than 200,000 speakers" (Mawkanuli 2005: 1). Anderson & Harrison (1999: 1) say that the Tuvans "live in the Tyva Republic, a constituent of the Russian Federation located in a large basin along the upper Yenisei River and its tributaries just to the north of western Mongolia". They do not mention the Tuvans of China, the Jungar Tuvans, on whom Mawkanuli's compilation of texts (2005) is based. Mawkanuli says that Jungar Tuvan is a dialect of Tuvan "spoken by approximately 4,000 people living in the Jungar-Altay region of the Altay Prefecture, Ili Kazak Autonomous District, Xinjiang Uygur Autonomous Region of China" (2005: 1), around Lake Kanas, a place surrounded by the borders of Kazakhstan, Russia and Mongolia. This is indeed a place very close to the Russian Tyva Republic.

Within the Turkic family, Tuvan is classified by Johanson (1998: 82-83), based on a combination of genetic and geographical criteria, in a Siberian branch. He divides this branch in a northern and a southern subbranch, with Tuvan included in the latter. Within the southern subbranch, Johanson places Tuvan inside a Sayan group, consisting of Tuvan and Tofa. Menges (1995: 60-61) proposes a slightly different classification by putting Tuvan in a group D with two subgroups: Central-South-Siberian and East, with Tuvan included in the latter together with Tofa, which he calls *Karaγas*. The Central-South-Siberian subgroup is formed by Khakas (called *Xaqas* or *Abaqan* by Menges) and Šor. In this way, he does not include Altay dialects and Chulym in the same group as Tuvan, as Johanson does. Besides, Menges (1995: 61) says that the East Siberian group (Tuvan and Tofa) is a link between Central-South-Siberian (Khakas and Šor) and North-East-Siberian (Yakut and Dolgan).

2.2. LAO

As Enfield points out very well (2007: 7), "Lao is among a small minority of the world's languages to have achieved national language status", namely it is the official language

of Laos, where it is spoken by over 4 million people. It is also spoken by a minority in north-eastern Cambodia and a large minority in north-western Thailand (at least 10 million according to Enfield). Therefore, Lao is a majority language in Laos (4 million speakers out of 6.5 million Laotians), but a minority language in Thailand (10 million speakers out of 67 million Thais)[9]. In this way, Lao presents a sort of paradox in that most of its speakers are in a country where they are a minority, but around one quarter of its speakers give the name of its ethnic group to a country where they are the majority and their language is the only official one.

Lao is part of the Tai language family, which is included in a bigger family called Tai-Kadai (Smyth 2006). Enfield (2007: 17) includes Lao in a Southwestern Tai group, to which also Thai belongs. In this regard, Enfield (2007: 18) states that Lao and Thai "are for all intents and purposes (i.e., in structural linguistics terms) dialects of a single language". However, he says that they should be treated as different languages for a number of reasons, namely because "this favors the political objective of neither Thai nor Lao nationalism". This is thus a clear case in which political or ethnic reasons play a more important role than purely linguistic features in establishing language borders.

2.3. FIGUIG BERBER

Figuig Berber, as its name well describes, is a Berber language spoken in Figuig. Figuig is an oasis in Marocco at the fringe of the Sahara Desert and next to the Algerian border. The oasis is composed by seven villages with a total population of 14,280 inhabitants in 1982 (Kossmann 1997: 2). According to Kossmann, Figuig Berber is spoken in all the villages of the oasis, although in four villages dialectal Arabic is spoken by a part of the population as a mother tongue. Figuig Berber may be receding in these villages, but that Berber speaking people are proud of their language (Kossmann 1997: 3).

2.4. ITONAMA

Itonama, also called by its speakers *dihnipadara* (dih-ni-padara, 1PI-REL-speak), *dihpadara* (dih-padara, 1PI-speak) or *sihpadara* (sih-padara, 1PE-speak) (Crevels 2012: 213), is an isolate language spoken in the Amazonian lowlands of north-eastern Bolivia (Crevels 2011: 577, 2012: 213). Crevels (2011: 577) reports only one last speaker of the language. Therefore, this is a moribund language that will be extinct with the death of

[9] Data on inhabitants are taken from the *Encyclopædia Britannica* online 2014.

the last speaker, if he has not already died and it is already a totally extinct language. Crevels was documenting the language with a group of elder last speakers in the town of Magdalena during the last decade, but these elders had not acquired the language as its first language, which was Spanish: they had learnt it when they were child listening to their grandparents while hiding, because their parents did not want them to know the language; and practising it with other elders of the village (Crevels, personal communication).

Regarding its genetic affiliation, Greenberg classified Itonama as belonging to Paezan, a subbranch of a Macro-Chibchan phylum, but this classification remains today unsupported and "Itonama is still considered an isolate or unclassified language" (Crevels 2011: 577).

2.5. SAVOSAVO

Savosavo is a Papuan[10] language spoken on the tiny island of Savo (Solomon Islands) by around 2,500 people. According to Wegener (2012: 1), "in 1999 Savo Island was home to 2,549 people (Solomon Islands Census Office 1999), most of whom had Savosavo as their first language".

Savosavo is classified tentatively in a Central Solomons family by Ross (2005: 30) together with three more Papuan languages spoken ON the Solomon Islands (Bilua, Baniata and Lavukaleve). In his book chapter, Ross tries to classify the Papuan languages in families based on pronoun similarities. However, Dunn & Terrill (2012: 3)[11] show that lexical cognates are due in a great part to Oceanic borrowings in the four languages. Therefore, it is not widely accepted that these languages form a genetic unity in spite of their geographic distance from the rest of the Papuan languages. Savosavo shows the greatest linguistic similarity with Lavukaleve: 13.7%; but it shows also up to 20% shared vocabulary with neighbouring Oceanic languages (Tryon & Hackman 1983: 456, 460, 464; cited by Wegener 2012: 2-3).

[10] The term *Papuan* is used here as referring to non-Malayo-Polynesian languages spoken in New Guinea and some nearby islands, not to a genetic language family.

[11] Dunn & Terrill do not use the term *Baniata*; they call this language *Touo*. There is no doubt that it is the same language because there are only four Papuan languages in the Solomon Islands.

2.6. MADURESE

Madurese is the native language of Madura Island (Indonesia), next to the densely populated island of Java. Madurese is spoken in Madura but also in other parts of Indonesia due to the migration of Madurese people, overall in Java, Kalimantan and Sumatra (Davies 2010: 2). According to the 2000 Indonesian census, there are 6.8 million Madurese speakers, which makes them the third largest ethnic group of Indonesia after Javanese (83.8 million) and Sundanese (31 million) speakers (Davies 2010: 1).

Regarding its genetic affiliation, Davies refers to Adelaar (2005a), who classifies Madurese in a group called Malayo-Sumbawan, inside Western Malayo-Polynesian, inside the Malayo-Polynesian branch of the Austronesian family. Adelaar (2005a: 358) postulates a division of Malayo-Sumbawan in three splits, one of which is Madurese, the other two being Sundanese and Proto-Malayo-Chamic-Balinese-Sasak-Sumbawa. In this last split is included Malay, whose standardised versions are official languages of Indonesia, Malaysia, Brunei and Singapur.

2.7. SANTALI

Santali is a Munda language spoken in Eastern India. The people who speak it are the Sandals. Neukom (2001: 1) cites Ghosh 1994, who uses the figures of the 1991 census, and says that most of the Sandals live in the states of Bihar (2.12 million), West-Bengal (1.63 milion) and Orissa (0.53 million). “The main area centers around Puruliya (West-Bengal) and extends up to Bhagalpur in the north, Hazaribag in the west (both Bihar) and Keonjhar in the south (Orissa)” (Neukom 2001: 1). This makes a total of 4.28 million. The Ethnologue (Lewis et al. 2014) gives the more recent figure of 5.94 million in India (2001 census) and 6.23 in all countries (including some areas in Bangladesh and Nepal).

Santali belongs to one of the two branches of the Austroasiatic family, namely to the branch of the Munda languages spoken in India, the other branch being the Mon-Khmer languages of Indochina. Nonetheless, each of these two branches belongs to different linguistic areas: the Mon-Khmer languages are part of the Indochina linguistic area (Enfield 2005), while the Munda languages lie inside the Indian linguistic area (Emeneau 1956). This implies that Munda languages have many features in common with the neighbouring Indo-Aryan and Dravidian languages, such as retroflex consonants, conjunctive participles, numeral classifiers, absence of the verb

have and word order features SOV, AN, GN, demN, Po (Masica 1976: 187-190). Neukom (2001: 1) classifies Santali as being part of the Kherwari group of the North-Munda branch. He divides the Kherwari group into two subgroups: one is Santali and the other one is formed by Mundari, Korwa and Ho, which are thus the languages closest to Santali.

CHAPTER THREE
ARGUMENT MARKING, WORD ORDER AND THEIR INTERACTION

In this chapter, the chosen languages are examined in detail in order to describe how they mark the arguments, which their basic and other word orders are and if there is any interaction between these two features, i.e. if different word orders imply any change in marking.

3.1. TUVAN

3.1.1. ORTHOGRAPHY

In the examples, I have used Mawkanuli's (2005) orthography, even when using examples from Anderson & Harrison (1999). The only difference is mainly in the representation of [ɯ], for which Anderson & Harrison use *ɨ* and Mawkanuli uses *ı*. I have chosen to use *ı* because this is the character typically used in Turkic languages. Another difference is that Anderson & Harrison use a grave accent (`) to represent a low pitch that can occur only on a syllable-initial vowel (e.g. *èet* 'meat'). I use the grave accent in examples from Anderson & Harrison because Mawkanuli does not represent this low pitch in any way[12]. Mawkanuli does not give explanations for the characters he uses; he only says that the texts "were transcribed in a traditional phonemic transcription" (2005: 2). A character of his whose value is not clear is *ä*. Given the vowel harmony that can be observed in words where this character occurs, there is no doubt that it is a front vowel. Consequently, it must represent [æ] or [a], while *a* must represent [ɑ]. Anderson & Harrison (1999: 7) say that initial bilabial and alveolar plosives contrast for most speakers in aspiration rather than in voicing, so that, in word-initial position, the values of *p*, *b*, *t* and *d* are [pʰ], [p], [tʰ] and [t] respectively.

Other characters used in the examples that differ from those of the IPA are shown here:

ž ~ [ʒ] ı ~ [ɯ] š ~ [ʃ] y ~ [j] ö ~ [ø] č ~ [ʧ] ü ~ [y]

[12] We have to take into account that Anderson & Harrison's grammar is based on data gathered in the Russian Tyva Republic, while Mawkanuli's compilation of texts is based on Jungar Tuvan, the dialect spoken in China. Therefore, there can be some differences.

3.1.2. MAIN FEATURES

Turkic languages are often used as a prototypical example of agglutinative languages. Tuvan also shows a prototypical agglutinative morphology with suffixes being attached to the lexical stems. As is typical for an agglutinative language, morphemes have a low degree of fusion.

Anderson & Harrison (1999: 70-71) say that the basic constituent order is SOV (APV in my approach), but that "under certain instances of emphasis or focus, orders different from the basic SOV are encountered". Nonetheless, Anderson & Harrison do not extend more on the subject and do not say which other orders are possible. After examining some texts, a strict verb-final position seems to be evident, although an instance of change of order placing P before A has been detected (see (6)). This subject is treated below.

A short sentence showing the mentioned Tuvan features is given in (3) (Mawkanuli 2005: 210):

(3) men giži-niŋ ö-ö-ŋgö bar-ıp gil-di-m
I person-GEN house-3POSS-DAT go-CONV come-PAST-1SG
'I went to someone's house and came back'.

3.1.3. ARGUMENT MARKING

The Tuvan noun has a case inflection that is expressed through suffixes. Nominative has a zero marker, and the other cases are accusative, genitive, dative, locative and allative (Anderson & Harrison 1999: 14). Regarding the marking of A and P, nominative and accusative cases do not mark them straightforwardly: the accusative case marks the P, but only when it is definite or specific (Anderson & Harrison 1999: 17). Indefinite P bear the Ø mark of the nominative, which is used also for the A/S (Anderson & Harrison 1999: 15). The use of the same marking for the A/S and the indefinite P could lead to ambiguity, but Anderson & Harrison say that "indefinite direct objects frequently come in immediately preverbal position" (1999: 70). According to this, the immediate preverbal position of an indefinite P should identify it as P and avoid ambiguity, while the accusative mark identifies definite P's. This mark is -NI[13].

[13] Suffixes in Turkic languages are usually given with capital or small capital letters. This means that their phonetic realisation in a given word is subject to the rules of vowel harmony and consonant assimilation, which are triggered by the preceding syllable. In this case, -NI can be realised as [n], [d] or [t] plus a high vowel. The examples in this chapter will show the actual phonetic realisation of the suffixes.

The different case marking for definite and indefinite P's is shown in (4) (Anderson & Harrison 1999: 17):

(4a)	men	nom-Ø	nomču-du-m		(4b)	men	nom-nu	nomču-du-m
	I	book-NOM	read-PAST-1SG			I	book-ACC	read-PAST-1SG
	'I read a book'.					'I read the book'.		

Pronouns in the accusative are also formed by adding -NI to the nominative form. In this way, the corresponding forms of *men, sen, ol, bis, siler* and *olar* (1SG, 2SG, 3SG, 1PL, 2PL, 3PL) are in the accusative *meni, seni, onu, bisti, silerni* and *olarnı*. P pronouns are always marked because they cannot be indefinite.

3.1.4. WORD ORDER

As is said above, Tuvan clauses show a rigid verb-final position. This rigidness can be observed most clearly when several subordinate clauses are chained inside a long sentence in which the main verb occurs always at the end. These long sentences formed with several subordinates are translated in English trying to imitate the subordination chain, what, not being ungrammatical in English, is not common. A good example of a long complex sentence is (5) (Mawkanuli 2005: 28). The delimitation of subordinates with square brackets is mine, as well as the bold marking of verbs:

(5) ınžaŋgaš [bir mıŋ tos žüs üžünči žıl-dar-ı
therefore one thousand nine hundred thirtieth year-PL-3POSS

moŋguliya-nan arnawlı bilim-nig giži-ler **gıyde-p**] [[xom xanas
Mongolia-ABL special knowledge-ADJ person-PL **call-CONV** Kom Kanas

šigi dıva-lar-nıŋ göböy **ornalaš-kan**] žer-i-ŋge bičii
like Tuva-PL-GEN many **be.settled-PSTPRT** place-3POSS-DAT small

semiyä-de šagın mektep **až-ıp**] bir bölüm duva
family-LOC small school **open-CONV** one part Tuva

anıyag-dar-ın **ööröt-kön**.
young.people-PL-3POSS.ACC **teach-PSTINDF**.

'Therefore in the 1930s, they invited specially designated knowledgeable people (teachers) from Mongolia and opened small family schools in places like Kom and Kanas where Tuvas were concentrated, (and) had them teach one part of the Tuvan youngsters'.

The structure of the subordinate clauses is as follows (the verb of each subordinate is given inside the brackets): [*gıydep*] [[*ornalaškan*] *ažıp*] *öörötkön*.

