Contact-induced change in Dolgan : an investigation into the role of linguistic data for the reconstruction of a people's (pre)history

Stapert, E.L.

CHAPTER 7 WORD ORDER

7.1. INTRODUCTION

An additional domain in which Dolgan differs from Sakha is word order. Compared to Sakha, Dolgan shows greater variation in the arrangement of constituents in transitive sentences, allowing more freely for orders other than the standard Turkic (and Sakha) SOV, in particular for SVO. This is not to say that Turkic languages show no variation at all and that Dolgan is the exception within the language family. Many Turkic languages do allow for variation, usually associated with particular discourse pragmatic functions such as topicalisation. To give an example, regardless of the fact that SOV is its pragmatically least marked order, some scholars claim that Turkish is essentially a language with free word order (Kornfilt 1997: 91). In addition to language-internal reasons for non-SOV constituent order, many Turkic languages spoken in the vicinity of languages belonging to other families (such as Slavic), have acquired greater flexibility due to contact with their neighbours. For example, word order in Khakas has become more flexible under the influence of Russian (Anderson 1998: 71), and in the West Rumelian dialects of Turkish spoken in Macedonia, SVO has become the unmarked word order under the influence of Macedonian (Friedman 2003: 66).

Rather than treating word order change by itself as an exotic phenomenon, the issues of interest for the current study are a) an investigation of the difference

in word order variation between Dolgan and Sakha; and b) how this difference can be explained. First I will show on the basis of quantitative analysis of word order patterns that the higher degree of flexibility in Dolgan is very unlikely to be due to chance. It cannot be attributed to the idiolect of certain individuals, nor does it correlate with a certain text genre or age category. I take this as evidence that this tendency is pervasive throughout the entire language, and that the present variation could eventually become established as a change.

After a review of some of the main ideas on word order change in the literature, I will argue that the variation in Dolgan word order is the result of Russian influence brought into the language by bilingual speakers of Dolgan. To substantiate this claim the sociolinguistic situation in the different Dolgan communities will be discussed and I will postulate that while this change is ongoing, the change is best explained in terms of two underlying processes of contact-induced change, depending on the linguistic dominance of the speaker: imposition in Dolgan people whose dominant language is Russian (i.e. typically the younger generations), and borrowing in those people whose dominant language is Dolgan, but whose way of speaking is influenced by the constant exposure to Russian.

7.2. WORD ORDER IN TURKIC LANGUAGES

Since the structure of Turkic languages is predominantly head-final, it follows that the unmarked word order in most languages is SOV. Within this statement, O needs to be understood as any kind of object, and V as any kind of predicate rather than only as a direct object and a verb, for which these abbreviations are normally used. This applies to finite (7.1, 7.2, 7.3) as well as to non-finite (7.4) clauses and is observed particularly strictly in the latter category (Johanson 1998: 57). In the following examples the object is marked in bold so the different clause orders can be spotted more easily:

TURKISH

(7.1) *Hasan* ***kitab-ı*** *oku-du*
Hasan book-ACC read-PST
'Hasan read the book.' (Kornfilt 1997: 89)

KHAKAS

(7.2) *min* *taŋda* *paba-zɨna* ***pu*** ***kniga-nɨ*** *pir-e-m*
I tomorrow father-3.DAT this book-ACC give-FUT-1
'Tomorrow I will give this book to his father.'
(Anderson 1998:72, morpheme breaks mine)

UYGHUR

(7.3) *saen* ***suet*** *ich-t-ing*
You milk drank-PST-2.SG
'You drank milk.' (De Jong 2007: 101, glossing mine)

UZBEK

(7.4) *Åybek-niŋ* *bu* ***kitåb-ni*** *yåz-γåni-ni* *bilå-mån*
Aybek-GEN this book-ACC write-CV-ACC know-1SG
'I know that Aybek has written this book.'
(Johanson 1998: 60, glossing mine)

However, in most Turkic languages variation in word order is not uncommon. Typically, a non-standard arrangement of constituents correlates with certain discourse-pragmatic functions. Constituents in sentence initial position normally have the interpretation of topic, whereas the focused element is found directly before the predicate (Johanson 1998: 58-59). In a pragmatically unmarked sentence these positions roughly correlate with the grammatical functions of subject and object, but this pattern may be reversed when other constituents are assigned the function of topic or focus, as in example 7.5. In this example, the sentence-initial position is occupied by the object *ıstakozu* 'lobster' instead of the subject *Hasan*, because the lobster is the topic of this sentence.

TURKISH

(7.5) ***ıstakoz-u*** *Hasan* *Ali-ye* *ver-di*
lobster-ACC Hasan Ali-DAT give-PST
'(Speaking of) the lobster, Hasan gave (it) to Ali.' (Kornfilt 1997: 200)

An additional 'postpredicative position' in Turkic languages, which is not included in the description of unmarked SOV sentences, is reserved for information that is not new, such as already activated topics, defocused constituents, or afterthoughts (Johanson 1998: 58). This is illustrated with an example from Turkish in 7.6, where

the sentence-final subject 'Hasan' represents shared background information. Kornfilt explicitly says that in this language the constituent in post-predicative position does not represent afterthoughts, but rather encodes shared knowledge or 'backgrounding' (Kornfilt 1997: 206), which is only compatible with the first two functions (topic and defocused constituent) described for Turkic by Johanson. However, the descriptions for Turkic and Turkish have enough in common to illustrate the function of postverbal slot with a sentence from Turkish.

TURKISH

(7.6) *Ali-ye* *kitab-ı* *ver-di* ***Hasan***
Ali-DAT book-ACC give-PST Hasan
'He gave the book to Ali, Hasan.' (Kornfilt 1997: 206)

7.3. WORD ORDER IN SAKHA

In Sakha standard word order and its possible variants closely resemble the general Turkic pattern described above. Sakha typically employs the standard Turkic SOV order for unmarked transitive clauses (Stachowski & Menz 1998), but very often only O and V are overtly expressed due to the fact that Sakha is a pro-drop language, as can be seen from example 7.7. Full SOV sentences, in which all three core constituents are overtly expressed, are in fact very rare in spontaneous narratives (only 0.8% of all counted transitive clauses in the corpus).

