

Black and white Languages

Leon Stassen

1 Introduction

Seen from a certain perspective, the cross-linguistic variation which language typologists aim to elucidate can be distinguished into two different kinds. First, there are parameters of cross-linguistic variation whose distribution is most likely to be rated as *free*, be it geographically, or typologically, or both. An often cited example of such a parameter is the distribution of the dual across languages. As is well known, there are languages which mark their noun phrases (or at least some of them) for dual number, whereas in other languages no such marking is available. Now, it seems that the languages which do have such dual marking do not, in all probability, form a natural class, neither with respect to their general typological status, nor with respect to their geographical distribution. Thus, dual marking appears to occur in languages that are widely different from each other in their typological make-up. Hence, there are no language features on the basis of which the occurrence or non-occurrence of dual marking in a language can be predicted, and, conversely, the occurrence or non-occurrence of dual marking in a language does not seem to predict any other structural characteristics. Moreover, the spread of the phenomenon of dual marking does not correspond to geographical patterns: dual marking seems to be ‘sprinkled’ across the globe without any discernible regularity, and there are hardly any major language families in which dual number marking is mandatory for all members. For these reasons, at the present state of our knowledge it seems best to view the phenomenon of dual number marking as some typological ‘extra’. That is, it is a feature that languages may ‘want’ to have, but its occurrence is probably not licensed by any considerations of a genetic, areal, or typological nature, and it is therefore largely unpredictable, if not to say whimsical.

In contrast, linguistic typology has also identified quite a few parameters whose settings do seem to fit into more general typological or areal pictures. In such cases, the particular setting on a given parameter can be shown to be correlated to the setting on one or more other parameters, so that an implicational relation between various different linguistic features can be established. To give just one simple example, the following implicational statement, originally formulated in Greenberg (1963), appears to hold between the options that languages have in their encoding of verb placement and adpositional phrases:

- (1) If a language has verb-initial word order, it will have prepositions.

Clearly, by statements such as these the randomness of parameter setting in the encoding of a certain grammatical construction (in this case, adpositional phrases) is constrained to a certain degree: if statement (1) is true – and it seems that it is; no counter-example has been found as yet – we can conclude that, for some reason, at least some languages are ‘forced’ to have prepositions instead of postpositions. Thus, verb-initial word order and prepositional phrases appear to ‘go together’. Another way of formulating this insight is to say that these encoding options form a typological *cluster*. Discovering typological clusters can be seen as the descriptive core business of language typology. In the last fifty years a considerable number of statements such as the one in (1) have been proposed and, in many cases, their validity has been established on the basis of extensive documentation.

When typological parameters form a cluster, it will commonly be the case that one of the parameters can be seen as *primary* or ‘first-order’. That is, the value settings on this parameter do not seem to be determined by anything else, and they therefore represent some sort of ‘basic structural decision’. Again, basic word order can be used as an instance of such primary parameters. As far as we know, there is nothing in the structure of a language which ‘forces’ that language to select verb-initial word order instead of, say, verb-final word order. All we can say is that, from the word order options available, a language has to choose at least some option, but the actual choice which a language (or a language family) makes in this respect is probably random. On the other hand, a typological cluster will also contain one or more *secondary* or ‘second-order’ parameters whose value settings can be said to be determined by the value settings on some other parameter. A case in point here is the value setting on the adposition parameter described above. In constructing adpositional phrases, a language may opt for prepositions or postpositions, but this choice is not completely random. As the

correlation formulated in (1) illustrates, the choice between prepositions and postpositions is restricted to prepositions for those languages which have selected the verb-initial option on the primary parameter of basic word order. In the following sections, we will encounter several other parameter settings which appear to be determined by a 'previous' value setting on a primary parameter. Thus, it turns out, for example, that the choice between verbal or nonverbal encoding of predicative adjectives is not random for a language: instead, this choice is determined by the value setting of that language on the (primary) tensedness parameter (see Section 3).

Since settings on primary parameters are essentially selected at random, it is commonly assumed that primary parameters cannot form clusters among themselves. Thus, if a language has a value setting A on one primary parameter (say, basic word order) and a value setting Z on another primary parameter (for instance, tensedness), the collocation of these two particular value settings is usually rated as a matter of coincidence: the combination of the features A and Z in this language might, in principle, have been otherwise. Now, the point of the present paper is to cast some doubt on this assumption, by demonstrating that the settings on at least a number of primary parameters show mutual restrictions and interdependencies to such a degree that the idea of random value selection on these parameters becomes highly unlikely.

In the following sections, I will discuss five typological parameters which, at the present state of our knowledge, are commonly held to be primary. These parameters will be defined in a binary fashion, so that for each parameter a choice between a "yes" and a "no" option is available. Basing my analysis on a sample of 410 languages, selected from families and areas from all over the globe, I will assign a value setting for the sampled languages on all five parameters, and plot the results for each parameter on a world map. A surprising conclusion that can be drawn from these maps is that these five parameters, which are generally understood to be structurally independent, show a remarkable similarity in the geographical distribution of their positive and negative settings. This result suggests interesting consequences for areal linguistics and linguistic typology alike; further discussion will be presented in the final section of this paper.

2 The basic word order parameter

The first of the parameters to be considered here is in many ways ‘iconic’ to the typological enterprise as a whole: it was the subject of the classic studies by Greenberg (1963, 1966) which are commonly seen as the starting point of modern-day linguistic typology. In these studies, Greenberg and his associates examined possible word order variation in a wide range of constructions and attempted to formulate correlations between the options which languages may choose across these constructions: the above statement (1) is, of course, an example of such a correlation. For the purposes of the current study, I will restrict myself to just a fragment of the domain of word order variation, namely, the options which languages may choose in arranging their basic word order.