The translation in English gives a coordination of three clauses, which correspond to the verbs *gıydep*, *ažıp* and *öörötkön*, with the clause headed by *ornalaškan*

being a relative clause referring to *žeriŋge* and within the subordinate headed by *ažıp*. The same chain of predicates can be coordinated in English and in many other languages, and this implies that all clauses are at the same level. However, in the Tuvan sentence, *gıydep* and *ažıp* are marked as converbs and only *öörötkön* has a tense suffix. This tense marking signals *öörötkön* as the main verb, while *gıydep* and *ažıp* are at a lower level. Besides the fact that converbs express subordination in Tuvan, it is important to note that *gıydep* and *ažıp* bear no tense or aspect marking, which is carried only by the main verb, *örötkön*. In this way, as happens with verb-final languages, the TAM (tense, aspect, mood) information is given only at the end of the sentence. A long complex sentence as (5) shows the strictness of the verb-final position in Tuvan because, independently of the length of the clause, TAM information is only given at the end of a series of pieces of information, which are chained around a main verb expressed at the end. In view of all this, the idea of TAM in Tuvan being expressed at the end of a *text* rather than at the end of a *sentence* could be an interesting discussion.

Regarding the order of A and P, I have been able to find one example in which the P occurs before the A. This is shown in (6), which is also a good example of a long chain of subordinates with the main verb at the end carrying the TAM information (Mawkanuli 2005: 31):

(6) baštawıš mektep-te duva ool-dar-ı mool dıl-ı-n
elementary school-LOC Tuva child-PL-3POSS Mongol language-3POSS-ACC

eki bil-be-en-dik-ten üžünči dörtünči kilas-ka
good know-NEG-PSTPRT-DERNOM-ABL third fourth grade-DAT

žed-ir **nom-nu** **mugalım** tuva dıl-ı-ŋga aŋnar-ıp
reach-PRSPRT **lesson-ACC** **teacher** Tuva language-3POSS-DAT translate-CONV

ool-dar-ga tüsün-dür-üp be-er.
child-PL-DAT understand-CAUS-CONV give-PRSINDF.

'In elementary school, because of the fact that the Tuvan children do not know the Mongolian language well, **the teachers** translate **the lessons** into the Tuvan language up to third and fourth grade to make the children understand'.

In the clause headed by the verb *aŋnarıp*, the A is *mugalım* and the P is *nomnu*. We can see that the P precedes the A. In this case, it seems that the subordinate clause headed by *žedir* is acting as a relative clause referring to *nomnu* ('the lessons reaching the third and/or fourth grade'), instead of the translation, in which this subordinate clause is treated as a prepositional phrase. The need to attach a relative clause to the P can be a good reason to place it before the A; otherwise the relative would be separating the A

from the rest of the clause. I have not been able to find another example of P before A in Mawkanuli's (2005) compilation of texts.

3.1.5. THE POST-VERBAL PRONOUN

As we have seen, the only possible change of constituent order in Tuvan consists in placing the P before the A, and the verb-final position is very strict. The only relation between order and marking that seems to occur is the nominative zero marking in indefinite P and their tendency to be placed in immediate preverbal position. The case system puts together S/A and indefinite P in the same category with the nominative zero marker, and in a different category with accusative marking, definite P. As regards the differentiation of S/A and indefinite P, this seems to be performed by the constituent order, i.e. with indefinite P in immediate preverbal position. Nevertheless, there is a post-verbal pronoun that breaks the verb-final strictness, which is explained below.

In spite of the strict verb-final position, there is a feature that seems to break this rigid rule, although this is not regarded in this way by Anderson & Harrison. Anderson & Harrison's grammar (1999: 39) say that Tuvan has two sets of verbal inflectional markers, that they call *Class I* and *Class II*. Class I markers are suffixes, while Class II is formed by enclitics. They say that Class I markers are used with subordinate clauses and class II with main clauses, except in PAST II (just PAST in Mawkanuli's terminology), which always uses Class I markers. First of all, the way Anderson & Harrison describe the use of both classes is somehow misleading, because, as we can see in (5) and (6), subordinate verbs take morphemes different from those of main verbs, i.e. they are inflected with participle or converb suffixes, while main verbs are inflected with TAM suffixes (tense in (5) and (6)). Therefore, it can be true that class I is used in subordinates and class II in main clauses except in PAST II, but this is due to the fact that verbs of these two kinds of clauses are inflected with different morphemes. I consider that a good explanation would be to state that participles and converbs, which are used in subordinate verbs, besides PAST II, take Class I markers; and the rest of verbal morphemes, used only in main verbs, take Class II markers.

As stated above, Anderson & Harrison say that Class II markers are formed with enclitics (1999: 39), but half a page down they say that "in the speech of certain Tyvans, Class-II markers are optionally enclitic, with a harmonic initial consonant". What they imply with this is that, when these markers are enclitics, the rules of consonant assimilation apply, whereas they do not apply when the markers are not enclitic.

Although not mentioned by Anderson & Harrison, we can assume that the markers are treated as independent particles when they are not treated as enclitics. This different treatment can also be seen in Anderson & Harrison's own examples, in which they use the enclitic gloss (=) with consonant assimilation, while, in other cases, they gloss class II markers as separate words with no consonant assimilation.

Be that as it may, the salient feature of Class II markers is that they have the same form as the full pronouns except in the 3rd person, which receives a zero marker. Table 1 shows the paradigm of Class II markers in comparison to full pronouns:

Table 1: Comparison of Class II markers and full pronouns in Tuvan (Anderson & Harrison 1999: 25, 39)

	CLASS II MARKERS		FULL PRONOUNS	
	SG	PL	SG	PL
1	men	bis	men	bis(ter)
2	sen	siler	sen	siler
3	Ø	(-LAr)[14]	ol	olar

If these markers are independent words —at least in most dialects, according to Anderson & Harrison (1999: 39)— and have the same form as full pronouns, they can be clearly considered as full pronouns. The only reason of not being considered as such by Anderson & Harrison may be that they are placed after the verb; however, this position, instead of making us consider them different from full pronouns, could lead us precisely to the conclusion that the constituent order PVA/VS occurs in clauses with these class II markers/pronouns.

Anderson & Harrison (1999: 39) say that "Class-II agreement markers constitute an archaic preservation of an earlier stage in the development of the Turkic languages". Menges (1995: 141-142), talking of these agreement markers in Turkic languages, says that "when enclitically attached, the personal pronouns are throughout identical with their original form. A certain development toward true suffixation has taken place inasfar as in a number of languages these enclitica have already undergone sound-harmony and/or assimilation and dissimilation of their initial consonants, while in the majority of the languages the accentuation still clearly indicates their enclitical nature". In Turkic languages, a straightforward way of identifying words is stress and vowel harmony, given that stress is usually placed on the final syllable and vowel harmony encompasses all the syllables of a word. Menges refers to *enclitics* when they are not

[14] L can be realised as [l], [n], [d] or [t] depending of the preceding phone. A represents a low vowel.

incorporated into the verb, but he says that, when undergoing vowel harmony, they become suffixes. On the other hand, as we have seen, Anderson & Harrison (1999: 39) say that "Class-II markers are optionally enclitic, with a harmonic initial consonant". According to this, undergoing vowel harmony implies, for Menges, that the pronouns become suffixes, and for Anderson & Harrison, that they become enclitics. Be that as it may, the real fact is that pronouns placed after the verb are in a process of grammaticalisation in Turkic languages, much more advanced in some languages than in others.

In Khakas, a Turkic South Siberian language like Tuvan, there are three classes of personal affixes (Anderson 1998: 25) —we have seen that there are two classes in Tuvan—, and their form resembles that of the personal pronouns, but in no case are they identical as in Tuvan 1st and 2nd person. The Khakas paradigm is shown in Table 2:

Table 2: Subject agreement suffixes paradigm in Khakas[15] (Anderson 1998: 25)

	CLASS I		**CLASS II**		**CLASS III**	
	SG	**PL**	**SG**	**PL**	**SG**	**PL**
1	-m	-BIS	-BIn	-BIS	-m	-BIS
2	-ZIŋ	-Z(Iŋ)ar	-ZIŋ	-Z(Iŋ)Ar	-ŋ	-ŋAr
3	Ø	-LAr	Ø	-LAr	Ø	(-LAr)

Looking at the paradigms of 1st and 2nd person verbal markers and pronouns in Tuvan (Table 1) and of Khakas suffixes, the logical inference is that both derive from full personal pronouns and have undergone a process of grammaticalisation: in Khakas, their form has changed and they have become suffixes, but in Tuvan, their form has remained unchanged and they have cliticised in some dialects and continue being separate words in most dialects. This is also stated by Schönig (1998: 412-413), who, referring to South Siberian Turkic languages, says that "a second set contains, in the first two persons, personal suffixes going back to postposed personal pronouns. The Sayan[16] Turkic suffixes are still identical with the pronouns".

In Mawkanuli's compilation of texts there are several examples of the use of these markers, sometimes glossed as a separate word (7) and sometimes as a suffix/clitic (8):

[15] The possible phonetic values of the archiphonemes shown in small capital are the following: Z can be [s] or [z], B can be [b] or [m], L can be [l] or [t], I represents a high vowel, and A represents a low vowel.

[16] Sayan Turkic languages are Tuvan and Tofa.

(7) al bis bo akkaba-da-gır olur-gan dıba žurd-u-n mončak
well we this Akkaba-LOC-RELCL live-PSTPRT Tuva people-3POSS-ACC Monchak

da de-er **bis**.
also say-PRSINDF **1PL**.

'Well, we also call the Tuva people who live in this Akkaba Monchak'.

(Mawkanuli 2005: 8)

(8) kogam-ga ün-ör-de iš ažılda-ar-da kazak-ša materiyal
society-DAT enter-PRSPRT-LOC work do-PRPST-LOC Kazak-EQU material

gel-er-de kazak-ša domaktan-ır-**bıs**.
come-PRSPRT-LOC Kazak-EQU speak-PRSINDF-**1PL**. (Mawkanuli 2005: 34)

'When we go into society or when we are at work, we speak Kazak when it comes to Kazak subject matter'.

In (7), *bis* is glossed as a separate word. It appears also at the beginning of the sentence, where it is glossed as 'we', while at the end it is glossed as '1PL', although the first *bis* could also have been glossed as '1PL'. In (8), *-bıs* is glossed as a suffix and undergoes vowel harmony. If we would take the existence of vowel harmony as a criterion to identify *bis* as a suffix/enclitic or a separate word, the fact is that, in (7), vowel harmony would result in the same form *bis*. However, in Anderson & Harrison (1999: 91) there is an example (9) with *men* after the verb[17] in which vowel harmony would result in *man* under the influence of the preceding syllable (*-ır*):

(9) àat sad-ıp al-ır **men**.
horse buy-CONV SBEN-PRSINDF **I**.
'I will buy a horse'.

In this example, *men* is glossed as 'I' instead of '1SG'. In any case, what it shows is that *men* is used here as a separate word because it does not undergo vowel harmony, which, as said above, is a good criterion to distinguish words in Tuvan and in Turkic languages in general. It is also remarkable that *men* is not uttered at the beginning of the clause, as is the case with *bis* in (7).

As we have seen, Tuvan has a very rigid verb-final word order, but it can be broken by 1st and 2nd person pronouns occurring after the verb. These personal pronouns can be cliticised, which is more common in some dialects, but they can also be independent words. In sentences with these post-verbal pronouns, the same pronoun can also occur at the beginning of the sentence, but this is optional. Since the obligatory occurrence of the pronoun is the post-verbal one, we must consider that

[17] *Al-* is an auxiliary verb with self-benefactive meaning (Anderson & Harrison 1999: 68).

these sentences show a PVA/VS constituent order. The post-verbal position of the pronoun is used only with what Anderson & Harrison call *Class II markers* (1999: 39), which occur with all TAM suffixes except PAST II markers (PAST in Mawkanuli 2005). This implies that they cannot occur in subordinate clauses, given that their verbs are inflected with participles and converbs. There is here a clear relation between a certain kind of markers (all TAM suffixes except PAST II/PAST) and a word order different from the basic one. This change is remarkable because it breaks a rigid rule of Turkic languages (verb-final position) and seems to be in a process of grammaticalisation. According to Anderson & Harrison, this feature is archaic in Tuvan (1999: 39). The comparison with the closely related language Khakas shows that these post-verbal pronouns are totally grammaticalised in Khakas and have become suffixes. In this respect, it is true that Tuvan is more archaic than Khakas because the process of grammaticalisation of its post-verbal pronouns is much slower, given that they occur both as enclitics and as separate words.

3.2. LAO

3.2.1. ORTHOGRAPHY

In the examples, I have used Enfield's orthography. Tones are represented with the following numbers:

1	mid level	3	low rising	5	low falling
2	high rising	4	high falling	ø	unstressed

Long vowels are represented doubling the vowel (e.g. *aa* ~ [aː]). The rest of the characters that appear in the examples and that differ from the IPA signs are shown below:

kh ~ [kʰ] ph ~ [pʰ] ù ~ [ɯ] ng ~ [ŋ] è ~ [ɛ] ò ~ [ɔ] ê ~ [e]

3.2.2. WORD ORDER AND CONTEXT AS ARGUMENT MARKERS

Lao is a good example of the isolating languages of South East Asia: every word is a monosyllabic morpheme. Examining Enfield's grammar, I have found some cases of polysyllabic words, usually compounds, but very scarce. If we compare it with Mandarin, which is often presented as a prototypical model of isolating language, we will see that Mandarin has much more polysyllabic words than Lao.

This implies that there are no morphemes to indicate the arguments in any word class, what should lead us to think that Lao, as English, relies on strict word order so as to mark the arguments. However, Enfield says that this idea is "weak at best, since extensive ellipsis and movement create widespread surface ambiguity, and without compromising communication" (2007: 272). Thus, Enfield describes Lao as a "pragmatically oriented grammar", and what he means with this is that the arguments are retrieved from context. Nonetheless, Lao has a basic word order, which is AVP or SV for transitive and intransitive sentences respectively. This is the order that a speaker would use when asked to utter a sentence, but this order is in no way a strict constraint as in English, as we will see below: what usually gives the information of who is *who* and *whom* in Lao is the context, while word order marks the arguments in case of ambiguity.

The combination of AVP and SV orders shows that Lao has a nominative-accusative alignment because both A and S are treated in the same way (Siewierska 1996: 153): both are placed before the verb and word order serves to distinguish the arguments. Regarding pronouns, they behave in the same way as noun phrases in that there are no different forms for A, P, T or R and their function in the sentence is governed by the same principles as the noun phrase. Their forms show distinctions in person, number and social level (Enfield 2007: 77).

The basic word order in Lao is shown in (10) for a simple intransitive sentence (SV) and in (11) for a transitive (AVP):

(Enfield 2007: 273)

(10) saan3 khon2 taaj3
three person die
'Three people died.'

(11) kuu3 jaan4 mùng2
1SG.B afraid 2SG.B
'I was afraid of you.'

3.2.3. CHANGES IN WORD ORDER

Enfield (2007: 273) says that changes in the basic word order are common and shows it with some examples that I reproduce here. In the intransitive (12), the normal order is reversed into VS; in (13), the order changes to VPA, and in (14), to PAV. The back slash represents "the onset of a prosodic mark-off, with lowered amplitude and pitch".

(12) taaj3 lèèw4 \ phòò1 hanø
die PRF father TPC.DIST
'(He)'d be dead, the father'. (Enfield 2007: 273)

(13) qaw3 mia2 \ haw2 niø
take wife 1.FA TPC
'Took a wife, I (did)'. (Enfield 2007: 273)

(14) lot1 haw2 laø bòø mii2
vehicle 1.FA PRF NEG have
'A car, I didn't have'. (Enfield 2007: 274)

Enfield (2007: 273) notes that "while it is formally apparent in these examples that something has been moved into an extraclausal position, there is no information about the semantic/functional role of the moved argument". In these cases, there is actually no possible ambiguity even without information on the role of the argument: a wife cannot take me (13) —in any case, a woman might take me as a husband— and a car cannot have me (14). But Enfield does provide us with some examples in which ambiguity can arise. A very good one is (15):

(15) (Enfield 2007: 276)
tamluat5 mak1 dêj2 \ phuø-saaw3 tòòn3 nan4
police like FAC.NEWS MC.HUM.girls time DEM.NONPROX
'Police$_i$, (they$_j$) liked (them$_i$) you know, girls$_j$ back then'. (Girls liked police).