SAKHA

(7.7) *Bu* *Uolba* *hir-itten* ***sü:rbe*** ***toγus*** ***kihi-ni***
This Uolba place-ABL.3SG twenty nine person-ACC
ilʤi-bit-tere, *bu* *kïrakïj* *baγajï* *deriebine-tten*
take.away-PST.PTC-POSS.3PL this tiny INTNS village-ABL
'From Uolba they took twenty nine people, from this very tiny village.'
(ARR: 022)

This sentence seems to be neutral with regard to the relation between the object *sü:rbe toγus kihini* 'twenty nine people' and the verb *ilʤi-bit-tere* 'they took', while the topic *Uolba* is placed in sentence initial position, and further specified as *bu kïrakïj baγajï deriebinetten* 'from this little village' as an elaboration and afterthought.

As in other Turkic languages, deviation from this basic word order pattern occurs for discourse-pragmatic reasons such as topicalisation, in which case the topic is fronted to clause initial position (Stachowski & Menz 1998). This is in agreement with data from the spoken corpus of Sakha. Within a set of 176 transitive clauses with overt expression of O and V (for details see Section 7.5.1) only 3 VO clauses are found (1.7%), reflecting the dominance of the OV pattern convincingly. In addition, 2 instances of OSV order are attested. All sentences with non-SOV word order have clear pragmatically marked connotations. Evidence for this is most clearly seen, or rather heard, in intonation patterns. In unmarked statements, sentence stress in Sakha normally comes on the final constituent. Since this is typically the verb, as a consequence of Sakha's SOV word order, in the average statement verbs are lightly stressed. Despite this being the unmarked prosodic pattern, sentences in which the final verb is in focus are still clearly distinguishable. In these cases the stress on the final verb is noticeably increased, and in addition the object can be moved to sentence-initial position to underline its topicality. This is exemplified in example 7.8, which displays OSV order. The sound recording reveals an unmistakable increase in stress on the verb *körbütüm* 'I have seen', showing that the focus of the sentence is the act of 'seeing', whereas the 'husband' in clause initial position fulfills the function of topic. This is further supported by the discourse context in which this sentence was produced. It is a story about a wedding, in which the participation of a wife (who is also the narrator) and a husband is typically presupposed. A third participant, who was ill and could therefore not come to meet the husband at the wedding, then said that he would not be able to meet the husband now, but that he has *seen* that husband before.

SAKHA

(7.8)	***En***	***kergeŋ-ŋin***	*min*	*kör-büt-üm*	*di:r.*
	2SG	spouse-ACC.2SG	1SG	see-PST.PTC-POSS.1SG	say.PRS.PTC
	'I've *seen* your husband, he said.'				(ARR: 273)

In example 7.9, clause order is VO, and the sound recording shows a clear break between the verb *emti:r* 'he treats' and the object *tugu barïtïn*, suggesting that the object 'what, everything' is produced as an afterthought.

SAKHA

(7.9) *Em-ti:r* ***tugu*** ***barï-tïn,*** *telepatija,*
medicine-VBLZR.PRSPT what.ACC all-ACC.3SG telepathy
vse takoe.
all.R such.R
'He treats what, everything, telepathy and all that.' (ARR: 256)

VO sentences for which the discourse-pragmatic function is undisputedly unmarked are not found in Sakha. From these data we must conclude that the unmarked word order pattern in Sakha is in line with the Turkic languages in general: transitive clauses are rather strictly SOV, and exceptions to this pattern occur only for particular discourse-pragmatic reasons.

7.4. WORD ORDER IN DOLGAN

In Dolgan, unmarked word order is also predominantly SOV, as is illustrated in examples 7.10 and 7.11. These sentences are neutral descriptions of what is usually done in preparation for migration (7.10) and how reindeer hides are prepared (7.11). They have a neutral intonation pattern in which none of the arguments is particularly stressed except for the light clause-final stress that, as in Sakha, characterises the unmarked prosodic pattern. Examples 7.10 and 7.11 also illustrate that Dolgan, like Sakha, is a pro-drop language, in which the S is frequently not overtly expressed within the clause.

DOLGAN

(7.10) ***taba*** *tut-a-bït* *buo* *oччoγo* *buollaγïna*
reindeer hold-SIM.CV-1PL PRT then PRT
boloχ-putugar ***aγïs*** ***taba-nï*** *kölüj-e-bit*
balok-DAT.1PL eight reindeer-ACC harness-SIM.CV-1PL
'We catch reindeer, and then for our balok, we harness eight reindeer.'
(IMA: 10)

(7.11) *taha:ra* *giniler-iŋ* *maŋnaj* *iti* ***tiri:*** ***üle-tin***
outside 3.PL-POSS.2SG first this skin work-ACC.3SG
üle-li:-ler
work–VBLZR.SIM.CV-PRED.3PL
'Outside they first do the work with the skin.' (ESB: 04)

However, data from the spoken corpus show that in Dolgan there is greater acceptance than in Sakha of word orders that differ from this standard constellation, in particular an acceptance of SVO. Importantly, the post-verbal object does not necessarily encode an already activated topic, defocused constituent, or afterthought, as was described for other Turkic languages, but can also occur in pragmatically neutral utterances. To substantiate this statement, in Dolgan 41 out of 175 sentences with overtly expressed verb and object are VO. These 23.4% contrast sharply with the 1.7% of VO -sentences just mentioned for Sakha.

Examples 7.12 and 7.13 are clear instances of objects in the position of an afterthought. In 7.12 the postposed *iǹe-ŋ haŋa-tïn* 'your mother's word' is the object of *iste-gin* 'you listen', and in 7.13 *ol tiri:-gin* 'that skin' is the object of *ïj-ï:l-lar* 'they hang'. This interpretation is corroborated by the fact that in both cases the verb is followed by the particle *buo*, which occurs at the end of a clause and is always followed by a pause. It has some kind of assertive meaning, displays a drop in intonation and turns the preceding clause into a closed unit. Everything following this particle is a new sentence, or an afterthought.