Basic word order can be defined as the linear ordering of the main parts of the sentence, viz. the verb (V), the subject (S) and the direct object (O); an additional part of the definition is that basic word order is limited to the ordering of these elements in declarative main clauses. Given the fact that a linear ordering of three elements can, in theory, give rise to six permutations, we conclude that the typology of basic word order will maximally consist of six different types. However, in our study I will reduce this number of options by leaving the position of the subject out of consideration. This decision is motivated by the fact that, as has been argued extensively in the literature that followed Greenberg, subjects have ‘a mind of their own’ when it comes to ordering principles. Most importantly, their behaviour in word ordering appears to be governed largely by their special status as sentential topics, and hence they are subjected to special motivations which do not hold for the other two basic elements. As a result, the parameter that will be considered here deals only with the ordering of the verb and the direct object. In this way, we arrive at a binary parameter with the orderings V-O and O-V as its possible options. It should be remarked that the VO/OV parameter is not an isolated typological distinction: it forms part of a typological cluster in that it can be shown to determine the settings of various other typological parameters. Above, we have seen that VO/OV settings at least partly determine the options in other realms of word order, such as the choice between prepositions and postpositions. Moreover, it has been suggested that the choice between OV order and VO order is an important determinant factor in the choice between prefixation versus suffixation of agreement items on verbs (Siewierska & Bakker 1996).

Both of the options on the VO/OV parameter can be shown to occur as the exclusive choice in at least some of the languages of the sample. Examples of the

V-O option (which covers the basic word order Types SVO, VOS, and VSO) are the following:

- (2) English (Indo-European, West Germanic)

John bought a newspaper.

- (3) Scottish Gaelic (Indo-European, Celtic)

Chunnaic sinn an tarbh.

saw we the bull

'We saw the bull.'

In contrast, Turkish is an example of a language in which the O-V option is mandatory:

- (4) Turkish (Altaic, Turkic)

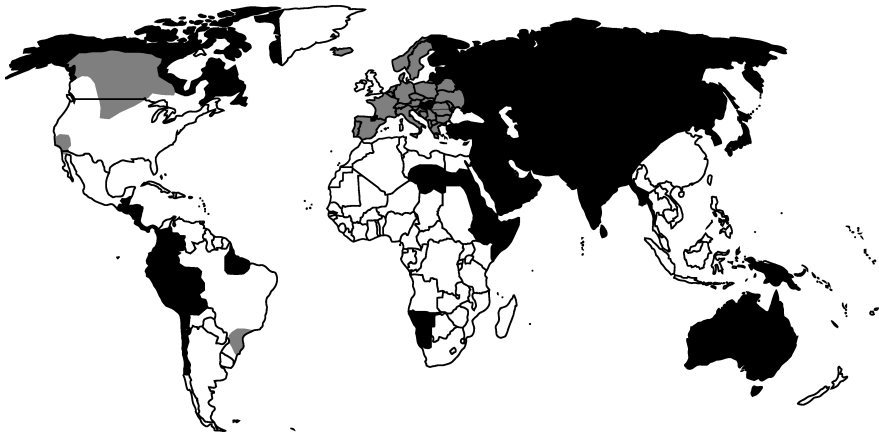
Hasan okü-zü aldi.

H. ox-ACC bought

'Hasan bought an ox.'

English, Scottish Gaelic, and Turkish are clear representatives of their respective types, in that their ordering of verbs and direct objects is rigid: divergence from the norm is virtually impossible, or acceptable only under highly marked circumstances. On the other hand, we find languages in which verb-direct object order (and often, word order in general) can be much less restrictive: Classical Latin, Hungarian, and many of the languages of Australia, are well-known cases in point. For some of these languages, a frequency count in text may help to establish the predominant ordering option, but there are also cases in which one has to concede that, apparently, both options are equally possible. Moreover, a different sort of typological indeterminacy may arise from the fact that languages may have undergone a diachronic change: this is, for example, the common analysis for a number of Western branches of Indo-European, where a drift from OV to VO has been hypothesized.

Map 1 shows the geographical distribution of the two possible parameter settings with respect to verb-object order. On the map, areas marked in black contain those languages in which OV is the only, or the clearly dominant option. Areas in which VO is the only or dominant option are marked in white. Shaded parts of the map indicate either areas in which both parameter settings are possible, or areas in which a drift from OV order to VO order can be argued with some degree of plausibility. Looking at this map, we see that there are at least three large, and



Map 1: Distribution of OV word order

practically uninterrupted, black areas. First, OV ordering is the norm in the mega-area which I will call Eurasia here, and which consists of North and Central Asia (including Japan), India, Iran, and Turkey, and eastern parts of Europe (including the languages of the Caucasus). The Eurasian black area spreads into America on its north-eastern flank, due to the immigration of Asian Eskimo-Aleut speakers into the north of Canada. The area on the western flank of Eurasia, the European peninsula, is mainly shaded, due to a possible diachronic OV-VO drift; exceptions here are Basque (which is clearly OV) and the Celtic languages, which have (and presumably always have had) basic VSO word order.

Apart from Eurasia, other notable 'black' areas are the mega-area which consists of Australia and New Guinea, and an area which covers the southern part of Middle America and the western part of South America (with the Andes mountain range as its eastern border). Smaller OV areas include the territory of the Southern Semitic, Cushitic, and Saharan languages in North-East Africa, the Khoisan languages of South-West Africa, and the Carib languages in the north-western part of South America; these latter languages have the extremely rare OVS pattern as their option for basic word order.