In (15), the word order is PVA. Enfield says that he presented the audio recording of this utterance, out of context, to some Lao speakers, and they interpreted that "the police liked the girls" (2007: 277), while the meaning of the sentence in its context is just the opposite. This shows two things: 1) the basic order AVP is assumed by Lao speakers; 2) context is more important than word order to determine the role of each argument.

A further illustration of how important context is in Lao is shown by what Enfield calls *ellipsis*:

(Enfield 2007: 274)

(16) lùùm2
forget
'(I have) forgotten (it).'

(17) hên3
see
'(I) saw (it).'

(16) and (17) are sentences with only one verb. Both have an A and a P, but they are omitted and retrieved from the context. Therefore, context is not only fulfilling the role of an inexistent marker to identify A and P, but also replacing words that are not present in the linguistic utterance.

In a language like Lao, we are confronted with an extralinguistic element (context) that may substitute the morphological, syntactic and semantic features to

which linguists love to devote their analyses: A, P, S, word order and even whole words. Such a linguistic situation should not leave much room for an analysis of the expression of the arguments in connection with word order. Yet Enfield's examples show something interesting regarding what he calls *movement* (a certain argument moves to a clause position different from its basic position) in (12), (13), (14) and (15). In (12), (13), and (15), we have seen that Enfield marks with a back slash what he calls "the onset of a prosodic mark-off, with lowered amplitude and pitch" (2007: 273). If we consider in what these examples are different from (14) (movement without prosodic mark-off), it can be observed that (12), (13), and (15) have the prosodic mark-off before the S or A (after V in the intransitive (12) and after a grouping of V and P in (13) and (15)), while in (14) the word order is PAV, with the P separated from the V by the A. Therefore, there is a prosodic mark that occurs before a S/A, and after the V in intransitive sentences and after a cluster of V and P (VP or PV) in transitives. This position of the prosodic mark (\) can be represented in this way: VP/PV/V \ S/A.

Indeed, although void of any phoneme, a prosodic *marker* is present. Given that Lao is a tonal language, a more detailed study could shed light on the nature of this prosodic marker in combination with the tones that the adjoining syllables bear, since Enfield does not devote much attention to it and limits himself to signal its presence.

In any case, we can consider that this is a *marker*, yet prosodic, which is identifying a change in word order: namely from the basic AVP/SV to VPA/PVA/VS, while it is not present with PAV. If we regard VP/PV as a cluster of V and P that can be represented by V_P, we can reduce the statement by saying that the prosodic marker occurs in transitive sentences when the order is V_PA instead of the basic AV_P and in intransitives when the order is VS instead of the basic SV. Therefore, we can identify a prosodic marker that serves to signal a change of word order in Lao.

We have to take into account that, in Lao, a change in the tone of a syllable can change the meaning of a word, in the same way as a change of a phoneme can do in any language. Therefore, the importance of a prosodic marker should be regarded with a higher importance in a tonal language as Lao, given that, for a speaker of Lao, this marker can have the same importance as the insertion of a phoneme.

3.3. FIGUIG BERBER

3.3.1. ORTHOGRAPHY

In the examples, I have used the same orthography as in Kossmann 2014, even in examples taken from Kossmann 1997, because Kossmann 2014's orthography makes a higher use of IPA signs, and the reading becomes thus more transparent in a work as the present, which treats several languages.

Geminate consonants are represented by doubling the consonant (e.g. *kk* ~ [kː]). The rest of the characters used in the examples that differ from the IPA characters are shown as follows with their IPA correspondence:

y ~ [j] ṭ ~ [tˁ] ẓ ~ [zˁ] ž ~ [ʒ] š ~ [ʃ]

3.3.2. A FUSIONAL LANGUAGE

Figuig Berber is a language that shows features typical of fusional languages, in which, according to Comrie's definition (1989: 44), "the expression of different categories within the same word is fused together to give a single, unsegmentable morph". Some of these fusional features are case, gender and number (in nouns), and person, gender and number (in verbs) expressed with the same affix, or aspect expressed through different forms of the verb root without the possibility to dissociate aspect affixes from a lexical root.

3.3.3. ARGUMENT MARKING

In Figuig Berber, the S is identified in the same way as the A: through verbal agreement. Therefore, it is a language with a nominative-accusative alignment. Kossmann talks along his grammar (1997) of *subject* referring to S and A, which is logical given that there is a subject category in which S and A are included. As is said above, S/A is identified through agreement in person, gender and number in the verbal inflection. This inflection can be observed in (18), where verbal affixes agreeing with the S/A are in bold:

(18) (Kossmann 2014: 31)

i-kkər	**i**-nsu	din,	ns-**ən**		lməxzən
3SG.M-rise:PF	**3SG.M**-spend.night:PF	there,	spend.night:PF-**3PL.M**		authorities
ttʃəssa-**n**	xf-əs.				
guard:I-**3PL.M**	on-3SG.				

'So he spent the night there, and soldiers spent the night guarding him'.

In (18) there are four verbs, and all of them are inflected with a S/A agreement affix that indicates person, gender and number.

Pronouns are expressed through a set of clitics that act as P, and through another set for the R. This is shown in (19), with a P enclitic on *ssəhwa* and a R one on *nna*:

(19) ssəhwa-n=**t**=id lməxzən l lmalik
make.descend:PF-3PL.M=**3SG.M.DO**=VNT authorities to king
i-nna=yas
3SG.M-say:PF=**3SG.IO**
'The soldiers brought him down to the king and he said:' (Kossmann 2014: 31)

3.3.4. INTERACTION OF WORD ORDER AND ARGUMENT MARKING

3.3.4.1. Clauses with non-pronominal arguments

The basic constituent order in clauses with non-pronominal arguments in Figuig Berber is VAP in the transitive clause (20) and VS in the intransitive (21):

(20) y-awy asən[18] dd ṛəppi ta-yd-ətt
3SG.M-amener:A 3PL.M.IO VNT Seigneur EL.F-chien-F.SG
[19]'Le Seigneur leur apporta une chienne'. (Kossmann 1997: 300)

(21) i-mmut u-rgaz
3SG.M-mourir EA.M-homme
[20]'L'homme mourut'. (Kossmann 1997: 250)

However, this basic word order can change through the process that Kossmann calls *anticipation* (1997: 301). It implies that an element of the clause is placed before the verb. Both A/S and P can be *anticipated*. If the P is the anticipated element, it has to be repeated with a pronoun after the verb. Anticipation serves to emphasise the anticipated element.

Regarding the relation between change of constituent order and marking, we can find very clearly in this anticipation a different marking for A/S and P whether they are anticipated or not: this is carried out through case inflection. Figuig Berber nouns can be inflected with case. There are two cases, which Kossmann (1997: 88)

[18] R pronouns are treated as clitics in Kossmann 2014, but as separate words in Kossmann 1997.
[19] In examples from Kossmann 1997, I will show the original translation in French and will give my translation to English in the footnote; first the words in the second line of the gloss and then the sentence. In this case: *amener* 'bring', *Seigneur* 'Lord/Sir', *chienne* 'female dog'. 'The Lord/Sir brought them a female dog'.
[20] Translation: *mourir* 'die', *homme* 'man'. 'The man died'.

names in French *état libre* (EL) 'free state' and *état d'annexion* (EA) 'annexation state'[21]. Yet this case opposition does not occur in all nouns: only "dans les mots berbères ou berbérisés à préfixe"[22, 23]. The EA is used with A/S after the verb —i.e. their basic position—, after all prepositions except *bla* 'without' and *al* 'until', and after the demonstratives *u* (M.SG), *ut* (F.SG), *at* (M.PL), and *siwt* (F.PL); while the EL is used anywhere else, i.e. with non-A/S, with A/S before the verb, in non-verbal sentences and after the prepositions *bla* and *al* (Kossman 1997: 88-89; 2014: 7). Therefore, the A/S has a different case marking depending on its position regarding the V, with the exceptions described above. This different case is shown in the following examples: in (22a), *urgaz* 'man' is in EA because it is after the verb, while in (22b), *argaz* 'man' is in EL because it is before the verb:

(22) (Kossmann 2014: 7)

a. i-mmutər **urgaz** tamәṭṭut nn-әs i tẓәqqa
3SG.M-see:PF man:EA woman:EL of-3SG in room:EA
'The man saw his wife in the room'.

b. **argaz** i-mmutər tamәṭṭut nn-әs i tẓәqqa
man:EL 3SG.M-see:PF woman:EL of-3SG in room:EA
'The man saw his wife in the room'.

The P (direct object in Kossmann's terminology) is always in EL independently of its position with regard to the verb. Nonetheless, as is said above, when the P is before the verb, a full pronoun after the verb has to refer to the P:

(23) Lunža y-uwәy **tt** u ʕәmmi-s
Lunja 3SG.M-amener 3SG.F.DO celui.de oncle-3SG
[24]'Quant à Lunja, le fils de son oncle l'a amenée'. (Kossmann 1997: 302)

In (23), *tt* is the direct object pronoun, which refers to *Lunja*, the anticipated patient. Although the P noun phrase does not receive a different marking when it is anticipated, there is a new element, the pronoun, which appears in the clause. Given that both A and P are in EL when they are before the verb, it may seem that the pronoun serves to disambiguate. Nonetheless, this disambiguation function can only occur in cases in which A and P have the same gender and number. In (23), the verb is marked with a

[21] 1997's grammar is written in French, but 2014's paper is in English and Kossmann continues using the same French terminology, that is why I do not translate these terms here.
[22] My translation: 'in Berber nouns or nouns berberised with a prefix'.
[23] Figuig Berber has a very big amount of nouns of Arabic origin, which in some cases, as in this one, do not use the same morphology as Berber nouns.
[24] Translation: *amener* 'bring', *celui.de* 'that.of', *oncle* 'uncle'. 'Regarding Lunja, her uncle's son has brought her'.

3SG.M reference subject prefix, which means that the A is masculine and cannot be *Lunja*. Therefore, the pronominal element is here redundant, as in most clauses (all with A and P not having the same gender and number). This appearance of the pronoun can be regarded as a case of marking in relation to change of order, given that a new marker, the P pronoun, appears when the P noun phrase is placed before the verb.

This process described for the P (direct object) is used identically for the R (indirect object) (Kossmann 1997: 303).

3.3.4.2. Clauses with pronominal arguments

Besides this different marking of noun phrases, Figuig Berber shows also something similar in direct and indirect object pronouns. Direct object pronouns have three series, which Kossmann calls A, B and C (1997: 177). Series A and B are used when the pronoun follows the verb, while series C is used when the pronoun precedes the verb. The paradigm of the three series is shown in Table 3:

Table 3: Direct object pronouns paradigm in Figuig Berber (Kossmann 1997: 177)

	Series A	**Series B**	**Series C**
1SG	yyi	yyi	yyi
2SG.M	š ~ šekk	š ~ šekk	š
2SG.F	šem	šem	šem
3SG.M	ss ~ t	i	ss
3SG.F	tt ~ tet	it	stt ~ tt
1PL	anex ~ ayax	anex ~ ayax	danex ~ dayax
2PL.M	šnim	šnim	šnim
2PL.F	šnimti	šnimti	šnimti
3PL.M	ten	in	sten ~ ten
3PL.F	tent	int	stent ~ tent

The choice between the series A and B is based on the phonetic ending of the verb and is not going to be discussed here. Both series A and B are used after the verb, while C is used before the verb. Therefore, while the choice between the series A and B is mainly determined by phonetic reasons, the difference between the series A/B and C is a clear case of change in marking when a change in order is produced.

There is also a change in the indirect object (R) pronoun: a *d-* is added at the beginning of the pronoun[25] when it occurs before the verb. The indirect object pronoun in (24) would be *idd* if it were in its more usual position, i.e. after the verb.

[25] All indirect object pronouns start with a vowel.

(24) u **didd** i-nni walu
NEG 1SG.IO 3SG.M-dire.NP rien
[26]'Il ne m'a rien dit'.

3.3.4.3. Conclusions

We have seen that Figuig Berber has a basic VAP word order, but it can change by means of a process that Kossmann calls *anticipation*, which implies that an element of the clause is placed before the verb, mainly due to pragmatic reasons. This change of order is clearly marked in noun phrases and pronouns: in A noun phrases, through the use of the EL case; in P and R noun phrases, through a reference pronoun after the verb; in P pronouns, with a different paradigm, and in R pronouns, adding *d-* to the pronoun. We face therefore a language that marks overtly the change of order.

The subject reference affix on the verb shows clearly which gender, person and number the subject has. Therefore, case marking can serve to disambiguate only if object (P) and subject have the same gender, number and person. When a sentence as, for example, *The mother kisses the daughter* (same gender, number and person for A and P) is given in the basic VAP word order, who kisses whom can be determined through case marking (EA for A and EL for P) and word order. If the A were placed before the verb without case change, we could still identify if the element after the verb is A or P on the basis of its case. In this way, it can be said that the system is redundant because, if there were no case change, the case would show who is A and who is P. If there were no case change associated with word order change, EA could be renamed as *nominative* and EL as *accusative*, i.e. if the A were in EA independently of its position with regard to the verb —all this with the exceptions in which the EA is obligatory after certain prepositions or elements, as is explained above.

According to all this, we could reformulate the relation between case marking and word order as follows: there is an AVP word order, which does not receive differential marking; and there is a basic VAP word order, which receives differential marking. The grounds of this formulation are that, with AVP, both A and P are in EL, while with VAP order, the A receives a different marking (EA)[27]. Surprisingly, this resembles the Bugis system, in which the AVP word order receives no marking, while

[26] Translation: *dire* 'say', *rien* 'nothing'. 'He hasn't told me anything'.

[27] A PVA word order is also possible with different case for A and P, what means that this word order has differential marking and is non-basic, but I do not consider it here because my goal is to show the contrast between a basic order with differential marking and a non-basic order with no differential marking.

the VPA order receives differential marking and is the most common. In both cases, the existence of a verb between A and P can make it easier to identify them, while, if both are at the same side of the verb, there could be a higher risk of ambiguity, and this could be a functional reason for the differential marking. Be that as it may, Figuig Berber shows a marking related to word order that resembles the Bugis system in two features: 1) in the basic word order, the arguments receive a differential marking; and 2) A and P are after the verb in the basic word order and the verb separates them in the less basic order (AVP in both cases).

3.4. ITONAMA

3.4.1. ORTHOGRAPHY

I have used in the examples Crevels' orthography and glosses. Since Crevels 2012 is in Spanish and Crevels 2011 in English, when using examples of Crevels 2012, I have adapted the Spanish glosses to the English glosses in Crevels 2011.

Itonama bears the stress always on the penultimate syllable independently of the length of the word and the number of affixes. The signs used in the examples that differ from those of the IPA are as the following:

‘ ~ [ʔ][28] ï ~ [ɨ] ch ~ [ʧ] ty ~ [tʲ] y ~ [j]

3.4.2. ARGUMENT MARKING

Crevels defines Itonama as a polysynthetic language (2012: 246), “sumamente sintética, básicamente aglutinante y con una medida limitada de fusión”[29]. These features imply that the verbal complex encompasses much of the grammatical information of the clause. This is shown in (25), where the verb is the whole sentence: [30]

(25) k-a’-da-si-pa-k’ede-‘o
F-2SG-INT-INC-boca-mostrar-REP
‘¿Empezaste a avisarle de nuevo?’ (Crevels 2012: 246)

[28] When this sign is used after a plosive, it has the same value as in the IPA system, i.e. an ejective consonant.