DOLGAN

(7.12) *iste-gin* *buo,* ***iǹe-ŋ*** ***haŋa-tïn***
listen.SIM.CV-PRED.2SG PRT mother-POSS.2SG language -ACC.3SG
'You listen to your mother's word.' (ESB: 42)

(7.13) *iti...* *kimieχe* *taŋas* *ïja:n-ar* *kim-ner-ge*
this... who.DAT clothes be.hung-PRS.PTC who-PRED.3PL-DAT
ïj-ï:l-lar *buo* ***ol*** ***tiri:-gin,*** *taŋas*
hang-PRS.PTC-PRED.3PL PRT that skin-ACC.2SG clothes
ïja:n-ar.
be.hung -PRS.PTC
'Ehm... they hang it on a clothes hanger, the skin, a clothes hanger.' (ESB: 34)

On the other hand there are examples like 7.14 and 7.15, in which there is no indication that the object is separated from the verb in any sense, even though the object occurs in clause final position. In 7.14, the post-verbal object occurs in the combination *kötöχtö ginini* 'he lifted him', which is a clear syntactic and intonational unit. An interpretation of *ginini* as afterthought seems, in the absence of any semantic, syntactic or intonational cues, very unnatural. Rather, 7.14 is a semantically and pragmatically unmarked description of this lifting event, and the changed word order does not affect the interpretation, i.e. this sentence would have exactly the same reading as in a sentence where the order is *ginini kötöχtö*. The same holds for 7.15, where *ontugun* 'that' follows the verb *tutuoχta:χχïn* 'you should hold'.

DOLGAN

(7.14) *hïnńan-an* *χanń-an* *bar-an* *ke* *de* *ol*
relax-SQ.CV and.so.on-SQ.CV go-SQ.CV CONTR PRT PRT
kötöχ-tö ***gini-ni,*** *kötöχ-tüler* *krïltso-χa:m-mït*
lift-PST.3SG 3.SG –ACC lift -PST.3PL doorstep-DIM-1PL
ürdük-ke:n *e-te* *ürdük*
high –DIM be-PST.3SG high
'After relaxing and so on, well he lifted him, ehm they lifted him, our porch was high, high.' (TJP: 85)

(7.15) *üčügej-dik* *tut-uoχ-ta:χ-χïn* ***on-tu-gun,***
good-ADVLZR hold-FUT.PTC-PROP-PRED.2SG that -DER -ACC.2SG
ï:p-pat *kördük*
send-PRS.PTC.NEG similar
'You should hold that well, in such a way that it doesn't drop.' (ESB: 74)

Now, how meaningful is this observed difference? Is it only anecdotal, reflecting chance variation, or is the higher frequency of VO-clauses in Dolgan significantly different from Sakha? A quantitative analysis presented in the next section is intended to solve these questions. Three specific questions are addressed: a) Do word order patterns differ significantly across Dolgan and Sakha? b) Do word order patterns differ across text genres? c) Does the difference between Dolgan and Sakha hold for the language as a whole, or is the difference due to idiosyncratic language use of particular speakers?

7.5 QUANTITATIVE COMPARISON OF WORD ORDER PATTERNS IN DOLGAN AND SAKHA

7.5.1 METHODOLOGY

In order to make a quantitative comparison for SOV and SVO across the two languages, I coded transitive sentences for S, O and V in a randomly selected part of the corpora for both Dolgan and Sakha. For Dolgan, 5 narrative texts and 6 Pear Stories were coded for S, V and O, which yielded 512 utterances, produced by 11 different Dolgan speakers. These 512 utterances included intransitive sentences, transitive sentences with an unexpressed direct object and transitive sentences with an overtly expressed direct object. Since only the last category is relevant for the current analysis of V and O order, only these transitive sentences were included, totaling 175 sentences. Transitive clauses for which O was not overtly expressed (as exemplified in 7.16, where the verb has a subject, *iti* 'this' and an indirect object *uol oγoχo:nugar* 'to the little boy', but no direct object), were excluded from the analysis.

DOLGAN

(7.16) *iti* *uol* *oγo-χo:n-ugar...* *kim...* *ber-s-i-bit*
this boy child-DIM-DAT.3SG who give-RECP-PST.PTC
'To the little boy he... what-is-it-called gave (it).' (TIS: 11)

For Sakha the total number of coded utterances was 575, taken from 2 long narratives and 6 Pear Stories, narrated by 7 different speakers (1 of whom produced a narrative as well as a Pear Story). 176 utterances were transitive clauses with an overtly expressed object and were included in the analysis. The transitive clauses were further classified as OV and VO order, the frequencies of which were then calculated and compared across the two languages. After that, the potential significance of the frequency difference between Dolgan and Sakha was evaluated with the help of statistical models, which will be described below.

Before discussing the comparison in detail, a few points need to be made. First, so far word order has been discussed in terms of S, O and V. However, as mentioned before, in spontaneous speech the overt expression of S, O and V is the exception rather than the rule, in particular in pro-drop languages like Dolgan and Sakha. Therefore a more general, but more useful global categorisation was made of OV and VO order, where OV includes the theoretically possible patterns OV,

SOV, OSV, OVS, and VO includes VO, SVO, VSO and VOS. Second, since text genre potentially influences the frequency of transitive clauses in general, the text corpus used for this analysis consisted of spontaneous narratives, as well as the semi-spontaneous Pear Stories to control for text genre (see Section 1.2.2 for details). The idea behind this is that a story on the preparation of reindeer hide, or the construction of a boat will naturally include more transitive clauses than a story about one's family, since procedural texts typically involve agents acting on patients, which is the argument scheme for the prototypical transitive clause. Thus, the frequency of transitive clauses may be dependent on the chosen topic of the narrative. In addition, certain text genres may correlate with particular kinds of pragmatic structures, and may therefore favour a more frequent use of certain clause orders. By using semi-spontaneous texts it is possible to control the choice of topic and discourse pragmatic function of the narrative to some extent, thus increasing the degree of comparability across speakers. Although the interpretation of the film used to elicit the Pear Story can of course not be controlled and a certain level of variability will naturally remain, the uniform input considerably and sufficiently limits the divergence of the output (see Section 1.2.2).