Opposed to these 'black' OV areas, there are also a number of vast areas where VO order reigns supreme. First, we have the 'white' area of East and South-East Asia, which also includes the islands of the Indian and Pacific Ocean. The Middle East and Africa are also predominantly 'white', as is the case with almost the whole of North America (including Mexico and Guatemala), and the centre and

south of South America. With regard to this latter area, the Tupi languages of Brazil and Paraguay need some special mention. Of the nine sampled members of this family, eight clearly have OV order. However, the ninth member, the Paraguayan language Guaraní, has more than ninety-five percent of the total of Tupi speakers, and it definitely has VO order, which may or may not be attributed to influence from Spanish.

3 The tensedness parameter

The second parameter to be examined in this study concerns the notion of tensedness. Since this notion is not a standard one in linguistic typology, some explanation may be in order. The notion was introduced in Wetzter (1996) and Stassen (1997). This latter author provides the following definition:

- (5) Definition of a tensed language (Stassen 1997)

A language is tensed if

- a) predicates in main sentences are obligatorily marked for a past/non-past distinction, and
- b) this distinction is encoded by means of bound verbal morphology.

Thus, in order to be rated as tensed, a language must meet two structural conditions at once. First, it needs to make a systematic and obligatory distinction in its finite verb forms between marking for present (or non-past) time reference and marking for past time reference. (Of course, for one of these time references the marking may be zero; this is often the case for the present tense). Furthermore, this distinctive present-past marking must be effectuated by morphological means, rather than by, for example, adverbs, independent particles, or other non-morphological devices. In other words, a language is tensed if, by looking at the form of a finite verb, one can always decide unequivocally whether this verb form refers to present or past time.

A language which clearly meets the requirements of the definition in (5) is English. Here, we see that simplex finite verb forms come in two paradigms. In one of these paradigms, which is used for present time reference, the verb appears in its unmarked stem form. In the other, the past tense, the verb appears in a form which is morphologically marked, either by a suffix – *ed* (for so-called ‘weak verbs’) or by some internal alteration of the stem (for so-called ‘strong verbs’).

- (6) English (Indo-European, West Germanic)
- a. *John sees the dog.*
 - b. *John saw the dog.*

There is only one way in which a language can be tensed, but non-tensed languages come in a number of different varieties. First, there are languages like Mandarin Chinese, which have no (or hardly any) verbal morphology at all. Secondly, languages like Choctaw that do have distinctive verbal paradigms, but the distinction expressed by these paradigms is aspectual in nature rather than temporal. In a language like Burmese, verbal suffixation is used to distinguish a future form from a non-future form, which, in all probability, represents a modal distinction rather than a temporal one. And finally, a language like Tigak does have an obligatory marking for past versus non-past in all of its declarative main sentences, but this marking does not involve bound verbal morphology: it is effectuated by the use of two different sets of so-called ‘subject pronouns’. As a result, all of these languages fail to meet the conditions stated in (5), albeit for different reasons.

- (7) Mandarin Chinese (Sino-Tibetan, Sinitic)
- Ta pao.*
3SG run
‘He/she runs/ran/will run.’
- (8) Choctaw (Muskogean)
- a. *Pisa - li.*
look.at.PERF - 1SG.ACT
‘I see/saw it.’
 - b. *Pinsa - li.*
look.at.IMPERF - 1SG.ACT
‘I am/was looking at it.’
- (9) Burmese (Sino-Tibetan, Burmese-Lolo)
- a. *Ein pyan -thwa-te.*
home return-go -NONFUT
‘(He) goes/went home.’
 - b. *Li? - me.*
vanish - FUT
‘(I) will vanish.’

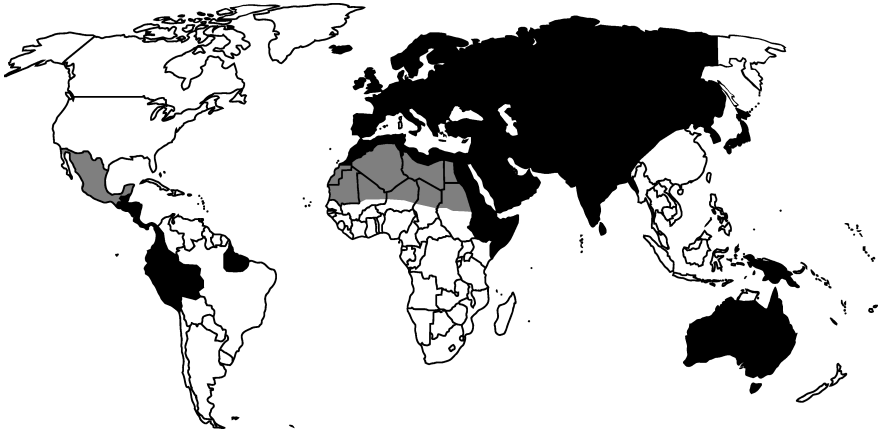
(10) Tigak (Austronesian, Melanesian)

- a. *Gi* *ima.*
3SG.PRES come
'He is coming.'
- b. *Ga* *ima.*
3SG.PAST come
'He came.'

There is evidence to show that tensedness is not an isolated parameter. In particular, Stassen (1997) has shown that it functions as a predictive factor in the cross-linguistic encoding of adjectival predicates: if a language is tensed, its adjectival predicates are almost always encoded in the same way as predicate nominals, whereas in non-tensed languages adjectival predicates are, in the overwhelming majority of cases, treated on a par with verbs.