[29] My translation from Spanish: ‘extremely synthetic, basically agglutinative and with a limited extent of fusion’.

[30] As far as the glosses in Spanish go, I will give the translation of the words in the second line of the gloss and then the free translation of the sentence. In this case: *boca* ‘mouth’, *mostrar* ‘show’. ‘Did you start warning him again?'

As is typical in a polysynthetic language, the arguments are indicated in the verb by affixes.

Itonama has two types of verbal clauses: independent and dependent (Crevels, 2011, 2012). The dependent clauses have a subordinating function. What is particular regarding these two types is that the arguments are marked in a different way in each type.

In independent clauses, 1st and 2nd persons are marked as a prefix when they have the role of S or A, and as a suffix when they are P (26), (27). However, the 3rd person is not marked in any case (28). The difference in dependent clauses consists only in the fact that the 3rd person is marked as S/A, but not as P (29).

(26) si-yalïs-na
1SG.S-be.hungry-NEUT
'I am hungry.' (Crevels 2011: 578)

(27) wase'wa de'-kewa-ne-mo
yesterday 2PL.A-face-see-NEUT-1O
'Yesterday you all saw me' (lit. 'you all saw my face'). (Crevels 2011: 579)

(28) u'-waba ohni siki'-na u-dame[31]
DV-chicha PRO:3SG traer-NEUT DV-maíz
'Para la chicha, él trajo maíz.' (Crevels 2012: 288)

(29) pa-nay-sewa-na-Ø
3SG-SUB-see-NEUT-3O
'When you see him/her/it.' (Crevels 2011: 583)

Regarding alignment, independent and dependent clauses behave in a different way. Both types show a nominative-accusative alignment because the S is marked in the same way as the A, but, in transitive independent clauses, when the A is 3rd person and the P is 1st or 2nd person, the P is marked with a prefix and the A is not marked, and the inverse marker *-k'i-* is added (30). The prefixes for 1st and 2nd person P have the same form as the S/A prefix in dependent clauses, with the exception of the 1st person singular, which has a specific prefix only used as P in independent clauses without the inverse prefix *-k'i-* (31).

(30) kumani a'-k'i-pachïhï'-ke kopone
last.night 2SG-INV-bother-PL rooster
'The rooster bothered you last night.' (Crevels 2011: 585)

[31] Translation: *chicha* 'meat', *traer* 'bring', *maíz* 'corn'. 'For the meat, he brought corn'.

(31) se'-buru'-tye ihwana[32]
1SG.INV-rascar-STAT Juan
'Me está rascando Juan.' Crevels 2012: 265)

This inverse system shows a person hierarchy in which 1st and 2nd persons are higher than 3rd. This hierarchy is present only in independent clauses and prevents a 3rd person morpheme (a zero prefix) from occurring before a 1st or 2nd person morpheme. The inverse system does not occur in dependent clauses, in which the P suffix is used independently of any hierarchy relation. The difference of the use of affixes can be seen clearly in (32), where the same arguments are used with a dependent and an independent clause.

(32) se'-ya-baya'-tye'-ka biluwa pa-nay-doh-na-mo
1SG.INV-mouth-lose-CNT-F.SG snake 3SG.F-SUB-bite-NEUT-1O
'The snake almost bit me' (lit. 'The snake mouth-lost me when she bit me').
(Crevels 2011: 588)

The whole paradigm of the different agreement affixes in Itonama is presented in Table **4**. The data are taken from Crevels (2011: 581-583; 2012: 262-264), but are put together in one table by myself. The addition of the inverse marker *-k'i-* is also mine.

Table 4: Paradigm of Itonama agreement affixes

Person	S/A in independent clauses	S/A in dependent clauses	P in inverse configurations	P in non-inverse configurations
1SG	si- / se'-	as-	se'-	-mo
2S	e'-	a'-	a'-k'i-	-we
2SG.F	ke'-	ka'-	ka'-k'i-	-we
3SG	Ø	ah-	n.a.	Ø
3SG.F	-ka	pV-	n.a.	Ø
1PE	se'-	sVh-	sVh-k'i-	-mo
1PI	de'-	dVh-	dVh-k'i-	-mo
2PL	de'-	dVh-	dVh-k'i-	-we
3PL	Ø	ah-	n.a.	Ø

3.4.3. WORD ORDER AND AFFIX ORDER INSIDE THE VERBAL COMPLEX

3.4.3.1. Word order

Regarding word order in clauses with noun phrases, Crevels (2012: 278) says that the basic word order in Itonama is VAP. This is illustrated in (33) (ibidem):

[32] Translation: *rascar* 'scratch'. 'Juan is scratching me'.

(33) i-doh-ne u-pa'u t'iyaya[33]
DV-morder-NEUT DV-perro niño
'El perro mordió al niño.'

3.4.3.2. Affix order inside the verbal complex and the person hierarchy

When one is examining a grammatical feature in languages typologically as different as those presented in this thesis, one faces the difficulty of comparing very different linguistic structures and sometimes the question even arises if there is actually much point in it. This idea comes out of comparing a language as Lao, in which every piece of grammatical information is a different word, with Itonama, in which most grammatical information is clustered in one word around one or even two lexical roots. Regarding the object goal of this thesis, this problem appears more clearly when the order of the arguments has to be examined and when one has to deal with pronominal reference. In Lao, word order is something that is evident in every clause, since the arguments, either noun phrases or pronouns, have a specific order. In Itonama, however, where the arguments are in most clauses attached to the verb and a clause is formed frequently by one single word, one has to dig in texts in order to find sentences in which talking about word order is something that makes any sense.

One could try to make an abstraction and try to avoid the difficult concept of *word*. Different definitions are given for what a word is, but the ways of identifying words in given languages is also quite different. In the case of Itonama, a word can be easily identified by the stress falling on the penultimate syllable (Crevels 2012: 244). Itonama is a moribund language, but, let us imagine that it had several dialects. Some of the dialects could change the stress pattern and this could lead not only to a phonological difference, but also to make it more difficult to identify word boundaries. This is obviously an abstraction, but I will give a real example of something happening in some southern Valencian varieties of Catalan. The written translation for English *I have talked* is in Catalan *he parlat* (he parl-at; have.1SG.PRS talk-PTC). The most usual pronunciation in Valencian dialects is [a paɾˈlat], with one stress in each word; but in some southern varieties, the pronunciation is [apaɾˈlat], with only one stress. We are speaking here of a language with a written tradition in which there is not much discussion on word boundaries, but a fieldworker who would be studying this language would say that, in the first case, the verb categories are expressed through an

[33] Translation: *morder* 'bite', *perro* 'dog', *niño* 'little boy'. 'The dog bit the little boy'.

analytic form, while, in the second case, there is only one word and the verb categories are expressed synthetically. I have used this example to show how fine and changeable the way of setting word boundaries in a language can be. In the same way, the fact that a pronominal morpheme is included in the phonetic scope of a stem or, on the contrary, receives independent stress, can cause that this morpheme is considered an affix, clitic or full pronoun, and therefore can be used or not for one or the other statistic. On the other hand, if a neighbouring dialect of the same language were considered, the data for that language could end up filling a different statistic. All that said and returning to the beginning of this paragraph, one could avoid altogether the concept of *word* and try to consider a clause as a chain of morphemes. In this way, the study of the order of constituents in Itonama would make more sense, because the order of pronominal morphemes included in the verbal complex could be fully compared with the order of Lao independent pronouns.

Nonetheless, the first thing to make clear is that the purpose of this thesis is actually not to make a comparison, but to look for changes in word order in relation to argument marking. In Itonama we have seen an inverse system that changes the order of agreement morphemes inside the verbal complex. This phenomenon cannot be studied under what is typically considered *word order*, yet I consider it important for the purpose of this thesis because it shows a change in order, not of words in the clause, but of morphemes inside the verbal complex. There is yet more reason to treat this phenomenon if we take into account that an Itonama verbal complex can form a whole clause.

In this regard, what we see in Itonama is a change of order —let us call it *pronominal reference morpheme order*— due to a person system in which 1st and 2nd persons are at a higher level than 3rd person. The verbal complex can be formed by many morphemes, but those with pronominal reference occupy a prominent place: in the first (S/A) or in the last position (P). When two pronominal references with different hierarchy levels occur, the person hierarchy in independent clauses does not allow that the lower reference occupies the first place and the higher reference occupies the last one. Given that the morpheme for the lower pronominal reference is marked with Ø (3rd person), we could explain the system in another way, i.e. stating that a pronominal morpheme cannot occur in the last position with the first position being empty. In any case, the very fact is that the P reference abandons its normal position (last) and comes to occupy the first position, usually reserved for the S/A, due to a person system, and, when this happens, the P morpheme adopts a different form.

The inverse system in Itonama shows therefore an example of change of order (inside the verbal complex) that needs special marking. As we have seen, Itonama has prefixes for A/S and suffixes for P, and the prefix forms are different from those of the suffixes. This means that, if the P suffixes were used as prefixes, there would be no ambiguity, because it would be clear that the reference is to the P. However, the change of order triggers special forms for P prefixes. Itonama shows thus a clear example of special marking in relation to change of order, in this case of morphemes inside the verbal complex instead of words inside the clause, but which is worth to consider for the reasons exposed above.

3.5. SAVOSAVO

3.5.1. ORTHOGRAPHY

The orthography used in the examples is the same as used by Wegener in her grammar of Savosavo (2012) —all the examples are from this grammar. The characters that differ from the IPA signs are these:

v ~ [ß] gn ~ [ɲ] gh~ [ɯ] ng~ [ŋ] gn ~ [ɲ]

The stress lies on the first syllable of disyllabic roots. In trisyllabic roots, the second syllable is stressed in around 75% of the cases, while in the remaining 25% of the cases, the stress falls on the first syllable. Four-syllable roots are stressed on the penultimate syllable. Words with more than four syllables are rare and their stress behaviour is complex (Wegener 2012: 24). When suffixes or enclitics are added, the stress can remain in its place or can shift to the right, and there is no rule that predicts the shift. In any case, the stress can never fall on the fourth or higher syllable from the end.

3.5.2. ARGUMENT MARKING

Wegener (2012: 6) defines Savosavo as “rather agglutinating”. Indeed, as we will see in the following lines, morphemes take the form of affixes and clitics.

Savosavo marks A and S in the same way, which means that this language has a nominative-accusative alignment (Wegener 2012: 198). Savosavo has a system of six marked cases (nominative, genitive, locative, ablative, comitative and benefactive) and one unmarked case (accusative) (Wegener 2012: 131-132). The fact that A/S (nominative) is marked and P (accusative) is unmarked contradicts Greenberg’s universal 38 (1966):

"Where there is a case system, the only case which ever has only zero allomorphs is the one which includes among its meanings that of the subject of the intransitive verb." It is indeed widely known that accusative and ergative are normally the marked cases in nominative-accusative and ergative-absolutive systems respectively. Therefore, this feature of Savosavo constitutes a rarity. Nonetheless, the P is obligatorily marked on the verb, while the A/S is not, as we will see below. This gives P a distinctive status, and it could be argued that this makes it the more marked argument.

Cases are expressed through suffixes and enclitics, which can be attached also to pronouns. This implies that pronouns are not inflected with case: they have different forms depending on number (singular, dual and plural) and gender (masculine and feminine) (Wegener 2012: 78). The A/S is marked with a nominative enclitic, whose forms are expressed through a paradigm based on person, number and gender. Nonetheless, this paradigm is reduced, as is shown in Table 5:

Table 5: Paradigm of Savosavo nominative markers (Wegener 2012: 135)

<table>
<tr><th></th><th>SG</th><th>DU</th><th>PL</th></tr>
<tr><td>1 + 2</td><td rowspan="2">na</td><td>na</td><td rowspan="3">na</td></tr>
<tr><td>3 masculine</td><td rowspan="2">tona</td></tr>
<tr><td>3 feminine</td><td>kona</td></tr>
</table>

The reduction of forms can be noticed in the facts that 1st and 2nd person have the same form, which is also the same for 3rd masculine in the singular paradigm; there is no difference in gender in dual, and there is only one form in plural for all persons and genders.

The use of the nominative enclitic is illustrated in the following examples. Square brackets to delimit nominal phrases are reproduced as in Wegener's examples. In (34), the nominative enclitic is used to denote the S of an intransitive sentence; in (35), the nominative is identifying the A of a transitive sentence; (36) and (37) are examples of the nominative enclitic in its 3rd person feminine and dual form respectively.

(34) [To mau]$_{NP}$=**na** ka ave-i.
3DU[GEN] father=**NOM** already die-FIN
'Their father had died already.' (Wegener 2012: 134)

(35) [No]$_{NP}$=**na** ka [korikori bulake]$_{NP}$ l-oi?
2SG=**NOM** already korikori bulake 3SG.M.O-eat.FIN?
'Have you eaten korikori bulake yet?' (Wegener 2012: 134)

(36) Zu [ko gnuba]$_{NP}$=**kona** oma ave-i.
but DET.SG.F child=**NOM.F** not die-FIN
'But the child (F) had not died.' (Wegener 2012: 135)

(37) To edo kola=zalo=**tona** lo polo=gha ze kata.
DET.DU two tree=DU=**NOM.DU** DET.PL pig=PL 3PL[GEN] bushwards
'The two trees (are) bushwards of the pigs.' (Wegener 2012: 135)

The S/A may also be expressed with a set of enclitics that are attached to the first constituent of a clause, i.e. they are Wackernagel enclitics[34]. Wegener (2012: 79) describes them as "nominative pronouns which can only be used for syntactic subjects". (38) is an example of the use of the first person singular nominative enclitic (in bold). The NP meaning 'only my war shield' is the P and the nominative enclitic is attached to this NP because it is the first constituent of the sentence:

(38) [Ai-va kise sua ghora pono]$_{NP}$=**gne** te mata-li(-i).
1SG.GEN-GEN.M fight ATT.SG.M shield only=**1SG.NOM** EMPH want-3SG.M.O(-FIN)
'I want only my war shield.' (Wegener 2012: 79)

In nominalised and some types of subordinate clauses, e.g. relative clauses, the S/A is marked with the genitive case, as in (39), where the A of the relative clause is *lo-va* 'he':

(39) lo-**va** k-au bo-tu ko adaki
3SG.M-**GEN.M** 3SG.F.O-take go-REL DET.SG.F woman
'the woman he took away' (Wegener 2012: 138)

The P is expressed through the accusative case with a zero mark (Wegener 2012: 136-137), and is also marked obligatorily on the verb through a prefix, a suffix, a prefix plus a suffix or stem modification. Affixes and stem modification agree with the object, which can be the P, the T or the R[35], in person, number and, in the 3rd person singular, gender. The choice of way of marking depends on the verb class, i.e. each verb adopts only one way of marking the object. The different paradigms are illustrated in Wegener 2012 (51-54). Examples (35) and (38) above show the use of the object agreement marker (glossed as O), in both cases 3SG.M, in (35) with a prefix and in (38) with a suffix. The 3SG.M marker is the most used since masculine is the default gender for all referents that do not have a discernible sex, i.e. inanimates and animates of unknown sex. An example of the use of the object marker in the 2nd person is provided in (40):

[34] Wackernagel enclitics are placed in the second position of the clause (after the first constituent in Savosavo), and they receive their name from Jacob Wackernagel, "who noted that enclitics in Indo-European languages usually occur in second position" (Hopper & Traugott 2003).
[35] Therefore, in the case of Savosavo, when referring to *object*, a verbal marker that agrees with P, T or R is meant.