Differences in word order frequency can be evaluated in several ways. The most straightforward way would be to compare percentages of occurring orders across languages and across genres, but the downside of this method is that it does not provide any information with respect to the significance of the different percentages. A much better result with regard to this issue can be achieved by applying statistical models, which are designed for dealing with just this task. The best model for the evaluation of the linguistic data in this study is a so-called Generalised Linear Mixed Model (GLMM)[1], more specifically a Poisson model[2]. A GLMM allows you to control for speakers' behaviour by including individual

[1] I am grateful to Roger Mundry, Matthias Urban and Michael Danneman for choosing the correct statistical models and applying them to my data.

[2] The Poisson model employed for these calculations used a log-link function and was built in R (R Development Core Team 2009), using code by Bates and Maechler (2010). This means that the data needed to be log transformed in order to fit the model. In this model, the total number of utterances was included as a log transformed offset term, controlling for effort. In a Poisson distribution, the mean is equal to its variance. If the variance is greater than the mean, or if it is dependent on the observed value, we speak of overdispersion and the model would not be appropriate for use. However, in the present study there was no issue with overdispersion ($\chi^2 = 12.94$, df = 16, p = 0.68, dispersion parameter = 0.81).

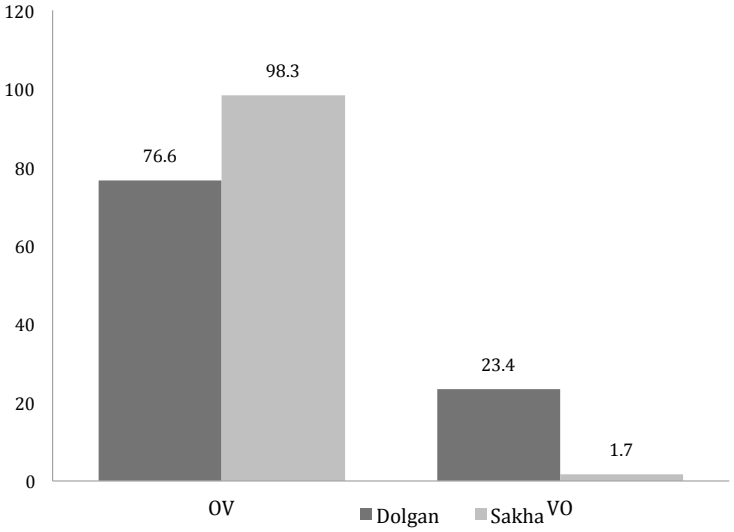
speakers as so-called random effects, which are factors that cannot be controlled by the experimenter. The inclusion of such random effects was a necessary requirement to the model, because not all data points were independent of each other, due to the fact that some of them were produced by one and the same speaker. If the statistical model does not account for this fact, there is a danger that the data might be biased in one way or another by the behaviour of an individual speaker. Since the independence of data points is mandatory for standardised models for significance such as chi-square or Fischer exact tests, these could not be used for the current purpose. Other factors that were implemented into the model as fixed effects were language (Dolgan vs. Sakha) and genre (Pear Story vs. narratives).

However, for a legitimate application of a GLMM to these data, it was necessary to carry out a test for model stability first. In principle, GLMMs also work only with independent data points. Since the present data set contained one non-independent data point (i.e. the speaker who participated in the production of language data for both genres) it needed to be proven prior to the choice for the Poisson model that the effect of this data point was no different from the effect of the independent data points. For this purpose a Generalised Mixed Model was used, showing that the single non-independent data point did not cause any different effects than the independent ones. Therefore, the GLMM model could be applied with clear conscience.

7.5.2 RESULTS

Figure 7.1 shows a summary of the distribution of OV and VO order across Dolgan and Sakha. In this figure, data from spontaneous narratives and Pear Stories were collapsed because it turned out that text genre had no significant effect on the clause order in Dolgan or Sakha (see below for details). The proportions of OV and VO order are calculated relative to the total number of transitive clauses with overt expression of V and O in the texts. The dark grey columns reflect the proportions in Dolgan, whereas the light shade represents the results for Sakha.

Figure 7.1 Proportions of OV and VO-order in Dolgan and Sakha



As can be seen from this figure, there is a noticeable difference in proportions of OV and VO occurrence across the two languages. In Sakha, the ratio of OV, the typical word order for the Turkic language family, is 98.3%, whereas for Dolgan this is only 76.6%. On the other hand, VO order occurs only in 1.7% of the clauses in Sakha, whereas the proportion of 23.4% in Dolgan is much higher. The exact numbers, specified for language as well as for text genre (i.e. for spontaneous narratives and Pear Stories separately) are provided in Table 7.1 below.

Table 7.1 Numbers and proportions of OV and VO clauses in Dolgan and Sakha

	Dolgan				Sakha			
Word order	OV		VO		OV		VO	
	no.	%	no.	%	no.	%	no.	%
Narrative	93	78.2	26	21.8	101	98.1	2	1.9
Pear Story	41	73.2	15	26.8	72	98.6	1	1.4
Total	134	76.6	41	23.4	173	98.3	3	1.7

The numbers in the table suggest that the text genre does not have much influence on the distribution of OV versus VO orders. In Dolgan, OV order is attested for 78.2% of the transitive clauses in spontaneous narratives, and 73.2% in the semi-spontaneous Pear Stories. For VO order a comparable similarity between

the text genres is found, namely 21.8% for narratives, and 26.8% for the Pear Stories. In Sakha, OV order occurs in 98.1% of all overt transitive clauses in spontaneous narratives, and 98.6% in the Pear Stories. VO clause order occurs in 1.9% in the narratives, and 1.4% in the Pear Stories. While these numbers suggest homogeneity in word order distribution across text genres, and a different distribution across languages, statistical tests are needed to establish whether this intuition is correct, in other words, whether the frequency differences between languages and across text genres are significant or whether they are likely to reflect chance variation.