It should be remarked that the distinction between tensed and non-tensed languages is not a completely discrete one: there can be 'undecided' or 'diffuse' cases. A main source for this diffusion is the fact that aspectual distinctions (mainly, the one between perfective and imperfective aspect) show a tendency of evolving into a temporal past/non-past distinction over time (see Bybee & Dahl 1989), but this diachronic process does not need to have reached its full completion in all languages. In Map 2, areas with languages in which such an 'intermediate' value on the tensedness parameter has been documented are represented as shaded. In contrast, areas with clearly tensed languages are marked in black, whereas areas with clearly non-tensed languages are marked in white.

There are striking similarities between this tensedness map and the map of OV/VO word order. Most importantly, the three black mega-areas on the word order map (viz. Eurasia, New-Guinea-Australia, and Meso-American-Andean) by and large repeat themselves as black areas on the tensedness map, while the major white areas on the Word Order map (viz. East-South East Asia and the Pacific, North America, the heartland of South America, and sub-Saharan Africa) turn up as white areas on the tensedness map as well. There are a number of discrepancies between the two maps (Hebrew and Arabic, two Semitic languages of the Middle East and Northern Africa are VO, but tensed; the languages of North-East Siberia, and the Eskimo-Aleut languages of North-America, are OV, but non-tensed, as are the Khoisan languages of South-West Africa), but it can be seen that these 'conflicting' areas are typically situated on what might be called 'fault lines', that



Map 2: Tensed vs. non-tensed languages

is, places where major black and white areas meet and ‘bump into each other’, so to speak.

4 The casedness parameter

Like tensedness, casedness is not a generally employed notion in typological linguistics. It has been devised specifically for this study, and it can be defined as follows:

(11) Definition of a cased language

A language is cased, if it has morphological (dependent) marking to indicate the difference between subjects and direct objects, at least for pronouns.

Casedness thus represents a specific strategy that languages may employ to keep the two core grammatical functions (Subject vs. Direct Object, or Agent vs. Patient) apart and identifiable. As is well known, the differentiation between these two functions can be effectuated by a number of different means. Some languages, such as Bahasa Indonesia, use fixed word order to this effect, as is shown in example (12):

(12) Bahasa Indonesia (Austronesian, West Indonesian)

a. *Saya memeluk dia.*

1SG embrace 3SG

‘I embrace(d) him/her.’

- b. *Dia memeluk saya.*
3SG embrace 1SG
'He/she embrace(d) me.'

Other languages use a 'head marking' strategy (Nichols 1992), in which core functions are made identifiable by means of a system of agreement affixes on the verb. Casedness is the opposite of this head marking strategy: here, the function of a core argument can be identified in isolation, by looking at the specific form which the two relevant argument NP's have. In languages which are 'cased' in this sense, it frequently happens that one of the core arguments (typically, the subject) remains formally unmarked, whereas the other has overt marking. The specific way of marking may take different formal forms: some languages have an affixational case system, whereas others use structurally independent case markers or adpositions to achieve the discriminatory effect. Examples are:

- (13) German (Indo-European, West Germanic)

Der Mann sah den Hund.
ART.MASC.SG.NOM man saw ART.MASC.ACC dog
'The man saw the dog.'

- (14) Japanese (Altaic, Japanese)

Taroo ga tegami o katta.
T. SUBJ letter OBJ wrote
'Taroo wrote a letter.'

A special feature of definition (11) is that dependent case marking on pronouns is specified as the minimal requirement for casedness. The reason for this is that dependent case marking systems show a cross-linguistic tendency to 'wear off' over time: an example of this development can be found in various Germanic languages such as English, Dutch, or Swedish, where an erstwhile case marking on noun phrases has gradually vanished. Now, it turns out that the abandonment of core case marking takes place earlier and more radical with full nominal arguments than with pronominal arguments: the above-mentioned Germanic languages no longer differentiate full lexical subjects and direct objects by case marking, but when the subjects and/or direct objects are pronominal they still do.

- (15) Dutch (Indo-European, West Germanic)

a. *Het meisje zag de hond. / De hond zag het meisje.*
the girl saw the dog the dog saw the girl
'The girl saw the dog. / The dog saw the girl.'

- b. *Het meisje zag mij.* / *Ik zag het meisje.*
 the girl saw 1SG.ACC 1SG.NOM saw the girl
 'The girl saw me. / I saw the girl.'

Pronouns thus appear to be the 'nec plus ultra' of dependent case marking. If they are not marked for grammatical function (as is the case in Bahasa Indonesia; see example (12) above), the language will be rated accordingly as non-cased.

As far as I am aware, the casedness parameter cannot yet be brought into connection with other typological parameters. That is, I do not know of any typological correlations in which the cased or non-cased status of a language predicts anything else, and neither do I know of any correlation in which the cased or non-cased status of a language is predicted by anything. It has sometimes been suggested that the presence of dependent case marking makes it easier for a language to have relatively free word order, and that, conversely, absence of such marking will lead to stricter and more rigid sentential word order. It remains to be seen, however, whether this idea, attractive though it may sound initially, will stand the test of thorough typological examination. Map 3 pictures the distribution of cased and non-cased languages in my sample. Cased areas are marked in black, and non-cased areas are marked in white.