(40) te=gho "zu agni=na ata; agni **n**-eghe-tu"
CONJ=3SG.F.NOM "but 1SG=NOM here; 1SG **2SG.O**-see-PRS.IPFV

tei(-i) ko=na.
say(-FIN) 3SG.F=NOM

'And she: "But I am here; I am watching you" —she said.' (Wegener 2012: 332)

As said above, the R and the T are marked in the same way as the P. The only verb that can take three arguments (A, T and R) is *l-ame-li*[36] 'give'. Although prefix and suffix constitute the same morpheme in verbs where the object is marked with a prefix and a suffix, *l-ame-li* is the only exception, because the prefix refers to the T and the suffix to the R (Wegener 2012: 166). Other actions that require three participants are expressed with a serial verb construction using the same verb *l-ame-li* (Wegener 2012: 193). The use of this verb is exemplified in (41), where the prefix *z-* agrees with the T ('two things') and the suffix *-mi* agrees with the R (the addressee):

(41) edo erongo=gha ai kati **z**-ame-**mi**-ghu=e
two something=PL 1SG.GEN CERT **3PL.O**-give-**2PL.O**-NMLS=EMPH
'…two things I will give you.'

3.5.3. WORD ORDER

Regarding the constituent order, in clauses with S or both A and P represented by a noun phrase, the order is APV in transitive and SV in intransitive clauses (Wegener 2012: 198). However, Wegener says that clauses with A and P as noun phrases are quite rare (2012: 203) due to the fact that "objects that are clearly identifiable from the context are usually dropped and thus only represented by the object agreement on the verb […] Subjects, on the other hand, are only very rarely dropped" (2012: 199). She also says that word order is rigid when A and P are represented by noun phrases, but it is more flexible when they are pronouns, and that "there is a tendency to have the verb complex at the end, and a restriction that object NP always have to precede the verb complex" (2012: 203).

In ditransitive sentences[37], the order of T and R in relation to each other is free, as shown in (42):

---

[36] In Wegener's grammar, Savosavo verbs out of context are shown with their object agreement affixes in 3SG.M. In this case, the root is *-ame-* and this verb takes a prefix and a suffix as object agreement markers.

[37] Although I do not treat ditransitive clauses in other languages, I do in Savosavo because T and R are marked with the same markers as P.

(42) (Wegener 2012: 201)

a. Mapamapa=ze lo pepa l-ame-li(-i).
RECIP=3PL.NOM DET.SG.M paper 3SG.M.O-give-3SG.M.O(-FIN)
'To each other they give the paper.'

b. Pa ringi pono=lo mane Basilio
one ring only=3SG.M.NOM consecutively Basilio

l-eme-li(-i).
3SG.M.O-give-3SG.M.O(-FIN)

'Only a ring he gave Basilio.'

In (42a), the R (*mapamapa*) occurs before the T (*pepa*), while in (42b), the T (*pa ringi*) is before the R (*Basilio*). Both R and T are marked with the same accusative zero marker and cross-referenced in the verb. Since in both examples in (42) T and R are 3SG.M, the marker on the verb does not serve to distinguish them, but the context allows no ambiguity due to the fact that the R is human and the T is inanimate in both cases.

According to all this, the word order system of Savosavo is a rigid APV/SV system with noun phrases, with clauses with A and P noun phrases being rare. In clauses without noun phrases, since objects are usually dropped and subjects are represented by Wackernagel enclitics, the order of constituents could be represented very schematically as shown in (43) for the transitive (a) and the intransitive (b) clause, where *x* is any first constituent of the clause and *p* represents a P affix, suffix, both or a stem modification:

(43) a. x=A Vp (43) b. x=S V

In spite of Wegener's statement "the constituent order in affirmative declarative verbal clauses is quite flexible" (2012: 203), my view is that Savosavo can be considered a language with a very rigid word order. In the rare cases in which all arguments are represented by noun phrases, Wegener admits that "the constituent order is more restricted" (2012: 203). It is to clauses with pronouns to which she attributes the word order flexibility due to the fixed position of the enclitic subject pronoun and the frequent dropping of object noun phrases. The position of the enclitic subject pronoun on the first constituent can cause a verb to be in initial position, as in (44), where it could be said that there is a VS order, but also that the subject enclitic is part of the verb complex and thus the sentence is formed only by a single verb.

(44) Samu-a=gho
have.meal-SIM=3SG.F.NOM
'[She will come later.] She is eating (at the moment).' (Wegener 2012: 201)

In fact, the section from where (44) is taken is called "Verb-initial minimal clauses" (section 7.1.1.1, 200-202), but all the examples provided are formed only by a single verb plus some elements of the "two layers of morphology" as defined by Wegener herself (2012: 161)[38].

The enclitic subject pronoun has thus a very rigid position: on the first constituent of the clause independently of its grammatical category. Such a *constituent* can be a verb (44), a conjunction (40) or a noun phrase with several words (38), i.e. the enclitic does attach to the first *constituent*, not to the first *word*. In this way, we can see that every constituent has its place very well established: noun phrases following the APV order, verbs at the end of the clause, subject enclitic attached to the first constituent, and P, when not a noun phrase, expressed through object affixes on the verb. The only permitted change seems to be between T and R. With such an order system, we can draw the conclusion that there is no change of constituent order in Savosavo because every constituent has its place very well fixed, with the only exception of T and R. But even T and R are marked with the same accusative zero marker and verbal object affix, what implies that there is no syntactic or morphological difference between them and thus we cannot speak of their word order in syntactic terms, but only in semantic terms. T and R together are marked only on the verb *I-ame-Ii*, where there is a morphological differentiation (prefix for T and suffix for R), but also in this case each marker has a fixed place.

According to all this, my conclusion is that Savosavo shows no change of constituent order because every constituent has a rigidly fixed place in the clause. Obviously, if there is no change of order, we cannot speak of argument marking in relation to change of order: there is no marking in relation to change of word order because there is no change of word order.

3.6. MADURESE

3.6.1. ORTHOGRAPHY

The orthography used in the examples below is as used in the grammar where they are taken from (Davies 2010). In my view, this is an orthography that does not do justice to the Madurese phonology.

[38] Wegener describes two layers of morphology for the Savosavo verb: an inner layer, closest to the stem, and an outer layer.

The orthography used by Davies 2010 is the standard Madurese orthography adopted in 1973 (Davies 2010: 53, 55). The unsuitability of this orthography for a descriptive grammar lies in the fact that some letters are used to represent different phonemes. Such an orthography may be good for native speakers, given that they know the phonology of each word, in the same way as European languages use the same letter for different phonemes, but it is not convenient for a descriptive grammar that has to be used by linguists who do not know the language. This orthography is hiding many phonemes under the same letter, which makes it useless for researches in which phonology plays am important role.

The letters that represent several phonemes are plosives and vowels. In the following, I show the phonetic values of the letters used in the examples that differ from the IPA symbols:

a ~ [a] or [ɤ]
e ~ [ə] or [ɛ]
o ~ [ɔ]
‘ ~ [ʔ]

b ~ [p] or [pʰ]
dh ~ [ɖ]
ng ~ [ŋ]

y ~ [j]
g ~ [g] or [kʰ]
j ~ [ɟ] or [cʰ]

As can be seen, some letters have different values, so that the reader cannot know which phoneme is actually used in the examples. There is even one character, *d*, that is used to represent three different phonemes (/d/, /tʰ/ and /ʈʰ/). It may be a good idea to choose an orthography that is already in use, but one should adapt it and make some changes in order to use it in a descriptive grammar written as a reference for linguists.

3.6.2. MARKING THROUGH VOICE IN TRANSITIVE CLAUSES

The Madurese way of marking arguments is carried out through the process that characterises the languages that Himmelmann (2005: 112) calls *symmetrical voice languages*. He includes in this group most languages of Western Austronesia (Taiwan, the Philippines, Malaysia, Madagascar and Western Indonesia) and says that “the defining characteristic of these languages is the presence of at least two voice alternations marked on the verb, neither of which is clearly the basic form”. This feature is also mentioned by Davies (2010: 249) in his grammar of Madurese: “Far and away the most striking and unique feature of Western Austronesian languages is the morphological means for identifying on the verb the most prominent argument in a clause.” The feature in question is best illustrated with a Madurese example:

(45) (Davies 2010: 249)

a. Ale' noro' ebu.
younger.sibling AV.follow mother
'Little brother followed mother.'

b. Ebu e-toro' ale'.
mother OV-follow younger sibling
'Little brother followed mother/Mother was followed by little brother.'

In both sentences in (45), *ale'* is the A and *ebu* is the P, and the actor voice[39] is indicating that the element before the verb is the A (45a), while the object voice indicates that the element before the verb is the P (45b).

According to Blust (2014), there are two schools who interpret this system in a different way: one says that this is a system of voice, which is inflected on the verb; the other one says that the clause has a focus and the marking on the verb indicates which argument role the focus is playing. Following the former approach, we should say that (45a) is an active and (45b) a passive sentence; when following the latter approach, we should say that the focus in both sentences is the first element of the clause and the verb indicates if the focus is A or P.

Davies (2010: 249) says that, in other languages, elements with a role similar to the first element of the clause in Madurese have been referred to as *subject*, *topic* and *trigger*. Davies chooses to call it *subject* arguing that this is the most neutral term. The reason for calling this *subject* argument in spite of it being A or P seems to be not only its pre-verbal position. Subjects (as defined by Davies) have a restriction in that they either must be definite or be introduced by the existential *badha* 'exist'. This restriction applies to A with the verb in active voice as well as for P with the verb in object voice (both called *subjects* by Davies). This is exemplified in (46) with a verb in active voice. Davies gives no example of the use of *badha* with object voice, but his explanation makes it clear that this restriction applies to all subjects, i.e. A with active voice and P with object voice.

(46) Badha burus ngekke' tokang pos.
exist dog AV.bite worker postal
'A dog bit the mail carrier.' (Davies 2010: 180)

Since the object of this thesis is to deal with the concepts of A and P in several languages, I will avoid a categorisation of this first element and will consider the elements in their role of A and P, independently of any theoretical approach to this

[39] In this example, the actor voice is marked with *n-*, which replaces the root consonant *t-*.

feature in this group of languages. According to this view, what we have in (45) is a sentence with AVP word order (45a) and a sentence with PVA order (45b) with the verb receiving a different marking in each case. With this simple statement, the research question of the thesis, i.e. if there is an interaction between argument marking and word order, should be easily solved in the case of Madurese, given that the verb receives a different marking with different constituent orders. Yet there are other ways of marking the arguments, and the intransitive clauses and the question of the basic word order must still be tackled.

3.6.3. WORD ORDER IN TRANSITIVE CLAUSES

Davies (2010: 149) says that the "canonical word order in Madurese is SVO", but he is considering as subject the A in clauses with the verb marked with actor voice and the P in clauses with the verb marked with object voice. Therefore, following the approach of this thesis, the word order must be posited as AVP and PVA in the two mentioned clause types respectively.

Regarding which order should be considered as basic, Davies (2010: 257) says that "in Madurese narratives, approximately 50-60% of all voice marking on transitive predicates is object voice". With this proportion, none of both clause types can be said to be basic, therefore my conclusion is that Madurese has no basic word order (parting from the point of view described above), or has two basic word orders: AVP and PVA. This leads to the question of what causes the choice of each clause type. Davies (2010: 258) does not give a clear answer to this question. He only says that "it has been proposed that in certain narrative discourse, actor voice and object voice clauses play different functional roles: actor voice providing background information and object voice for advancing the plot line". Anyway, as we have seen above, there is some controversy on how to explain this system in the group of languages that have it, and one of the explanations would be that the first element of the clause is placed before the verb because it is the focus.

3.6.4. INTRANSITIVE CLAUSES

Intransitive clauses are marked in two different ways depending on the control that the subject exercises on the predicate, i.e. verbs that imply a volitional control of the subject are marked differently from those in which there is no volition: verbs that imply volition are marked with the active voice, while predicates (verbal or adjectival) that do not imply volition receive no marking (Davies 2010: 157-158). This is exemplified below.

In (47) there is an adjectival predicate, and in (48) the predicate is a non-volitional verb. In both cases the verb bears no marking. In (49) the verb is volitional and is marked with the active voice prefix *a-*. In all cases, the word order is SV.

(47) Kopi reya manes.
coffee this sweet
'This coffee is sweet.' (Davies 2010: 157)

(48) Buku-na elang.
book-DEF disappear
'The book is lost.' (Davies 2010: 158)

(49) Na'-kana' rowa a-berka' ka toko.
RED-child that AV-run to store
'Those kids ran to the store.' (Davies 2010: 158)

Although Davies describes the verbs that bear no marking as non-volitional, he gives a short list of verbs of this class in which some of them may imply a volition of the subject (2010: 158): *ambu* 'stop', *dhapa'* 'arrive', *dhateng* 'come', *entar* 'go', *maso'* 'enter', *molae* 'begin', *mole* 'go home', *toju'* 'sit (down)'. Davies (2010: 159) also gives a list of verbs that bear active voice marking, which I show here for comparison: *abber* 'fly', *bala* 'say', *jalan* 'walk', *kejung* 'sing', *lonca'* 'jump', *maen* 'play', *rangka'* 'creep', *tanya* 'ask', *tangdang* 'dance'. The list of verbs without marking shows some verbs of movement that imply volition in most contexts (*stop, arrive, come, go, enter*), while some verbs that bear marking are also verbs of movement (*fly, walk, creep*), although they express specific kinds of movement. Therefore, we can observe that the semantic nuances that distinguish both classes may be difficult to discern.

Davies (2010: 159) says that the neutral word order is SV, but "the verb often precedes the subject for discourse-related reasons". He gives four examples of sentences with VS order, and in all of them the verbs bear no marking. Two of them have adjectival predicates and the other two have verbs of movement: one is *dhateng* 'come' and the other one is a two-verb construction, *lebat potpot*, translated as 'pass walk' in the gloss and as 'go' in the free translation. For the purpose of this thesis, it would be interesting to know if volitional verbs marked with active voice lose their marking when moved to initial position, but this is something that cannot be found out through Davies' examples because *lebat* and *potpot* may belong to the group of movement verbs that bear no marking. Since nothing is said in Davies' grammar in this respect, I will assume that intransitive verbs marked with active voice keep this marking when moved to initial position. The example with *lebat potpot* is shown in (50):

(50) **Lebat potpot** Tandha Serrat, (Davies 2010: 159)
pass walk Tandha Serrat

e baba-na gunong nojju dha' Klampes.
at under-DEF mountain AV.point to Klampes.

'Tandha Serrat went right around the mountain toward Klampes.'

3.6.5. MARKING THROUGH PREPOSITIONS

Besides the way or marking arguments described above, the use of prepositions can be combined with or replace voice marking. Davies (2010: 160) says that there is a group of predicates that "characterize the experience of the subject" that do not receive marking and whose P is marked with a preposition:

(51) Wati enga' dha' jawab-ba guru.
Wati remember to answer-DEF teacher
'Wati remembered the teacher's answer.' (Davies 2010: 160)

Davies (2010: 160) says that, in this kind of sentences, there is an experiencer argument and a stimulus argument. I will call them A and P respectively in order to avoid complicating things and to fit them in the terminology used in this thesis.