The calculations made by the GLMM Poisson model reveal that the frequency of occurrence of VO order indeed differs significantly across Dolgan and Sakha. It shows that VO order occurs significantly less in Sakha than in Dolgan. This is evidenced by the significance of the so-called estimate value, which is an estimation made by the model with respect to the relative frequency of a certain result in Sakha and Dolgan (estimate = -2.63, $p < 0.0001$). At the same time they confirm that genre makes no significant difference for the occurrence of OV or VO order within languages ($p = 0.599$) suggesting that the proportions of OV and VO are stable, regardless of whether the text was a spontaneous narrative or a semi-elicited Pear Story.

Finally, a stability test was carried out to see whether one of the data points, i.e. speakers, could be the cause of the observed patterns. This is tested by constructing models in which one of the data points is removed at a time, and comparing the range of estimates for those models with that of the original GLMM. This shows neither a strong impact on the estimate for the differences between languages (range from -3.11 to -2.47, estimate for the original model: -2.63) nor on that for genre (range from 0.01 to 0.33; estimate for the original model: 0.165).

Summarising we can conclude that Dolgan and Sakha differ significantly from each other with respect to the frequency of VO clause order, regardless of speaker and of text genre. The proportion of atypical VO clauses is much higher in Dolgan than in Sakha, which could foreshadow an ongoing change in Dolgan. The next section will be concerned with possible explanations for this development.

7.6. DISCUSSION

Three possible scenarios come to mind for the explanation of increased VO order in Dolgan. First, it could be a language-internal development. Second, it could be motivated by contact with Russian, and third, it could have developed under the influence of contact with neighbouring indigenous languages. Of these three possibilities, the third can quickly be dismissed, since the neighbouring Tungusic language (Evenki), as well as the Samoyedic languages (Nganasan, Enets, Nenets) have the same SOV basic word order as Dolgan. Therefore, influence of these languages in this linguistic domain would have no noticeable effect. As a result, a stimulating role of Tungusic or Samoyedic speakers with respect to the increase of SVO order in Dolgan can be confidently excluded.

7.6.1.1 LANGUAGE-INTERNAL MOTIVATIONS FOR WORD ORDER CHANGE

With respect to language-internal development, a change from SOV to SVO is cross-linguistically not uncommon. Dik (1997) explains this in terms of the 'Principle of Increasing Complexity'. In his words, this means that there is a preference for ordering constituents in an order of increasing complexity' (Dik 1997: 404), where the concept of 'complexity' roughly corresponds to concepts like the 'Gesetz der Wachsenden Glieder ('law of increasing parts') formulated by Behaghel (Behaghel 1909: 139) or 'heaviness' (Hawkins 1983: 90, Mallinson-Blake 1981: 158). In Hawkins' terms, 'heaviness' is a composite notion defined in terms of: a) the length and quantity of morphemes; b) quantity of words; c) syntactic depth of branching nodes; and d) inclusion of dominated constituent. The heavier a constituent, the more likely it is to be placed to the right of the head of the clause. Dik (1997: 410) even dedicates a special 'principle' to it, namely 'Specific Principle 6', which states that

> [t]he Prefield is universally less hospitable to complex material than the Postfield. Prefield languages may thus be expected to possess strategies for relieving the Prefield of excessive complexity.

In this quotation, the prefield is the position preceding the head, and the postfield the position following the head of the sentence. He goes on to say that one can

distinguish between 'strict' and 'liberal' prefield languages (that is, SOV languages), where the more liberal prefield languages allow for a certain amount of 'leaking' of constituents beyond the head (Ross 1973, as cited by Dik 1997: 410). This, he argues, may diachronically lead to a gradual change from an SOV language into a 'Prefield-derived SVO language'. These are SVO languages, which have retained a number of prefield properties, such as Karaim, which has adopted the relative pronouns and right branching relative clauses from Russian, but has kept participles and converbs to the left of the verb (Johanson 2002a: 131-137).

Although the preceding account has shown that several scholars employ concepts such as 'heaviness' or 'complexity' to explain tendencies in internally motivated language change, there is no unequivocal explanation for the existence of the principles themselves. For a long time it was assumed that the attested tendency was motivated by general psycholinguistic principles and constraints on processing. Left branching structures in general were assumed to put a heavier burden on memory in production (Yngve 1961, in Johanson 2002a: 120) as well as in comprehension, and would therefore be disfavoured. The argument is that in production the speaker needs to plan the entire sentence before he can even start producing it, due to the fact that the head is in final position. Likewise, the hearer needs to remember all the details and modifications before the eventual head is revealed at the end of the clause. However other studies show that the processing complexity of left-branching structures is not any greater than for right-branching structures (Frazier & Rayner 1988, in Johanson 2002a: 120).

Despite contradictory results in psycholinguistic research, the tendency in languages to position longer and more complex constituents towards the end of the sentence, and thus of 'leaking', remains a fact. Because of this natural tendency in one direction, it is not surprising that the opposite direction of internally motivated language change, from SVO to SOV, is cross-linguistically less common. There are even claims that go as far as to say that a change from SVO to SOV can only occur as a result of contact, and would never happen as a language-internal process (Ross).

These facts leave open for consideration the possibility that the increase in SVO clause order in Dolgan could be a language-internal development, following a universal tendency in language change. However, as can be seen from examples like 7.14 and 7.15, heaviness or complexity can certainly not always be adduced as a motivation for SVO order. The object pronouns are in fact, apart from omission, the shortest possible way of expressing objects. In addition, if it were a change

independent of contact, one would expect it to be pervasive throughout the whole area where the language is spoken, and to occur irrespective of geographical location or sociolinguistic setting. However, the next section will show that this is not the case.