Map 3: Cased vs. non-cased languages

As was the case with the two previous maps, we see that the geographical distribution of the settings on the casedness parameter corresponds to a patterning in mega-areas, and that this patterning is conspicuously similar to the one depicted in Map 1 and Map 2. Again, we find a split between the same ma-

major 'black' areas (Eurasia and North Africa, New-Guinea-Australia, and Meso-American-Andean) and major 'white' areas (East and South-Asia, North America, sub-Saharan Africa and Central South America). Areas that are 'white' on the tensedness map but black on the casedness map cover the Khoisan languages of South-West Africa, and the North-East Siberian languages (including Eskimo-Aleut). A change from black to white can be seen in the Carib languages of North-East South America: these languages are OV and tensed, but they do not have a dependent case system.

5 The AND-WITH parameter

The AND-WITH parameter, which was proposed in Stassen (2000), has to do with the cross-linguistic variation in the encoding of noun phrase conjunctions. It entails a distinction between AND languages and WITH languages, and can be defined as follows:

- (16) Definition of AND languages and WITH languages (Stassen 2000)
- a. A language is an AND language if
there is a structural or lexical distinction between the encoding of noun phrase conjunctions and the encoding of the comitative case.
 - b. If there is no such distinction, the language is a WITH language.

An obvious example of an AND language is English. In order to express a situation in which two participants are involved together in one action, this language has the choice of using either one of the following constructions. In the first, the two participants, encoded here as full noun phrases, are constructed on the same structural rank. Thus, they form – in this language at least – constituents of a conjoined noun phrase and are – again, in this particular language – connected by the conjunctive particle *and*. In the second construction, there is no equality of structural rank between the two noun phrases: while one of these noun phrases is a core argument (in this case, the subject), the other noun phrase is constructed as part of an adverbial phrase, marked by the comitative preposition *with*.

- (17) English (Indo-European, West Germanic)
- a. *John and Mary went to see a movie.*
 - b. *John went to see a movie with Mary.*

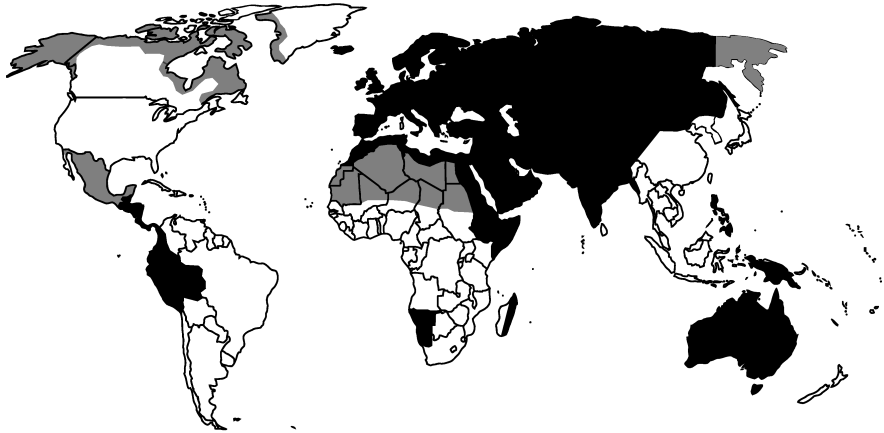
In contrast to AND languages such as English, there are languages in which only one option is available. To be specific, such languages have only the second option – that is, the option in which the two noun phrases are not of equal rank – at their disposal; an equivalent way of stating this situation is that such languages lack the option of conjoined noun phrases. Examples of languages which are to be rated as WITH languages on this criterion are:

- (18) Samoan (Austronesian, Polynesian)
Ua sau Paulo ma Maria.
 PROG come P. and/with M.
 ‘Paulo and Maria are coming / Paulo is coming with Maria.’
- (19) Akan (Niger-Kordofanian, Kwa)
Kwasi nye Amba a-ba.
 K. and/with A. PAST-come
 ‘Kwasi and Amba have come/Kwasi has come with Amba.’

Like the casedness parameter, the AND-WITH parameter is not known to participate in any established typological clustering. Moreover, just like the other three parameters discussed so far, settings on the AND-WITH parameter are not necessarily discrete: ‘intermediate’ values can be observed for quite a few languages in the sample. The motivation behind this indeterminacy is, again, mostly of a diachronic nature. Stassen (2000) discusses a number of cases in which an erstwhile WITH language gradually reanalyses its WITH strategy into something that resembles a ‘true’ conjunctive construction to a lesser or greater degree. In such languages, an additional *and* structure can be seen to arise, although typically the same conjunctive item will continue to be used in both constructions. Furthermore, several WITH languages seem to have borrowed an additional AND construction from neighbouring dominant languages. This is apparently the case in a number of languages from Siberia, which have added the Russian conjunction *i* ‘and’ to their repertoire of conjunctive strategies.

Map 4 shows the AND languages in my sample marked in black. Areas that contain WITH languages are marked in white, and intermediate cases are shaded.

I trust that, by now, the stratification of this map will look familiar to the reader. We see the same major ‘black’ and ‘white’ areas here as we have seen on the three previous maps, and what is more, we also note the same ‘swing’ areas. North-East Siberia and Eskimo-Aleut are shaded on this map, for reasons that were exposed above. Khoisan is back as a black area here, but Carib and Tupi



Map 4: AND vs. WITH languages

in South America are white, as they are definitely WITH languages. Newcomers to the ‘black’ areas are to be found among Austronesian languages. Malagasy (the westernmost Austronesian language, spoken in Madagascar) is an AND language, as are the languages of the Philippines: all these languages are ‘white’ on the three previous maps. Conversely, Japanese and Korean are ‘white’ on this map, whereas they are ‘black’ on all other maps presented thus far.