Sentences with this kind of verbs can also receive active voice with the addition of the suffix *-e*, whose status, according to Davies (2010: 259), is "somewhat uncertain". An example of this kind of sentences is shown in (52):

(52) Hasan ng-entar-e Bibbi'.
Hasan AV-go-E aunt
'Hasan went to (visited) Auntie.' (Davies 2010: 259)

In clauses marked with object voice, a preposition can be optionally added to the A. The options with and without preposition are shown in (53):

(53) (Davies 2010: 256)

a. Maleng rowa e-tangkep polisi.
thief that OV-catch police
'The police caught the thief.'

b. Maleng rowa e-tangkep so polisi.
thief that OV-catch by police
'The police caught the thief.'

Davies (2010: 256) says that "the preposition adds little to the structure" and does not talk of any nuance in meaning, but the preposition becomes obligatory "if the agent is separated from the verb". He gives two examples of this: in (54), an adverb (*bari'*

'yesterday') separates the verb from the A, and in (55), the A has moved to initial position, thus having as a result an APV word order:

(54) Ale' e-kekke' bari', bi' embi' rowa.
younger.sibling OV-bite yesterday by goat that
'That goat bit Little Brother yesterday.' (Davies 2010: 257)

(55) So polisi maleng rowa e-tangkep.
by police thief that OV-catch
'The police caught the thief.' (Davies 2010: 257)

Therefore, (55) shows us another possible word order: APV, which requires a preposition before the A and object voice marking on the verb.

3.6.6. ALIGNMENT

Regarding alignment, we have seen that the common way of marking S, A and P is by means of a voice marker on the verb, but A and P can also be marked with a preposition. I will consider firstly the marking with voice.

We have seen that some S have their verb marked with active voice, the same as A, but other S (those with non-volitional verbs or verbs not controlled by the subject) have an unmarked verb. This system shows us three ways of marking arguments:

1) S of volitional verbs and A are marked with active voice on the verb;
2) P are marked with object voice on the verb[40], and
3) S of non-volitional verbs have their verb unmarked.

This can be considered roughly an active alignment system as defined by Siewierska (1996: 153) and Dixon (1987: 4; 1994: 70-71). According to Siewierska, the active alignment occurs when the S is sometimes treated as the A and sometimes as the P. Dixon defines active or split-S marking in the same way (1987: 4), but adds that "those S which are semantically similar to A (exerting control over the activity) will be S_a, marked like A, and those S which are semantically similar to O (being affected by the activity) will be S_o, marked like O". The Madurese system fits this description but differs in that the S is never marked as the P. In cases where active alignment/split-S languages mark the S like the P (with non-volitional or non-controlled verbs), Madurese treats the S with a special marking for this kind of clauses: with no marking on the verb. This kind of clauses represent what Siewierska (1996: 153) calls *tripartite*

[40] The way of marking might also be formulated saying that A and P can be marked putting them after the verb and with object or actor voice respectively, but for the purpose of examining the alignment, I am considering both arguments only in the same position as the S in the basic intransitive word order (SV), i.e. preceding the verb.

alignment: “each argument is treated differently”. Indeed, taking into account intransitive non-volitional clauses, the S is treated in one way (no marking on the verb) and A and P in a different way each (with active and object marking respectively). Dixon (1994: 71-78) devotes seven pages to the description of split-S systems, but does not identify a system like that of Madurese. He describes split-S systems as having two ways of S: S_a (marked as the A) and S_o (marked as the P), the S_o being used with the same kind of verbs that receive no marking in Madurese. Therefore, Madurese S with unmarked verbs occur with the same kind of verbs as Dixon’s S_o, but differ in that they are not marked as the P. Furthermore, Dixon (1994: 110) says that “there is no example where a special marking is used just for some S and not for A or O”, but Madurese belongs precisely to this type.

According to this, the alignment in Madurese can be considered a mixture of an active and a tripartite type: it behaves like an active type, because S and A are treated identically in intransitive clauses with volitional verbs, but in intransitives with volitional verbs, an active alignment system would be expected to treat the S like the P, and Madurese treats it in a way different from A and P, which fits the system of a tripartite alignment. We find in this way an alignment that does not fit the canonical typological alignment types.

But there is yet a minor way of marking the arguments that we have seen above: by means of prepositions. In (51), there is a preposition marking the P in a clause with AVP word order; in (53b), the preposition is marking the A with PVA order, but the preposition is optional here, and in (55), the preposition, now obligatory, is again marking the A with APV word order. The S is never marked with prepositions. As a result, if we consider the group of clauses marked with prepositions, we can see that A, P and S are each treated in a different way, because, although both A and P can be marked with a preposition, the word order is different depending on which argument bears the preposition, thus the word order differentiates A from P. In this way, in this prepositional marking system, we can speak of a tripartite alignment because each of the three arguments is treated differently.

3.6.7. INTERACTION OF ARGUMENT MARKING AND WORD ORDER

Regarding the object of this thesis, Madurese offers us several examples of the interaction of marking with word order, given that each possible word order receives a different marking. Table 6 shows the different word order combinations discussed above with the marking they receive and the number of the corresponding examples:

Table 6: Combinations of word order and marking strategies in Madurese

1	AVP	Active voice on verb	(45a), (46), (52)
2	AVP	Preposition on P (experience verbs)	(51)
3	PVA	Object voice on verb	(45b), (53a)
4	PVA	Object voice on verb plus preposition on A	(53b), (54)
5	APV	Object voice on verb plus preposition on A	(55)
6	SV	Active voice on verb (volitional verbs)	(49)
7	SV	No marking (non-volitional verbs)	(47), (48)
8	VS	Active voice on verb or no marking	(50)

Table 6 shows a system in which there is a high interaction of word order and marking. Combinations 1 and 3 are the basic orders in transitive sentences. In this system, word order is used for pragmatic purposes, as definitions in the literature of *topic* and *focus* for this kind of languages (symmetrical voice languages, as referred to by Himmelmann 2005: 112) indicate, and some marking is needed in order to know which argument is A and which is P. This implies that, when word order changes due to pragmatic reasons, the marking must change. In Madurese, as in other Western Malayo-Polynesian symmetrical voice languages, the verb carries this marking in the form of a twofold (AV and OV) voice morpheme.

The use of prepositions to mark the arguments is used in two different ways depending on which argument carries the preposition. In the case of combination 2 in Table 6, the preposition precedes the P of an experience verb. In this way, this kind of verbs is not treated syntactically as transitive because of their specific semantic content. In the case of combinations 4 and 5 in Table 6, the preposition precedes the A because it is separated from the verb, and this marking can be accompanied by a change of order as in combination 5.

In the case of intransitive clauses, the change of order from basic SV to VS implies no change of marking and is only due to "discourse-related reasons" (Davies 2010: 159).

The system of Madurese is of capital importance for reasons of comparison with Bugis, given that Bugis is also classified as a Western Malayo-Polynesian language (Adelaar 2005b: 10). The interaction of word order and argument marking in Bugis as described in section 1.1 and in Pedrós 2014 (35-44) is very different from that in Madurese. The similitude lies in the fact that there is a strong interaction between word order and argument marking, as opposed to other languages discussed in this thesis, in which the interaction is rather weaker. Himmelmann (2005: 113) roughly describes the geographical scope of his so called symmetrical voice languages and mentions "the

northern half of Sulawesi", thus excluding the southern half, where Bugis is spoken. The brief description of Bugis given in chapter 1.1 and in more detail in Pedrós (2014: 35-44) does not fit the system of this type of languages either. Nonetheless, although the language does not have a voice system, the strong interaction between word order and argument marking in Bugis may point at a possible origin of this system in relation with the system of symmetrical voice languages.

3.7. SANTALI

3.7.1. ORTHOGRAPHY

The orthography is as used in Neukom's grammar (2001), in which the IPA characters with the following exceptions are used:

– Aspiration in consonants is marked with *h* instead of IPA [ʰ] (e.g. *khange*)

– Retroflex consonants are signalled with a low point under the consonant (e.g. *ṛ*).

– *j* represents *ɟ* and *y* represents *j* (e.g. *jənum-iɲ*, *khəṭi-ge-y-e*).

Regarding stress, Neukom (2001: 8) says that stems are monosyllabic or disyllabic and rarely trisyllabic, and disyllabic stems have the stress on the first syllable unless the second syllable is heavy (CVV or CVC) and the first syllable is light ((C)V); in this case the stress falls on the second and last syllable. Neukom does not give more information on the stress' behaviour when affixes are added to the stem.

3.7.2. ARGUMENT MARKING

Santali has a case system with seven cases (nominative, dative, allative, instrumental/allative, ablative, locative and genitive) (Neukom 2001: 23), but none of them serves to differentiate A, S and P, because the three arguments are marked with the zero marker of the nominative case. The way in which these arguments are marked is through an enclitic for A and S and a verbal suffix for P, both in the form of a paradigm inflected with person and number (shown in Table 7). This means that Santali has a nominative-accusative alignment, given that A and S are marked in the same way. However, Neukom (2001: 115, 117) states that only animate arguments are marked with the subject (S/A) enclitic or the object (P) prefix, while inanimate arguments are not marked; but there is an exception: the suffix for inanimate objects *-k'-*, which "appears only in the applicative construction" (Neukom 2000: 97). The applicative construction "allows a participant other than the patient to be cross-

referenced on the verb" Neukom (2000: 105; 2001: 121). Neukom (2000: 108) calls this other element *dative object* (2000: 108). Nonetheless, Neukom shows examples with the suffix *-k'-* in verbs in which this dative object rather has the function of a P than of a R, as in (56):

(56) daka cak'-em kurud-a-**k'**-kan-a, jɔm-me.
food what-2SG.S hate-APPL-**INAN**-IPFV-FIN eat-2SG:IMP
'Why do you hate the food? Eat.' (said to children who don't eat)

(Neukom 2000: 108)

Therefore, the inanimate suffix can function as a P cross-reference marker only in applicative constructions, although it can also occur in a more R-like function, as in (57), with the construction *add to the statement*:

(57) katha alo-m juṭuc'-a-**k'**-a. (Neukom 2000: 108)
speech PROH-2SG.S add-APPL-**INAN**-IPFV-FIN
'Don't add anything to the statement.'

Table 7: Paradigm of subject enclitics and object suffixes in Santali (adapted from Neukom 2001: 113, 115)

	SUBJECT ENCLITICS			**OBJECT SUFFIXES**		
	SG	**DU**	**PL**	**SG**	**DU**	**PL**
1 excl	=ɲ	=liɲ	=le	-iɲ-	-liɲ-	-le-
1 incl		=laɲ	=bo(n)		-laŋ-	-bon-
2	=m	=ben	=pe	-me-	-ben-	-pe-
3	=e	=kin	=ko	-e-	-kin-	-ko-

An animate noun in Santali is "any entity that has a soul […], which includes not only human beings and animals, but also sun, moon, stars and spirits" (Neukom 2001: 22). The general lack of marking for inanimate nouns could result in ambiguity: even though transitive clauses with two inanimate arguments are rather infrequent in any language, they do exist. Ambiguity can also occur with animate nouns when A and P have the same person and number. In these cases, although not specified in Neukom's (2001) grammar, it is reasonable to think that word order, explained below, must disambiguate when the arguments are not retrievable from context.

The subject enclitics are usually attached to the word preceding the verb, as can be seen in (58). They can also be attached to the verb, but this occurs mainly in one-verb clauses, as in (59) (Neukom 2001: 114).

(58) khange Jhɔ̃ṛɛ̃-dɔ=**e** dəṛ-ket'-a. (Neukom 2001: 113)
then Jh.-TPC-**3SG.S** run-PAST.ACT-IND
'Then Jhore ran away.'

(59) cala-k'-a-**m** —hɛ̃, cala-k'-a-**ɲ** (Neukom 2001: 113)
go-MID-IND-**2SG.S** —yes, go-MID-IND-**1SG.S**
'Will you go? —Yes, I shall go.'

The object suffix can have two functions: the most frequent is to mark the patient of an action (60); the other one is to mark the experiencer (mainly with physical sensation) of an experience verb (61) (Neukom 2001: 116):

(60) teheɲ-dɔ-ɲ ɲam-akat'-**ben**-a. (Neukom 2001: 116)
today-TPC-1SG.S find-PRF:ACT-**2DU.O**-IND
'Today I have found you two.'

(61) jənum-iɲ rɔk'-en-te, (Neukom 2001: 116)
thorn-1SG.S stitch-PAST:MID-CONV

cabac'cubuc' haso-ed-**iɲ**-kan-a jaŋga.
stabbing.pain hurt-IPFV:ACT-**1SG.O**-IPFV-IND foot

'Since I got a thorn, my foot is aching me with a stabbing pain.'

However, Neukom (2001: 116) says that this way of marking an experiencer is not frequent and that it is more usual to mark the experiencer with the subject enclitic using the middle voice, as in (62):

(62) uni haṛam-ren kimin-ko-dɔ-**ko** əṛis-**ok'**-kan-a.
that.AN old.man-GEN.AN daughter-in-law-PL-TPC-**3PL.S** worry-**MID**-IPFV-IND
'The old man's daughters-in-law are worrying.' (Neukom 2001: 116)

3.7.3. WORD ORDER

3.7.3.1. Basic word order and change through an afterthought

Regarding word order, Neukom (2001: 174) says that "in the simple sentence, subject and objects normally precede the verb, the object follows the subject", and that "the position after the predicate can be used as afterthought". The categories described by Neukom as *subject* and *object* coincide with A/S and P respectively. Therefore, the word order in Santali is APV and SV, although the order can change when an afterthought occurs after the verb. Neukom (2001: 174) gives only one example of afterthought in which there is a change of constituent order —in other examples, the afterthoughts are not arguments. In **(63)**, the afterthought is the T:

(63) onate iɲ-hɔ̃-ɲ sɛn-len-khan, jənic' khəṭi-ge-y-e
therefore I-also-1SG.S go-IRR:MID-when probably surely-FOC-Y-3SG.S

em-a-ɲ-a **abon ləgit' kicric' ar jɔm-ak'-dɔ.**
give-APPL-1SG.O-IND **1PI for clothes and eat-NMLS:INAN-TPC**

'Therefore, when I go there as well, he will quite probably give me clothes and food for us.' (Neukom 2001: 174)

Example (63) is a clear case of change of place of an argument: the object, in this case the T because this is a ditransitive clause where the R is cross-referenced on the verb[41] (*a-ɲ*, 'APPL-1SG.O'), is placed after the verb. Neukom does not give an example of P after the verb as an afterthought. However, (63) shows us that an argument, the T, is moved from its normal pre-verbal position to post-verbal position, and no marking different from the usual in ditransitive sentences can be observed. The most logical deduction is that the P can occur after the verb as an afterthought also without change of marking.

We should take into account what an afterthought is: something that is thought *after*. This means that the sentence is uttered and then something is remembered and added. Therefore, it is very questionable if an afterthought can be considered a part of a sentence, given that the sentence is formed without intention of including this element in it, and it appears after the sentence has been built with the rules of the grammar of the language. In (63), the afterthought is the T and the meaning of the sentence cannot be understood without it unless by context, just for that reason it is possible that the T is implicit in the context, but, after uttering the sentence, the speaker thinks that maybe it is not so implicit and then he or she utters the T for clarification. My opinion is that an afterthought, considered as something that is remembered after the sentence is uttered and added thereafter, cannot be considered a part of the sentence and therefore should not be considered in a study of word order, unless an afterthought in a given language is something different. Since no explanation in that sense is given in Neukom's grammar (2001), I will not consider that Santali word order changes because of an afterthought.

[41] In ditransitive clauses, the R and not the P is cross-referenced on the verb with the applicative suffix *-a-* and a paradigm of person/number applicative suffixes. I will not discuss this system here because ditransitive clauses are not the subject of this thesis.