7.6.1.2 LANGUAGE-EXTERNAL MOTIVATIONS APPLIED TO DOLGAN

An investigation of the distribution of VO-order across the villages reveals that its frequency varies depending on geographical location. More specifically, it correlates with the sociolinguistic situation prevailing at the geographical location with respect to the use of Dolgan and Russian. As was described in Section 1.3.2, the villages where I recorded the Dolgan narratives differed considerably with respect to the balance of linguistic dominance between Dolgan and Russian, as well as the attitude towards use of each language. It was mentioned that the Dolgan language is most vital in the villages that are furthest away from the Russian-dominated centers, and that its use gradually decreases as one comes closer to the towns, in particular Dudinka. As can be seen from Table 7.2, the proportion of Turkic OV and Slavic VO-clause order in the speech of the language consultants correlates with this difference in sociolinguistic setting, in particular with the increase of Russian dominance.

Table 7.2: Percentage of OV and VO clause order per community

	OV	VO
Syndassko	90.1%	9.9%
Kheta	70.0%	30.0%
Dudinka	70.7%	29.3%

To put it concretely, there is an increase in VO structures when travelling from east to west, i.e. from Syndassko over Kheta to Dudinka. In Syndassko, VO structures constitute 9.9% of the transitive sentences, in Kheta 30.0%, and in Dudinka 29.3%. These results show that the higher occurrence of VO structures in communities with a strong social and linguistic representation of Russian could be due to transfer of such structures from the dominant Russian into non-dominant Dolgan.

Another possible factor influencing the distribution of word order patterns could be speaker age. If apparent time[3] gives a realistic representation of ongoing language change and if one follows the general idea that children are the main locus of language change, or the less widespread idea that innovative structures are predominantly promoted by preadolescents (Ross forthcoming), one would expect younger speakers use innovative SVO structures more frequently than older speakers do. Since it was not possible for me to find enough speakers in each age group for a reliable sample, it is currently not possible to say with certainty whether age plays a role or not. However, the impressionistic data shown below go against this expectation.

Table 7.3: Percentage of OV and VO clauses per age group

Age	Location	OV	VO	Total OV	Total VO
75	Dudinka	70.7%	29.3%	75%	25%
	Syndassko	84.2%	15.8%		
40	Kheta	70.0%	30.0%	77.3%	22.7%
	Syndassko	92.9%	7.1%		
14	Syndassko	93.3%	6.7%	93.3%	6.7%

Table 7.3 includes three age groups, roughly corresponding to three generations, and the percentage of OV and VO clauses they produced in their spontaneous speech. Due to the labour-intensiveness of manual word order counting, the number of individuals per age group is only two. For the youngest age group only one individual was included because the narratives from the other children in this age group were unsuitable for the current purpose, either due to the absence of transitive clauses, or to interference of the parent. The table shows that the distribution of OV and VO order within the age groups is far from homogenous. In the age group of 40, one speaker uses VO order in 30.0% of the transitive clauses, whereas the other uses it in only 7.1%. A similar situation, though less extreme, applies to the age group of 75. Since the number of individuals is so low, this diversity could of course be due to chance, but it seems that word order patterns cannot be correlated with a particular generation. Second, the average frequencies

[3] 'Apparent time' is the idea that language variation between speakers of different age groups at a particular moment in time is representative of the development of the language through time. According to this idea, synchronic language variation can be used to study diachronic language change (Labov 1994: 28-29, Chapter 3).

for age groups (found in the columns 'total OV' and 'total VO') are not distributed across these groups in the way expected for ongoing language change, if credibility is given to apparent-time predictions[4]. This could be taken as an argument against language-internal innovation for this particular feature in Dolgan.

7.6.2.1 LANGUAGE CONTACT AND WORD ORDER CHANGE

As it turns out, the literature about the nature of word order change is ambiguous. From the perspective of language contact studies, word order is characterised as a linguistic feature that is affected in contact situations relatively easily, whereas the literature on language acquisition classifies word order as a 'deep' structural feature that is supposedly very resistant to influence from other languages. Thomason and Kaufman write about word order change in contact situations:

> The evidence we have collected does not support the often implicit assumption, in the literature on word order change, that word order patterns constitute a fundamental deep structural feature relatively impervious to foreign influence. On the contrary, word order seems to be the easiest sort of syntactic feature to borrow, or to acquire via language shift. (Thomason and Kaufman 1988: 54-55)

They explain this by the idea that SOV and SVO patterns both fulfill the syntactic function of identifying a subject and object with respect to each other and to the verb. They illustrate their argumentation with examples from Finnish, which changed from SOV to SVO order under the influence of Indo-European languages and Austronesian languages of New Guinea, which show a change in the opposite direction (SVO to SOV) due to contact with Papuan languages. These cases seem to be instances of heavy copying of structure in a situation of language maintenance, and as far as I can tell they do not give examples of word order change in language shift situations. However, the message remains valid nonetheless: word order patterns do change under the influence of contact.

---

[4] I realise, of course, that a larger sample of texts and informants would potentially provide a different picture. Multiple speakers for every location would have been desirable, but were not always possible to find.

This view is confirmed in later work by Thomason (2001: 88) in which she repeats that basic word order is among the most common features that are affected by structural interference. While she labels this phenomenon primarily as a "replacement of native linguistic features by new interference features" (ibid.: 87), she adds the possibility that change in word order may be the result of convergence, which she defines as "any process through which two or more languages in contact become more like each other" (ibid.: 89). The term convergence for her implies that it is impossible to clearly define a source language and a recipient language. Rather, the languages converge towards each other, being both source and recipient language at the same time. In this context she adduces the example of Kadiwéu, spoken in Brazil, which in natural discourse shows six different word orders (including SVO), but in translations from Portuguese displays an unusually high frequency of SVO word order, copying the unmarked Portuguese order of constituents (Sandalo 1995 in Thomason 2001: 89). Although the adaptation is unidirectional, Thomason prefers to characterise this case as convergence, since SVO word order was already present in Kadiwéu, and therefore it would be inappropriate to call Portuguese the source language and Kadiwéu the recipient language. According to Thomason, this does not necessarily represent a change in Kadiwéu, although she admits that it could eventuate in it, but rather is an example of how changes can start through a shift in frequency of particular constructions[5] (see also frequential copying (Johanson 2002b: 292, Heine and Kuteva 2005: 47) as discussed in Section 6.4).