6 The absolute parameter

The final parameter to be discussed in this study can be called the absolute parameter. This parameter has to do with the morphosyntactic variation which languages exhibit in the encoding of clause linkage. This is a structural domain in which several parameters interact. For our present purpose we do not have to consider all these parameters, and we can restrict ourselves to a somewhat simplified picture, in which we will take into account only those constructions in which the linked clauses at issue have different subjects.

If two such clauses are linked in, for example, a sequence that expresses the simultaneous occurrence of two events, a distinction can be made between two different strategies of encoding. In the first strategy, the predicates in the two clauses both have the form that predicates in main clauses have, the so-called ‘finite form’. English examples of sequences in which this situation holds are the following:

- (20) English (Indo-European, West Germanic)
- a. *Mary sang and John played the piano.*
 - b. *While Mary sang, John played the piano.*

In the English constructions in (20), which represent cases of clause linkage between two clauses that have different subjects, we can see that the predicates in both clauses retain the form of a main clause predicate. Since Stassen (1985) the term ‘balancing construction’ has come into use for clause linkage constructions of this kind. This term is meant to reflect the fact that both predicates in the construction have main predicate form and thus ‘balance’ one another in terms of their structural rank within their clause.

Unlike encoding in a balancing construction, clause linkage may also be achieved by a strategy that was labelled ‘deranking’ in Stassen (1985). Under this strategy, one of the predicates in the sequence keeps its main predicate form, but the predicate in the other clause takes a non-finite form which is typically reduced in its verbal categories when compared to finite verb forms. Such ‘deranked’ predicates take a number of different morphosyntactic shapes across languages and the terminology to refer to them has not been standardized; we find labels like ‘participle’, ‘gerund’, ‘infinitive’, ‘action nominal’, ‘converb’, and several others in the literature. For our purposes, the cross-linguistic variation in the morphosyntactic encoding of deranked predicates need not detain us. What is important for us here is the fact that all deranked predicates, irrespective of their actual morphosyntactic make-up, have a form which cannot be used for a predicate in a main clause.

In addition to its balancing option, English also has the option of deranking for linked clauses, but this option is restricted to sequences in which the two clauses have the same subject. This is demonstrated by the examples in (21). We see that in sentence (21a), where the two clauses have the same subject, it is possible to derank one of the predicates by means of a non-finite verb form called the present participle. If, however, the two clauses have different subjects, deranking of one of the predicates leads to ungrammaticality (see example (21b)).

- (21) English (Indo-European, West Germanic)
- a. *Mary was up on the stage, playing a violin.*
 - b. **Mary was up on the stage, John playing a violin.*

In traditional grammar, deranked sequences with different subjects are known as ‘absolute constructions’. Sentence (21b) demonstrates that English is a language

that lacks the possibility to form constructions of this kind. On the other hand, however, quite a few languages in my sample actually allow such absolute constructions, be it as the only option for different-subject sequences, or as one of the options in addition to a balancing construction. Examples of languages in which absolute constructions are particularly frequent are Classical Latin and Finnish.

- (22) Classical Latin (Indo-European, Italic)

Serva cantante dominus
slave.girl.ABL.SG sing.PCP.PRES.ABL.SG master.NOM.SG.
bibit.
drink.PRES.IND.3SG
'The master drinks and/while the slave girl sings'

- (23) Finnish (Uralic, Fenno-Ugric)

Kalle-n tu -le -ssa Pekka lahti.
K.-GEN come -INF -INESS P.-NOM leave.PRES.IND.3SG
'When Kalle came, Pekka left.' (lit. 'In Kalle's coming Pekka left.')

On the basis of this contrast between languages like English, on the one hand, and languages like Classical Latin and Finnish, on the other, the concepts of absolute and non-absolute languages can now be defined as follows:

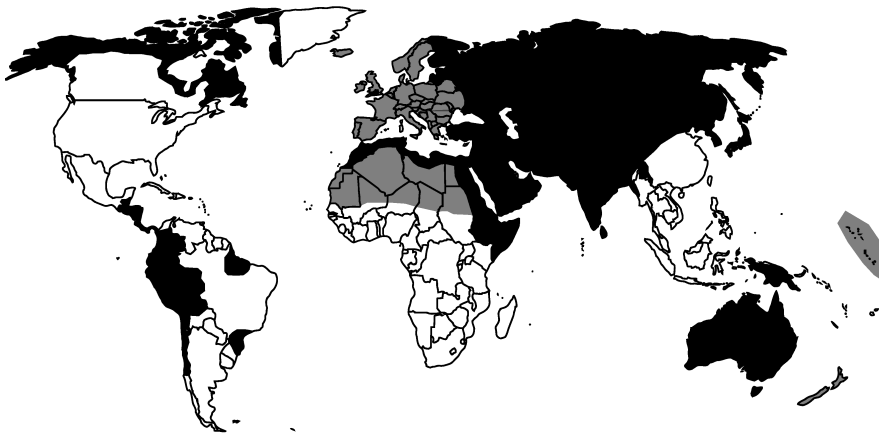
- (24) Definition of absolute and non-absolute languages

- a. A language is an absolute language if, in a sequence of two clauses with different subjects, one of the predicates can take a deranked form.
- b. A language is non-absolute if deranking in a sequence of two clauses with different subjects is not possible.