3.7.3.2. Passive constructions

There is yet a kind of sentence in which the order can change: the passive. A passive construction similar to English has not appeared in the languages discussed until now, but it does occur in Santali. If we are dealing with the order of A, V and P, there is no doubt that in an English passive sentence, such as *The clothes were found by them*, the P (*the clothes*, syntactic subject) is in the first place, and the A (*by them*, prepositional phrase) is after the verb, resulting thus in a word order PVA, although the order of the syntactic arguments continues being the basic English SVO. If we take the semantic arguments A and P into account, there is no doubt that the English passive implies that A and P change order. Therefore, I will discuss shortly the Santali passive here.

The Santali passive is formed with the middle voice, which has also other uses (intransitive, reflexive, basic meaning, inchoative), "an exceptionally wide range of meanings" according to Neukom (2001: 103), which are not going to be studied here. There are only three examples of passive sentences with nominal phrases as A and P in Neukom's grammar (2001), but all of them show a word order different from the basic Santali APV:

(64) oṛak' rəput'-en-a hɔe-dak'-te.
house break-PAST:MID-IND wind-water-INST
'The house was damaged by the rainstorm.'

(65) uni rəni-dɔ uni sɔdagɔr-ṭhɛn ti-re-y-e
that.AN queen-TPC that.AN merchant-DAT hand-LOC-Y-3SG.S

sap'-oco-y-en-khan...
seize-CAUS-Y-PAST:MID-when

'When the queen was caught in the arm by the merchant...'

(66) ḍhiŋki-dɔ mit' hɔṛ-te baŋ uskəu-k'-a.
rice-stamp-TPC one person-INST NEG move-MID-IND
'The rice-stamp cannot be moved by one person.'

In (64), the order is PVA, and in (65) and (66), the order is PAV. In (64) and (66), the A is marked with the instrumental case, and in (65), with the dative case. What these three examples show us is that the P is placed first. This occurs also in passive constructions in European languages, which are normally used when the A is not considered important or the P is highlighted. Although no explanation about the reasons of the choice of the passive construction is given in Neukom 2011, it seems that it is used in a way similar to the passive construction in European languages.

3.7.3.3. Conclusions

According to all what is exposed above, Santali has a basic word order SV/APV and the only way to change it is with a passive construction —as said above, I am not considering the afterthought. After examining the two texts reproduced at the end of Neukom's grammar (2001: 202-223), the word order SV/APV seems quite evident and no changes have been found. Therefore, since a category *subject* is valid to refer to Santali S and A, a rigid word order SOV[42] can be postulated for Santali. Regarding the semantic arguments of S, A and P, in passive constructions, the P can be moved to the first position of the clause, and the A, marked with the dative or the instrumental case, can precede or follow the verb.

The passive construction can be considered a change of marking in relation to word order, given that the verb is marked with a special morpheme (middle voice in Santali) or a special verbal construction (passive periphrasis in English). Nonetheless, the passive is a well-studied grammatical occurrence cross-linguistically and, although it undoubtedly implies a change of marking in the verb involving change of word order, it should not be the goal of this thesis because a whole different thesis could be devoted to the study of passive constructions cross-linguistically. However, it is worth to mention here without entering in detail.

[42] Here, S takes a meaning different from that used elsewhere in the thesis: it represents Santali S and A.

CHAPTER FOUR
COMPARISON AND ANALYSIS

4.1. SUMMARY OF THE KEY DATA OF EACH LANGUAGE

Before summarising the data found in each language, it is important to note that all languages have a nominative-accusative alignment except Madurese, which has a complex alignment system described in section 3.6.6. This implies that, in all languages but Madurese, a syntactic category *subject* is identifiable as representing the semantic categories S and A. Therefore, a word order as described in this thesis, having into account the semantic arguments A and P (e.g. APV), is in these languages equivalent to a word order defined having into account the syntactic arguments S and O[43] (e.g. SOV). Besides the exception of Madurese, we have the case of the Santali passive construction, in which a clause with a word order defined semantically as PAV has a word order defined syntactically as SOV. This explanation is important so that the reader knows beforehand how the syntactic order in each language is according to the description given below in terms of semantic arguments. The key data obtained from each language are as follows:

Santali: Santali marks A and S with an enclitic on the first word before the verb or on the verb in clauses where the verb is the only word. The P is marked with a verbal suffix. Nonetheless, only animate S, A and P are cross-referenced in this way. Inanimate P are cross-referenced with a verbal suffix only in the applicative construction. The basic word order is APV and no changes are observed except in the passive construction, in which the order is PAV. Verbs are marked with an active or a middle voice in active or passive constructions respectively.

---

[43] As said in footnote 1, I have used along the thesis the terms S (subject), A (agent) and P (patient) to refer to the semantic roles defined by Comrie (1989: 58-59): S is the only argument of an intransitive clause, A and P are the most agent-like and patient-like arguments of a transitive clause respectively. In this chapter and in chapter 5, I will also use the terms S (subject) and O (object) as morphosyntactically identifiable arguments. In this way, as explained in section 3.7.3.2 with the discussion of the Santali passive, in an English passive sentence as *The clothes were found by them*, *the clothes* is the syntactic S and the semantic P, and *by them* is the semantic A and, as syntactic argument, a prepositional phrase. In the English active sentence *They found the clothes*, *they* is both A and S, and *the clothes* is both P and O. I use this twofold terminology here in order to make clearer the explanation of alignment and word order, both semantically and syntactically regarded, in the discussed languages.

Figuig Berber: In Figuig Berber, S/A and P are cross-referenced on the verb with different sets of affixes, but the arguments are also marked with case. In the basic VAP/VS word order, S/A and P are marked each with a different case, but in the non-basic word order AVP/SV, S/A and P are marked with the same case. This implies that the arguments in the basic order receive differential case marking, but that they are marked identically in the non-basic word order. This feature resembles the interaction between marking and word order in Bugis. Object pronouns (P, T and R) also receive different marking if they precede or follow the verb.

Savosavo: In Savosavo, A and S are marked on the noun phrase with a nominative suffix, while the P receives no case suffix. This contradicts Greenberg's universal 38 (1966). Moreover, the A/S is cross-referenced with a Wackernagel enclitic and the P with an obligatory affix on the verb. Word order is a rigid APV that cannot be changed. Therefore, there is no change of marking in relation to word order change because there is no change in word order.

Madurese: As said above, Madurese is the only studied language without a nominative-accusative alignment, given that it has a complex alignment described in section 3.6.6. Madurese transitive clauses have two basic word orders, AVP and PVA, and each is marked on the verb with a different voice morpheme or with a preposition on one of the arguments (see Table 6). Moreover, a non-basic APV order is possible and is marked with a preposition before the A. The basic word order for the intransitive clause is SV, but it can change to VS without any marking. The system of the transitive clause implies that Madurese shows a very high interaction between word order and argument marking, the highest of all the studied languages.

Itonama: Itonama marks A/S and P with agreement affixes on the verb. Word order with nominal phrases is VAP. Given that Itonama is a polysynthetic language in which a clause can frequently be formed only by a verb, the order of agreement affixes inside the verbal complex has been studied instead of the constituent order inside the clause. This approach is grounded in section 3.4.3.2. In independent clauses, 1st and 2nd person A and S are cross-referenced with a prefix in the first position of the verbal complex, and 1st and 2nd person P with a suffix in the last position, while 3rd person arguments are never marked. In dependent clauses, the 3rd person is marked only as

A/S with a prefix. However, in independent clauses, there is a person hierarchy that prevents a 1st or 2nd person P from being cross-referenced as a suffix when the A/S is 3rd person. When this person combination occurs, the 1st or 2nd person P is cross-referenced with a special set of prefixes. This represents a change of order inside the verbal complex due to a person hierarchy that implies a change of marking, given that there is a different set for P prefixes.

Lao: In Lao, the main indicator of which argument is A or P is context. Nonetheless, the AVP word order serves to disambiguate where the arguments cannot be retrieved from context. Word order changes are grammatical and frequent, but when the change is to VPA, PVA or VS (verb or cluster of V and P before A/S), a prosodic mark is inserted. The fact that Lao is a tonal language makes this prosodic mark more important than in non-tonal languages.

Tuvan: In Tuvan, P noun phrases are marked with an accusative suffix only when the P is definite or specific. Indefinite P bear no marking, the same as all A/S, or putting it in other words, A/S and indefinite P receive the zero marking of the nominative case. Tuvan has the rigid APV word order common in the Turkic languages, but has the particularity that, in main clauses, the S/A 1st or 2nd person pronoun is postposed to the verb with any TAM morpheme but with PAST II/PAST (different terminologies in the two works used, Anderson & Harrison 1999 and Mawkanuli 2005). Verbs in subordinate clauses do not receive TAM marking (they appear in the form of participles or converbs), so that the above statement "in main clauses" is interchangeable with "in clauses marked with TAM". This post-verbal pronoun is an archaic feature of Tuvan, given that, in other Turkic languages, it has become a suffix, and in Tuvan it is in process of grammaticalisation, because it is already an enclitic in some dialects. Therefore, the Tuvan system shows a basic APV word order[44] and a non-basic PVA order used with A/S 1st and 2nd person pronouns in main clauses and with all TAMs but PAST II/PAST.

To end with, it is also important to summarise here the features of **Bugis**, which led to the research question of this thesis. Bugis transitive clauses show two word orders: a

[44] The APV word order is considered basic because it occurs in most cases: all clauses with A/S non-pronominal phrases or 3rd person pronouns, subordinate clauses and main clauses in PAST II/PAST.

basic VPA and a non-basic AVP. The VPA order receives obligatorily a P suffix on the verb and an enclitic article on the A, while it optionally receives an A prefix on the verb and an enclitic article on the P, both of which have to co-occur, i.e. the occurrence of the A verbal prefix implies the occurrence of the enclitic article on the P. On the contrary, the AVP order receives no marking of any kind. In both cases, word order indicates which argument is A and which P, given that word orders *PVA and *VAP are ungrammatical. In intransitive clauses, the basic word order is SV, and the non-basic, VS. The latter is marked with a S agreement suffix on the verb and an enclitic article on the S. The SV order bears no marking. All these descriptions are valid for clauses without pronouns. The system with pronouns is too complex to be summarised here (see Pedrós 2014: 38-41).

4.2. COMPARISON

Regarding basic word order, there are three languages with APV/SV order (Santali, Savosavo and Tuvan), two with VAP/VS order (Figuig Berber and Itonama), one with AVP/SV order (Lao), and one with two basic orders in transitives, AVP and PVA, and one in intransitives, SV (Madurese)[45].

Santali and Savosavo have in common that they cross-reference their arguments in a similar way: Santali attaches an A/S enclitic on the first word before the verb, while Savosavo attaches a like enclitic on the first constituent of the clause. Furthermore, both languages have P agreement affixes. These two languages have also in common a rigid word order APV/SV, which can be changed in Santali only with the passive construction. In any case, both have a rigid SOV order of syntactic arguments. This word order (semantic APV/SV, syntactic SOV) is also shared by Tuvan, which is rigid as well but can be broken by a post-verbal 1st or 2nd person pronoun. In this way, we have three languages with a rigid syntactic SOV word order, which changes only in Tuvan by means of a post-verbal pronoun that is on the way to become a suffix through grammaticalisation, as described in section 3.1.5.

Figuig Berber shows a clear relation between argument marking and word order, given that the A receives a different case marker depending on whether it precedes or follows the verb. With the basic word order (VAP), the A is marked with a case different from the P, and when the non-basic order (AVP) occurs, A and P are marked

[45] I will consider Bugis for comparison with the other languages regarding only the interaction of word order and argument marking.

with the same case. This system shows a differential marking in the basic order, something that also happens with Bugis. Therefore, both languages have in common that the basic word order receives more marking. It has to be taken into account as well that, with the basic word order, A and P are at the same side of the verb in both languages, which can be a reason for the differential marking.

Madurese shows a high interaction between word order and marking due to the voice system that marks on the verb which argument is in the first position of the clause. It may be argued that this voice system must be regarded as a combination of active and passive sentences. Nevertheless, Davies (2010: 257) argues against considering the clause marked with object voice as a passive. Davies says that object voice clauses make 50-60% of all clauses, while this percentage is considerably lower in languages with prototypical passives (7.8% in oral and 17.9% in written English). Additionally, Davies (2010: 257-258) argues that, in prototypical passives, the percentage of clauses with overt agents is low (13-20% in English, 17.8% in German), while it is much higher in Madurese (60-70%). In addition to this, in a prototypical passive, the order of the semantic arguments is changed (e.g. English AVP>PVA), but the order of the syntactic arguments remains the same (e.g. English SVO). In Madurese, there is no point in talking of syntactic arguments because the arguments are not cross-referenced, unless one wants to call *subject* the element in first position, as Davies does, but in this case, if we want to talk about word order, we find a circularity in that the subject is defined through its position, so that there is no point in talking about the position of the subject in the clause. According to all this, the voice system of Madurese cannot be regarded as an active-passive opposition and thus cannot be compared with Santali, which appears to have a more prototypical passive.

In Itonama, the order inside the verbal complex changes due to a person hierarchy, which causes a change in the form of the subject (A/S) affixes. No similar hierarchy has been found in the rest of the studied languages.

Lao has also no similarities with the other languages, given that the change of order is accompanied by a prosodic marker. The importance of this marker for the speaker remains unknown, because Enfield (2007) only observes the presence of the marker. A more in-depth research could show if the inexistence of this prosodic mark makes the clause difficult to interpret by a speaker, i.e. if it is obligatory.

In this way, we have three languages (Figuig Berber, Madurese, and Lao) in which a change in order is optional for the speaker, but where a change in marking does appear when the order is changed. This change in marking is at the core of the

grammar in Madurese, is quite evident in Figuig Berber and has only a prosodic marker in Lao —as said above, the importance of this marker is unknown.

In the other four languages (Santali, Savosavo, Tuvan and Itonama), a change in word order is not optional. The change is not possible in Santali and Savosavo —in Santali, considering the syntactic order SOV due to the passive construction, in which the semantic arguments change order. In Tuvan and Itonama, a change is syntactically conditioned, i.e. the speaker has to perform the change when certain conditions concur in the clause (TAM marking and a kind of pronoun in Tuvan and person hierarchy in Itonama). The case of Tuvan resembles more that of Santali and Savosavo, given that, in Itonama, the fact of considering the order inside the verbal complex makes it special. Moreover, Santali, Savosavo and Tuvan share the same word order (APV), which is quite rigid in the three languages —in Santali, the syntactic SOV is rigid, not the semantic order APV, due to the passive construction.

CHAPTER FIVE
CONCLUSIONS

5.1. EVALUATING THE REPRESENTATIVENESS OF THE SAMPLE

In the preceding pages, some subjects that have been widely studied in the literature have been discussed, namely alignment and word order.

Regarding word order, Siewierska (2001: 16553) says that "the cross-linguistic investigations conducted since the 1960s identify the same basic word order preferences consistently; the most common basic order is SOV (+/-45 percent), followed closely by SVO (+/-35 percent), and then VSO (+/-10 percent). The three object-before-subject orders are all highly uncommon as basic orders, with VOS occurring slightly more frequently than OVS, and OSV being extremely rare." Taking into account the languages studied in this thesis, 43% of the languages are APV/SOV (semantic and syntactic arguments coincide in all languages but Madurese) and 29% are VAP/VSO. Madurese is AVP or PVA, so I will count half a language for each order. Then there are 21% AVP languages (half Madurese and Lao), or 14% SVO counting only Lao, and 7% PVA[46]. Comparing with Siewierska's data, SOV languages have a similar percentage here and in her data (43%/45%), while the percentages of VSO (29%/10%) and SVO (14%/35%, or 21%/35% if we include AVP) are quite different. The half a language PVA appears rare (7%), as OVS in Siewierska (2001).