This is in line with the ideas formulated by Thomason (2001: 69), who describes word order as among the "next easiest things to borrow", after the lexicon. However, Heine (2008) argues that in fact there is no case in which 'new' word order is completely new and unprecedented in the language:

> What frequently happens is that speakers draw on a minor use pattern – one that has a more marginal status, being used rarely and/or only in specific contexts only to build a new major use pattern by increasing the frequency of use and extending the range of contexts in which it may occur. (Heine 2008: 55)

[5] While Kadiwéu in this process is indeed (its own) source and recipient language at the same time, enhanced by contact with Portuguese, the change still seems to take a one-way direction form Portuguese to Kadiwéu. Whether Kadiwéu influences Portuguese in other linguistic domains is not further specified.

He admits that often this process does not lead to a complete change in word order, in which case it may result in a more flexible word order in the recipient language. As an example he gives Eskimo speakers of North America who are in contact with English speakers. He observes that instead of changing their own word order pattern completely on the strict SVO model of English, Eskimo word order has simply become more free (Heine 2008: 57).

These accounts give the impression that word order change is rather common and easily achieved in language contact situations, and the considerable number of case studies provides supporting evidence. However, these accounts are relatively unspecific with regard to one or more of the following factors: the sociolinguistic conditions in which they occurred (language maintenance or shift), the identity of the initiators of change (L1 or L2 speakers), and the underlying processes of change (borrowing or imposition). As was pointed out in Chapter 3, the combination of these factors is important for an accurate description and analysis of any contact situation and its linguistic outcomes (see Sections 3.1.3, 3.1.4). Through the complex interplay of each of these factors, different contact situations may lead to the same surface outcomes, and reversely, comparable contact situations may lead to different outcomes. Only detailed sociolinguistic information will make it possible to describe and/or reconstruct the events that underpin the outcomes of language change in the most realistic way.

7.6.2.2 LANGUAGE ACQUISITION AND WORD ORDER CHANGE

The literature on the acquisition of syntax presents a different view of word order change. Part of this may be caused by the fact that much of the work in this field has been dominated by ideas from generative linguistics. As was briefly mentioned before, a common assumption in this research tradition is that word order belongs to the so-called 'deep structure' of language, and thus cannot be changed after the relevant 'parameter' for this feature has been set (e.g. Lightfoot 1979). One argument in favour of this assumption comes from studies on L1 attrition in individuals who have lived abroad for a long time and hardly ever use their L1. Several studies show (Schmid 2002, Altenberg 1991 in Lucas 2012: 282) that the difficulties these individuals experience in comprehension and production of their L1, including word order, may be only "the result of temporary difficulties with access and retrieval". (Lucas 2012: 281). According to Lucas this indicates that in

fact the parameter setting, or competence, of the bilingual speaker has not changed, but only that the performance has changed temporarily due to high activation of L2 and low activation level of L1. Since in this view word order is strictly tied to parameter settings that are fixed in early childhood, the implication is that word order change can only be initiated by young infants, at a time when this parameter is set differently from that in the language system of their parents. The consequence of this idea would be that contact must be excluded as a potential cause of word order variation and change, because the only source of change (the infant) does not exactly actively engage in influential social contact, and it does not have the necessary network ties for new language variants to be spread across the community. Although Lightfoot does recognise language variation and contact as possible external factors in the process of language change (Lightfoot 1979: 374), the pathway by which he proposes that contact-induced variants reach, and eventually settle into the grammar of the infant is not entirely clear.

Additional work on language acquisition and change supports the skepticism towards the monopoly of infants in language change. For example Aitchison (1981: 180) supports the idea mentioned above that babies cannot be the main source of language change since they do not have the social significance nor the network ties needed for new variants to become favourable over others and spread through the community. Instead, she argues that language variation is only meaningful for the field of language contact when it occurs in children from 4 years onwards, when they begin to engage in social activities, and identify with certain social groups, which may be different from their parents. She puts special emphasis on preadolescents and adolescents because these are the groups whose members are easily influenced by peers and people a little older than themselves, but are, counter to generative convictions, still able to make fundamental changes to their language (c.f. Light Warlpiri and Gurinji Creole in contact with English based Aboriginal Kriol (O'Shannessy 2005, McConvell and Meakins 2005), German in contact with English (Clyne 1992), all examples adduced by Ross forthcoming).

As will be recalled from Section 3.1.5.2, this view is strongly supported by Kerswill's work, which shows that language changes throughout a person's lifespan, but that the different kinds of change are conditioned by a person's life stage. Partly in line with Aitchinson's conclusions, Kerswill attributes the greatest significance in the emergence of innovative grammatical patterns to preadolescents, and not to adults or in infants (Kerswill 1996: 198). While Kerwsill's

studies focus on the emergence of differences between dialects, in a recent paper Ross applies similar ideas in his explanation of change in contact situations involving different languages, when he argues that preadolescents are also crucial agents in the initiation of contact-induced grammatical change, including calquing of word order patterns (Ross forthcoming).

To recapitulate, the dominant idea in the acquisition literature that word order belongs to 'deep structure' and therefore cannot be changed after infanthood, thus tacitly implying that word order cannot change due to contact, is contradicted by multiple case studies. This, in combination with evidence from language contact theory now opens the way for an account in terms of language contact to explain the word order variation in Dolgan.

7.6.3 WORD ORDER VARIATION IN DOLGAN EXPLAINED

Returning to the data from Dolgan, it is clear that it would be wrong to assume that the increase in SVO structures is due to language-internal factors alone. The observations that the heaviness principle does not always apply in SVO structures and that SVO order does not correlate with age in the expected direction, in combination with the fact that high SVO frequency is found in an area of intense contact with an SVO language, argues against this explanation, and in favour of an account in terms of contact. Of course, the observed tendency to develop SVO structures through language-internal change may certainly have enhanced this process in Dolgan, but taking the fact that this change is cross-linguistically common as the single explanation, would ignore an obvious and significant aspect of the story, which became clear from the correlation shown in Table 7.2.