There is ample evidence to suggest that the absolute parameter, and clause linkage encoding in general, constitutes an important primary parameter. Options in clause linkage encoding play a determinant role in the typologies of other construction types, such as the encoding of comparative constructions (Stassen 1985), the encoding of predicative possession constructions (Stassen 2009), and the encoding of various constructions of secondary predication, such as the formation of manner adverbials (Loeb-Diehl 2006) and resultatives (Verkerk 2009).

Map 5 documents the geographical distribution of absolute and non-absolute languages. Here, a by now familiar caveat must be repeated. Assigning a value on the absolute parameter can be problematic for some languages, and again, this indeterminacy is mainly due to diachronic developments. In the Indo-European

languages of Europe in particular one can notice a gradual demise of the absolute construction in favour of balancing encodings, to the effect that absolute formations, if they are still in use at all, are seen as ‘old-fashioned’, ‘bookish’, or ‘formal’ in comparison to their balancing counterparts. Languages in which this diachronic drift towards balancing clause linkage can be made plausible are represented by shading in Map 5. For the clear cases, we will use the same colouring as on the other maps. Areas with a positive value on the absolute parameter (that is, areas with absolute languages) are marked in black, whereas areas with a negative value on the parameter (that is, areas with non-absolute languages) are marked in white.



Map 5: Absolute vs. non-absolute languages

As is clear from Map 5, the geographical distribution of the ‘black’ and ‘white’ values on the absolute parameter does not differ greatly from the distributions that are depicted on the other maps. Eurasia (including Japanese, Korean, North-East Siberia and Eskimo-Aleut, the Middle East, and North-East Africa) is again a black area here, with the exception of Europe, which, as we have seen above, is a shaded area on this parameter. New-Guinea-Australia and Meso-America-Andean are again the other two steadfast black areas. The same consistency can, by and large, be attested for the major ‘white’ areas on this map. North America, the heartland of South America, sub-Saharan Africa, and East and South-East Asia all contain clearly non-absolute languages, the only exception being the Polynesian languages, which happen to allow for absolute constructions and therefore show up on this map as a ‘black’ area for the first time. As for the ‘swing areas’,

we can note that the South-American Carib and Tupi languages turn up marked in black on this map. The South-West African Khoisan languages, on the other hand, do not tolerate absolute constructions and are therefore represented as a 'white' area here.

7 Discussion

As I see it, the results of the cross-linguistic investigation of the five parameters in the previous sections, and the maps that are based upon these results, raise some intriguing questions for both linguistic typology and areal linguistics. Starting with the typological side of things, we can conclude that the combination of value settings on these five parameters is almost certainly not random, even though all the parameters considered are, to the best of our knowledge, 'primary', and hence structurally independent of one another. However, if languages were free to select their value settings on these five parameters, the predicted number of different language types would be $(2 * 2 * 2 * 2 * 2 =) 32$. Now, we can see that this number of logically possible language types is severely restricted empirically. In fact, it appears that there is a strong tendency towards a dichotomy into 'mega-types', in which languages tend to align themselves into two sides which, on each parameter, have opposite settings. In keeping with the terminology used in the previous sections, I will call these two language types 'black languages' and 'white languages'. For these two types, the following clustering of parameter settings can be observed:

(25) Value settings for black and white languages

	White Languages	Black Languages
Word Order	VO	OV
Tensedness	Non-tensed	Tensed
Casedness	Non-cased	Cased
Nominal Conjunction	WITH	AND
Absolute Converb	Non-absolute	Absolute

What these results suggest is that the typological variation between languages may be far more restricted than has been assumed so far. It may be the case that languages align themselves in 'optimal' collocations of settings on a rather restricted set of 'primary' parameters, which largely determine the type to which the language belongs. In fact, based on the results of this study one might even

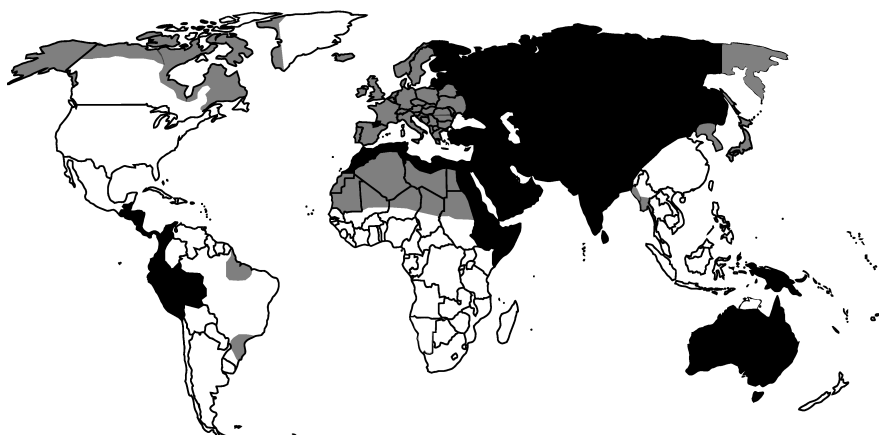
venture the – admittedly, totally wild – hypothesis that there are only two basic language types in the world, which are characterized by taking opposite choices on a number of fundamental structural decisions. At the same time, however, it should be conceded right away that the results obtained in this study raise various questions of their own, both of a descriptive and an explanatory nature. First, we can ask whether the five parameters that cluster in the way which is presented in (25) are all truly independent of one another: it may very well be that the empirically established cluster of these five parameters can be shown to have (some degree of) internal structure after all, so that some of these parameters are actually secondary. Furthermore, one can ask why it is just this set of parameters which gives rise to the differentiation into two structurally opposite language types. It should be kept in mind that the parameters discussed in this study were selected largely on the basis of my personal typological domains of interest, and that there is absolutely no guarantee that the list of these clustering parameters is complete; in fact, the opposite seems far more likely. In other words, the results obtained in this study can only be considered as a starting point for much more thorough further research, and the best one can say at the moment is that these results are intriguing and potentially promising.