Regarding alignment, all languages but one in this thesis have a nominative-accusative alignment, which makes 86% of the total. Dixon (1994: 40) says that the accusative alignment is the commonest in the languages of the world, the ergative alignment is not uncommon and the tripartite alignment is extremely rare. Nonetheless, in a more detailed study, Siewierska (1996: 155) gives the figures for accusative alignment in a sample of 237 languages as 37%, 27% and 55% in pronouns, nouns and agreement respectively. For neutral alignment, she gives the figures of 43%, 49% and 23%. These figures raise the suspicion that maybe she is considering languages as Lao, which mark the arguments only through word order, as neutral alignment. Unfortunately, although Siewierska (1996: 153) defines the different alignment types in terms of how A and P are treated in relation with the S, she does not

[46] In a study as Siewierska 2001, where syntactic arguments are taken into account, Madurese should count as a language with no basic word order.

explain what is meant with *being treated*, i.e. if word order is included or if she refers only to case markers, pronouns and agreement affixes.

The data for the languages in this thesis are compared with cross-linguistic data just to give an idea of how representative this sample of seven languages may be. In principle, a sample of only seven languages can show no cross-linguistic trend. Nevertheless, if a given feature exists in 80%-90% of the languages of the world and it does not appear in a sample of seven languages, the conclusion would be that the choice of languages has been a very unlucky one. The results of the languages of this thesis compared with cross-linguistic data of the well researched features of alignment (Dixon 1994, Siewierska 1996) and word order (Siewierska 2001) show that, although the data of my sample deviate from those of larger cross-linguistic samples, they do not deviate much. The commonest types in the languages of the world (accusative alignment and SOV word order) are also the commonest in this thesis' sample. In my view, the results obtained here can be interpreted as follows: if a trend is clearly detected, it is probable that this trend appears in much more languages, in spite of the fact that an unlucky coincidence can cause a choice of languages that present a rare feature.

5.2. WHAT TO CONCLUDE ON THE STUDIED FEATURE

The idea for this thesis comes from a feature of Bugis that in principle seems rare because it is unknown in other languages to me or to other people whom I have consulted. On the other hand, it seems that this feature has not been studied in detail in the literature. The feature in question, putting it shortly, concerns a clear correspondence between two different word orders and two different markings.

As explained above, in three languages of the sample (43%, Figuig Berber, Madurese and Lao), the order can be changed freely by the speaker and the change requires some additional or different marking. In two languages (29%, Itonama and Tuvan), there is an obligatory change of order under certain syntactic conditions that requires marking. In the two other languages (29%, Santali and Savosavo), the order is rigid and cannot be changed. This last group can be reduced only to Savosavo (14%) if we consider the Santali passive construction. In this case, Santali would form part of the first group, given that the order can be changed freely to form the passive construction and it requires special marking. This would increase the first group to 57% of the total.

The main conclusion to draw from these data is that most of the studied languages change marking when word order changes, whether the change is optional or syntactically motivated. Of the seven studied languages, only Savosavo shows no trace of differential marking for different word orders because its word order never changes, as was concluded in section 3.5.3. In languages in which it is optional to change order, the speaker can move an argument to first position to highlight it, i.e. to treat it as topic or focus, but this movement requires a differential marking that indicates that the order is changed. In this sense, the Itonama change seems to highlight obligatorily the higher-ranked persons in the hierarchy, namely 1st and 2nd person.

The Santali passive deserves some consideration. A passive construction is, as defined by Dixon (1987: 146)[47], a clause in which the syntactic subject is the P, the A is marked as a non-core argument and the verb bears a special marking. A similar definition is given by Siewierska (2013). The Santali passive construction has these features (the A is marked with a dative or instrumental case and the verb receives middle voice marking). This implies that a language with a passive construction needs to have a syntactic subject that correlates normally with the A, given that a clear subject function must exist to say that the P becomes subject in the passive clause. If we consider the word order with semantic arguments, the passive construction implies a change of order with a change in marking (APV changed to PAV in Santali, or AVP changed to PVA in English). However, if we consider the syntactic arguments, there is no change of order (SOV in Santali and SVO in English). Passive constructions exist in many languages[48] and are recognisable by the features mentioned above. Their cross-linguistic study could be the object of a different thesis; that is why they are not studied in detail in this thesis. Nonetheless, the passive subject overlaps with this thesis' subject and this short mention is convenient in order to clarify concepts.

It is important to remark that a differential marking occurring with different word orders is by no means necessary. Consider the following German examples:

(67) a. Der Mann küss-t die Frau.
DA.M.SG.NOM man kiss-3SG.PRS DA.F.SG.ACC woman
'The man kisses the woman.'

[47] Dixon's description is summarised here and adapted to the terminology of this thesis.
[48] Siewierska (2013) gives the figure of 44% of a sample of 373 languages. She adds that languages with passive constructions are most common in Eurasia and Africa and are regularly found in the Americas, particularly in North America.

(67) b. Die Frau küss-t der Mann.
DA.F.SG.ACC woman kiss-3SG.PRS DA.M.SG.NOM man
'The man kisses the woman.'

(68) a. Die Frau küss-t den Mann.
DA.F.SG.NOM woman kiss-3SG.PRS DA.M.SG.ACC man
'The woman kisses the man.'

(68) b. Den Mann küss-t die Frau.
DA.M.SG.ACC man kiss-3SG.PRS DA.F.SG.NOM woman
'The woman kisses the man.'

In the two examples in (67), the order is reversed without additional marking. The case of the masculine article —accusative and nominative have the same form in the feminine article— indicates that it is the man who kisses. In the two examples in (68), the order is also reversed without additional marking and it is clear in both cases that the woman is kissing. The normal order in German for these short sentences is SVO/AVP, but it can be reversed without additional marking and the arguments are identified by the same case marking in both orders.

Therefore, an additional marking with different word orders as we have seen in Figuig Berber, Madurese, Itonama, Lao and Tuvan can be considered redundant in most cases. The distinctive marking with a different word order could serve to disambiguate only when a language differentiates the arguments through word order, as in English; or when the marker on the verb indicates the arguments according to their position, as in Madurese. In the case of English, a change of order of the semantic arguments A and P requires a differential marking (a passive periphrasis) in order to indicate which is A and which P. In the case of Madurese, where the marking on the verb indicates whether the argument before the verb is A and P, any change of order requires change of marking so as to identify the arguments. In other languages, where the arguments are distinguished through case (German, Tuvan, Figuig Berber, Savosavo), through cross-reference affixes or clitics (German, Tuvan, Figuig Berber, Itonama, Savosavo, Santali) or through context (Lao)[49], a distinctive marking with different word orders is redundant, because there are other features that indicate who is A and who is P.

[49] The case of Lao could be challenged here. The prosodic marker that occurs with change of order may be considered a disambiguating element when the arguments cannot be retrieved from context. However, if the context indicates clearly which the arguments are, this prosodic marker should be considered redundant.

5.3. FINAL CONCLUSION

We have seen in a sample of seven languages that, in three of these languages (Figuig Berber, Madurese and Lao), the word order can be changed freely with a differential marking; in two of them (Itonama and Tuvan), there is a change —of affixes inside the verbal complex in Itonama— required by the syntactic environment and this change is also marked differently; in the other two (Santali and Savosavo), there is a strict word order that does not change, with the peculiarity of the Santali passive discussed above. In this sample, all languages that can change order clearly receive differential marking when changing order. The very small size of the sample does not allow cross-linguistic conclusions, but the facts shown here should raise the awareness that it is very likely that the researched feature occurs widely cross-linguistically.

The fact that a language signals a change of order, redundantly in many cases, seems to imply that the speaker is trying to call the attention to the fact that a change has happened. It seems that it does not matter that other grammatical features indicate who is A and P, nor that the clause is perfectly understandable with the change of order: in a way, it seems that the speaker intuitively knows that there is a basic order and feels the need to announce that it has changed.

The study of how changes of order are marked, in which cases and how often could shed interesting results when many more languages are studied. This could be thus a fertile field for future research, in terms of language structures and of functional motivations.

REFERENCES

ADELAAR, Alexander, 2005a. "Malayo-Sumbawan", *Oceanic Linguistics*, vol. 44, no. 2: 357-388.

—, 2005b. "A historical perspective", in *The Austronesian languages of Asia and Madagascar*, ed. by Alexander Adelaar and Nikolaus P. Himmelmann, 1-42. London: Routledge.

ANDERSON, Gregory D.S., 1998. *Xakas* [Languages of the world / materials, vol. 251]. München: LINCOM Europa.

ANDERSON, Gregory David, & HARRISON, K. David, 1999. *Tyvan* [Languages of the world / materials, vol. 257]. München: LINCOM Europa.

BLUST, Robert Andrew, 2014. "Austronesian languages", *Encyclopædia Britannica online academic edition.* Available online at www.britannica.com. Accessed on 4 May 2014.

COMRIE, Bernard, 1989. *Language universals and linguistic typology*. Oxford: Basil Blackwell.

CREVELS, Mily, 2011. "Who did what to whom in Magdalena (Itonama)", *International Journal of American Linguistics*, vol. 77, no. 4: 577-594.

—, 2012. "Itonama", in *Lenguas de Bolivia. Tomo II: Amazonía*, ed. by Mily Crevels and Pieter Muysken, 233-294. La Paz: Plural editores.

DAVIES, William D., 2010. *A grammar of Madurese* [Mouton grammar library, vol. 50]. Berlin: De Gruyter Mouton.

DIXON, R.M.W., 1987. "Studies in ergativity", in *Lingua*, vol. 71: 1-16.

—, 1994. *Ergativity*. Cambridge: Cambridge University Press.

DRYER, Matthew S., 1992. "The Greenbergian word order correlations", *Language*, vol. 68, no. 1: 81-138.

DRYER, Matthew S. & HASPELMATH, Martin (eds.) 2013. *The world atlas of linguistic structures online*. Max Planck Institute for Evolutionary Anthropology. Available online at wals.info. Accessed on 21 April 2014.

DUNN, Michael, GREENHILL, Simon J., LEVINSON, Stephen C. & GRAY, Russell D., 2011. "Evolved structure of language shows lineage-specific trends in word-order universals", *Nature*, vol. 473: 79-82.

DUNN, Michael, & TERRILL, Angela, 2012. "Assessing the lexical evidence for a Central Solomons Papuan family using the Oswalt Monte Carlo Test", *Diachronica*, vol. 29, no. 1: 1-27.

EMENEAU, M.B., 1956. "India as a linguistic area", *Language*, vol. 32, no. 1: 3-16.

ENFIELD, N.J., 2005. "Areal linguistics and mainland South-East Asia", *Annual Review of Anthropology*, vol. 34: 181-206.

—, 2007. *A grammar of Lao* [Mouton grammar library, vol. 38]. Berlin: Mouton de Gruyter.

GHOSH, Arun, 1994. *Santali - a look into Santal morphology*. New Delhi: Gyan Publishing House.

GREENBERG, Joseph H., 1966. "Some universals of grammar with particular reference to the order of meaningful elements", in *Universals of language*, ed. by Joseph H. Greenberg, 73-.113. Cambridge, Massachussets: MIT Press.

HIMMELMANN, Nikolaus P., 2005. "Typological characteristics", in *The Austronesian languages of Asia and Madagascar*, ed. by Nikolaus P. Himmelmann and K. Alexander Adelaar: 110-181. London: Routledge.

HOPPER, Paul J. & TRAUGOTT, Elizabeth Closs, 2003. *Grammaticalization*. Cambridge: Cambridge University Press.

JOHANSON, Lars, 1998. "History of Turkic", in *The Turkic languages*, ed. by Lars Johanson and Éva Á. Csató, 81-125. London: Routledge.

KEYES, E. Jane, 2014. "Thailand", *Encyclopædia Britannica online academic edition.* Available online at www.britannica.com. Accessed on 7 April 2014.

KOSSMANN, Maarten G., 1997. *Grammaire du parler berbère de Figuig (Maroc oriental)* [M.S. — Ussun amaziγ, vol 10]. Paris: Éditions Peeters.

—, 2014. Figuig Berber: basic grammatical structures. Unpublished material for the MA course "Samples of linguistic structures", March 2014, at Leiden University. Available online at tinyurl.com/omvvsn3.

LAFONT, Pierre-Bernard, 2014. "Laos", *Encyclopædia Britannica online academic edition.* Available online at www.britannica.com. Accessed on 7 April 2014.

LEWIS, M. Paul, SIMONS, Gary F., & FENNIG, Charles D. (eds.), 2014. "Santhali", in *Ethnologue: languages of the world, seventeenth edition*. Dallas: SIL International. Available online at www.ethnologue.com. Accessed on 12 May 2014.

MASICA, Colin P., 1976. *Defining a linguistic area. South Asia.* Chicago: The University of Chicago Press.

MAWKANULI, Talant, 2005. *Jungar Tuvan texts* [Indiana University Uralic and Altaic Series, vol. 170]. Bloomington: Research Institute for Inner Asian Studies, Indiana University.

MENGES, Karl H., 1995. *The Turkic languages and peoples. An introduction to Turkic studies* [Veröffentlichungen der Societas Uralo-Altaica, vol. 42]. Wiesbaden: Harrassowitz.

NEUKOM, Lukas, 2000. "Argument marking in Santali", *The Mon-Khmer Studies Journal*, vol. 30: 95-113. Available online at www.mksjournal.org.

—, 2001. *Santali* [Languages of the world / materials, vol. 323]. München: LINCOM Europa.

PEDRÓS CABALLERO, Toni, 2014. Grammar sketch of Bugis. Final work for the MA course "Fieldwork methods", at Leiden University, September-December 2013. Available online at tinyurl.com/odqsyoe.

ROSS, Malcolm, 2005. "Pronouns as a preliminary diagnostic for grouping Papuan languages", in *Papuan pasts: cultural, linguistic and biological histories of Papuan-speaking peoples*, ed. by Andrew Pawley, Robert Attenborough, Robin Hide and Jack Golson, 15-65. Canberra: Pacific Linguistics.

SCHÖNIG, Claus, 1998. "South Siberian Turkic", in *The Turkic languages*, ed. by Lars Johanson and Éva Á. Csató, 403-416. London: Routledge.

SIEWIERSKA, Anna, 1996. "Word order and alignment type", in *Sprachtypologie und Universalienforschung (STUF)*, vol. 49: 149-176.

—, 2001. "Word order", in *International encyclopedia of the social and behavioral sciences*: 16552-16555. Amsterdam: Elsevier. Available online at www.sciencedirect.com.

—, 2013. "Passive constructions", in *The world atlas of language structures online*, ed. by Matthew S. Dryer & Martin Hasplemath. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at wals.info/chapter/107. Accessed on 28 May 2014.

SIRK, U.H., 1979. *La langue bugis* [Chahiers d'archipel 10], translated from Russian by Nicole Pacory. Paris: Association Archipel.

SMYTH, D.A., 2006. "Tai languages", *Encyclopedia of language and linguistics*, ed. by Keith Brown: 481-482. Available online at www.sciencedirect.com. Accessed online on 7 April 2014.

TRYON, Darrell Trevor, & HACKMAN, Brian D., 1983. *Solomon Islands languages: an internal classification* [Pacific linguistics, series C, no. 72]. Canberra: Research School of Pacific Studies, Australian National University.

WEGENER, Claudia, 2012. *A grammar of Savosavo* [Mouton grammar library, vol. 61]. Berlin: De Gruyter Mouton.