The fact that SVO order was already an available, but pragmatically marked, structure even before contact with the Russians intensified also facilitated the extension of this construction into less marked contexts (see Johanson 2002a: 111-112, 2002b: 292, Heine 2008: 31, 43, 56-57). Nonetheless, contact with Russian seems to have been the main trigger for the introduction of the option of VO word order. Accepting contact with speakers of Russian as a primary explanation for this difference between Dolgan and Sakha, questions arise with respect to a) the relative status of the languages in contact; b) the initiators of the change (children, adults or preadolescents, L1 or L2 speakers of Russian?) and c) the process underlying the change (borrowing, attrition, imposition?).

The sociolinguistic situation on the Taimyr Peninsula leads me to think that this development in Dolgan must be the result of more than just one process of change on the level of the bilingual individual. With respect to the initiators of change, it was not possible to identify one age group in which SVO occurs consistently more than in others (see Table 7.3). Rather the innovative word order patterns seem to occur in speakers of all age groups who live in a Russian-dominated environment, and for whom Russian has become their dominant (in Van Coetsem's terms), and most highly activated (in Lucas' terms), language. In the remaining part of this discussion, I will focus on this group of speakers only.

While the individuals in this category are all Dolgan people who are in a Russian-dominant environment, even this group is anything but homogeneous, and includes people with very different levels of proficiency in Dolgan. Following Van Coetsem's theory, this would mean that the same result (SVO sentence structures) can be explained by two different processes of change depending on the linguistic dominance (typically correlating with age) of the speakers. The argumentation for this is rendered schematically in Table 7.4.

Table 7.4: Linguistic dominance and processes of change in different age groups

	Age group	
	> 70	< 40
L1 (dominant)	Dolgan	Russian
L2 (non-dominant)	Russian	Dolgan
Direction of transfer	L2 → L1	L1 → L2
Agentivity	L1	L1
Process of change	Borrowing	Imposition

Of course such a differentiation can only be made on the level of the individual, and while the change is in progress. Although it was argued before that age does not play a role in the *frequency* of occurrence of SVO structures, it is indirectly a distinguishing factor when it comes to the *process* underlying the appearance of these structures in Dolgan, because of its link with linguistic dominance. In the table above, the youngest age group, including children and teenagers, has been left out. Their Dolgan did not display much influence from Russian in the most isolated village of Syndassko, and in other villages children do not speak Dolgan anymore, thus making influence from Russian complete and predicting a shift to Russian in these communities in the near future.

The first group consists of the oldest generation (average age 75), who grew up in the 1940-50's. Although Russian influence was already strong on the Taimyr in this period (e.g. children were often forbidden to speak their native language in public), children would still have been brought up by Dolgan-speaking parents, who were often monolingual in this language. Thus one may assume that their L1 and dominant language was Dolgan in infanthood and early childhood, and that they later learned Russian as an L2 in school. Despite the increasing presence of Russian in their community, the Dolgan language would remain for them an important means of communication in the interaction with their parents and other members of the community in non-public settings. Disregarding individual exceptions for the sake of generalisation, one can say that this generation is bilingual in Russian, but has remained dominant in Dolgan, regardless of occasional higher activation levels of Russian, which are situationally conditioned. Against this background, the presence of SVO word order in their variety of Dolgan can be best explained through the process of borrowing, more specifically structural borrowing occurring in a situation of intense contact (Thomason and Kaufman 1988 and Section 3.4.1 of this thesis). Structures from the non-dominant source language (Russian) are transferred to the dominant recipient language (Dolgan) due to high exposure to the source language in the community. Possibly this happens to reduce processing costs for the speaker, as well as for the hearer for whom Russian is most probably also the most accessible language.

A different situation holds for the younger speakers (40 and younger). Growing up in the 1970's and later, these Dolgan individuals had bilingual parents, and were mostly settled in Russian-oriented villages. Even if they spent the first few years of their life in the tundra, they were brought to boarding school from the age of 5 where any initial Dolgan dominance would quickly disappear. The boarding schools were monolingually Russian, and the use of indigenous languages was not at all appreciated, if not forbidden. This led automatically to a change in attitude towards both languages. Russian was represented as prestigious and the language of education and development, and children would speak it to their teachers, but often also to each other. In the beginning they did this perhaps mainly so they would not get 'caught' speaking an indigenous language by a teacher, but later possibly because Russian became the more activated and therefore easier language, so that in addition to the social issues, retrieving Dolgan would mean greater psycholinguistic effort. For these people it is highly questionable whether Dolgan can be called their L1 and it is certainly not their

dominant language. They speak Russian like native speakers, whereas their use and knowledge of Dolgan is more limited and not as fluent. If they do speak Dolgan, code-switching with Russian is common. Therefore, for this generation Dolgan has often acquired the status of a non-dominant L2. It goes without saying that also to these statements there are exceptions, but the purpose here is to characterise the general tendency.

I deliberately chose to talk about L1, L2 and dominant language instead of referring to Russian or Dolgan as the 'native' language for the following reasons. Despite the often poor knowledge of Dolgan in the younger generation, many Dolgans would still say that their native language is Dolgan. This is completely justified considering the fact that they are Dolgans, and it was one of the languages they grew up with from birth. However, this choice seems to be based rather on factors such as ethnic identity and association with a certain ethnolinguistic group than on actual linguistic proficiency. The sociolinguistic features of the group under forty leads me to the conclusion, that word order change in these people (i.e. people whose dominant L1 is Russian and whose non-dominant language is Dolgan) is the result of the process of imposition. Their high exposure to, and psycholinguistic dominance in Russian causes them to project sentence structures from their L1 (Russian) onto their non-dominant L2 (Dolgan).

The differentiation between borrowing and imposition can only be made at the level of the individual speaker, while the change from SOV to SVO is in progress. Once it is completed, the detailed information on individual variation will no longer be available. Therefore, diachronically, and at the community level, this change is best explained by imposition by Dolgans who are dominant speakers of Russian.