Turning now to issues of areal linguistics, we can conclude that the results of this study have potential consequences for our conception of language areas as well. In the previous sections we have seen that the typological distinction between black and white languages tends to converge with the definition of black and white ‘mega-areas’. In Map 6, I have summarized the five maps for the separate parameters, according to the following format:

- If a language appears 4 or 5 times on a map as a black language, it will be marked as a black language on Map 6.
- If a language appears 0 or 1 times on a map as a black language, it will be marked as a white language on Map 6.
- If a language appears 2 or 3 times on a map as a black language, it will appear as shaded on Map 6.

As a result, the map of black and white language distribution looks like this:

As we have already noticed in our discussions of the various parameter maps, the distribution of black and white languages across the globe gives rise to the identification of a number of clear black and white mega-areas, which form relatively homogenous, uninterrupted stretches. The largest black area is Eurasia (or ‘The Old World’), which, in its maximal extension, covers Central and Northern



Map 6: Black vs. white languages

Asia, India, Pakistan, The Middle East, Northern Africa, and North-East Africa. Further mega-areas which have a consistent (or almost consistent) black encoding are New Guinea/Australia and the Meso-American-Andean area, which covers the south of Central America and the north-west of South America. In opposition, consistently white areas are found in North America/Mexico, most of South America, sub-Saharan Africa, China and South East Asia, and the Indic and Pacific Ocean. A further remarkable feature of Map 6 is that the shaded 'in-between areas' or 'swing areas' are commonly found at the edges of the Eurasian black area. First, there is Europe, which constitutes the far-western tip of the Eurasian land mass. (The exception here is the isolated Basque language, which is a steadfast black language). At the south-western border of the Eurasian mega-area we find the intermediate, 'shaded' area covered by the North-African Berber languages. Moreover, at the very north-eastern fringe of Eurasia, there are the so-called Paleo-Siberian languages, Japanese, Korean and Eskimo-Aleut which constitute a transitory area between 'black' and 'white' encoding. 'Swing areas' that are not situated at the borderline between black and white mega-areas do occur, but they are scarce and tend to consist of one language family only: in South America we can point to the Tupi languages and the Carib languages, and in South-West Africa we have the relatively small Khoisan family.

What this map suggests is that geographical contact may have been much more extensive than has been assumed up to now. The possibility of linguistic 'macro-areas', which supersede current genetic classification, should be seriously consid-

ered, and the time-depth for presumed geographical contact should perhaps be extended much further than is commonly accepted now in areal linguistics. Alternatively, one might reconsider the possibility that there exist ‘mega-families’ of the type proposed by Joseph Greenberg in various publications. Thus, Map 6 may provide some renewed credibility for concepts like ‘Amerind’ and ‘Nostratic’.

Whatever one may think about these suggestions, I feel it is safe to draw at least one general methodological conclusion from the data presented in this study.

In the generally accepted view, typological linguistics and areal linguistics are seen as two separate enterprises, and one should take care not to confound them. Although I agree that, in principle, this is sound practice, there are nonetheless indications that typological collocations and areal configurations of linguistic parameters have a tendency to converge, especially when macro-areas are considered. Therefore, the two linguistic sub-disciplines of language typology and areal stratification may be beneficial to one another to a degree that is much higher than is usually thought possible.

List of abbreviations

In the glosses of the example sentences the following abbreviations have been used:

1,2,3	first, second, third person
ABL	ablative case
ACC	accusative case
ACT	actor case
ART	article
FUT	future tense
GEN	genitive case
IMPERF	imperfective aspect
IND	indicative mood
INESS	inessive case
INF	infinitive
MASC	masculine gender
NOM	nominative case
NONFUT	non-future tense
PAST	past tense

PCP	participle
PERF	perfective aspect
PRES	present tense
PROG	progressive aspect

Bibliography

- Bybee, J. & Ö. Dahl. 1989. The creation of tense and aspect systems in the languages of the world. *Studies in Language* 13–1. 51–103.
- Greenberg, J. H. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In J. H. Greenberg (ed.), *Universals of language*, 73–113. Cambridge: MIT Press.
- Greenberg, J. H. 1966. *Language universals*. The Hague: Mouton.
- Loeb-Diehl, F. 2006. *The typology of manner expressions*. Nijmegen: Radboud University dissertation.
- Nichols, J. 1992. *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Siewierska, A. & D. Bakker. 1996. The distribution of subject and object agreement and word order type. *Studies in Language* 20.1. 115–161.
- Stassen, L. 1985. *Comparison and Universal Grammar*. Oxford: Blackwell.
- Stassen, L. 1997. *Intransitive Predication*. Oxford: Oxford University Press.
- Stassen, L. 2000. AND-languages and WITH-languages. *Linguistic Typology* 4. 1–54.
- Stassen, L. 2009. *Predicative possession*. Oxford: Oxford University Press.
- Verkerk, A. 2009. *Secondary predication in a typological context*. Nijmegen Radboud University MA thesis.
- Wetzer, H. 1996. *Nouniness and verbiness: a typological study of adjectival predication*. Berlin: Mouton De Gruyter.

Author

Leon Stassen
Faculteit der Letteren, Taalwetenschap
Radboud Universiteit Nijmegen
l.stassen@let.ru.nl