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 5 REGULARISATION OF PARADIGMS

5.1. INTRODUCTION

Apart from the lexicon, the morphology of Dolgan shows several points of divergence with Sakha. While the overwhelming majority of morphological paradigms is identical in the two languages, a number of differences can be observed, which require closer investigation because they seem to be restricted to the Dolgan-speaking area, and second because they can be subsumed under a common heading of paradigm regularisation, a phenomenon which is not unusual in internally, as well as externally motivated change.

The first phenomenon to be addressed in Section 5.2 is regularisation in the nominal paradigm. It appears that for nouns with a particular phonological structure the forms in Dolgan have a different underlying stem from their cognates in Sakha. More specifically, in Sakha these stems have an irregular declension paradigm, whereas in Dolgan the paradigm has become regular. Counter to previous discussions, in which this difference was assumed to be purely a result of language-internal phonological change, I will argue that this regularisation is the result of a more fundamental cognitive process of reanalysis, motivated, or reinforced by the presence of a substantial number of L2 speakers in the Dolgan-speaking community. The second example concerns the inflectional paradigm of the defective auxiliary verb *e-* 'to be' and is described in Section 5.3.

Here the inflectional suffix of the third person plural, which is normally an irregular form within a paradigm of which the endings are otherwise identical to the suffixes of possessive person marking, has been synchronised with the paradigm of possessive person marking, and thus has become more regular. I will argue that this instance of regularisation has occurred on the basis of perceived analogy between the inflectional paradigm of *e-* 'to be' and the paradigm of possessive person marking, and that L2 speakers may have played a significant role in the establishment of this change in the speech community. A careful evaluation of language-internal and language-external factors in the development of these changes will be pursued in Section 5.4.

5.2. REGULARISATION OF NOMINAL PARADIGMS

5.2.1 DESCRIPTION OF THE PHENOMENON

In Sakha, the majority of noun stems are inflected in a regular way. They have a transparent agglutinative structure, consisting of a clearly identifiable stem, followed by suffixes of case (ex. 5.1) possession (ex. 5.2), possessive case (ex. 5.3), or predication (ex. 5.4).

SAKHA

(5.1) ***oskuola-ɣa*** *ïːp-pa-ta*
school-DAT send-NEG-PST.3SG
'She didn't send me to school.' (ARR: 41)

(5.2) *kör-büt-üm,* ***oɣo-m*** *öl-ön* *χaːl-bït*
look-PST.PTC-POSS.1SG child-POSS.1SG die-SQ.CV RES-PST.PTC
'I looked, my child had died.' (ARR: 44)

(5.3) ***ïnaχ-pïtïn*** *tut-tu-lar*
cow-ACC.1PL hold-PST-PL
'They took our cow.' (ARR: 27)

(5.4) *(...) araːs* *buld-u* *barï-tïn* *bul-taː-bït* ***kihi-bin***
(...) various catch-ACC all-ACC.3SG catch-VBLZR-PST.PTC person-PRED.1SG
'I am a person who hunted all the various animals.' (AIC: 46)

However, in a small set of Sakha words (the so-called unstable stems) the stem is modified due to morphophonological rules in such a way that the surface form becomes ambiguous to the hearer and opaque with respect to the shape of the underlying stem. For example, the Sakha form *kennitten* 'from behind' consists of a stem *kelin* 'back part' and a possessive marked ablative suffix with the underlying form *-(t)IttAn*. Due to rules of consonant assimilation and vowel harmony (see Section 5.2.3.3.1) the combination of stem and suffix results in the surface form *kennitten*. In Dolgan, however, the corresponding third person possessive ablative form is *kennititten* (containing an additional syllable *ti*) which consists of a stem *kenni* and the suffix *-(t)IttAn*. The different forms and their underlying morphological structures are presented schematically in Table 5.1.

Table 5.1: Ablative of kelin 'back part' in Sakha and Dolgan

Language	Ablative	Stem	Translation
Sakha:	*kenn-itten* *kelin-(t)IttAn* back.part-ABL.3SG	*kelin*	'from behind'
Dolgan:	*kenni-titten* k*enni-(t)IttAn* back.part-ABL.3SG	*kenni*	'from behind'

Instead of treating *kennitten* as a word with an unstable stem *kelin,* which can only be inferred with the help of complicated rules, it seems that speakers of Dolgan have taken a more straightforward interpretation of the Sakha form *kennitten*. Due to the ambiguity of this surface form, they have taken the 'mutated' stem *kenni* as the basis for inflection and have derived the underlying structure directly from this form. This suggests that the inflected form encountered in Sakha has undergone 'reanalysis' in Dolgan.

The recognition of such a difference is one thing, but more interesting is the question what could have motivated this change. While seeking to explain this development in Dolgan, particular attention is paid to the question whether the most plausible explanation is found in language-internal processes of change, or whether this phenomenon is better explained by language-external motivations, such as second language learning and language contact. Before addressing these issues in depth, some theoretical background is given on the characteristics of reanalysis in the next section.

5.2.2 REANALYSIS

Reanalysis is an important mechanism of change in syntax and morphology. Aikhenvald defines it as

> (...) a historical process whereby a morphosyntactic device acquires a different structure from the one it originally had with little or no change to its surface form or semantics. (Aikhenvald 2006: 30)

Harris and Campbell add that reanalysis "depends upon surface ambiguity or the possibility of more than one analysis." (Harris and Campbell 1995: 3). One example comes from Udi, a Lezgian language from the East Caucasian language family. For Proto-Lezgian the verb structure is reconstructed as a verb stem preceded by a vowel[1] and a prefix for gender-class according to the following scheme:

(5.5) class marker + vowel + verb stem

Schulze (1982: 148, cited in Harris and Campbell 1995: 67) demonstrates that a number of verbs that had such a structure originally, are nowadays treated by speakers as an unanalysable stem. Thus, a verb like *b+o+q* 'love, want', in which *b* is the gender class of 'other living things', *o* is the inserted vowel and *q* the original verb stem, would nowadays more accurately be represented as a single unit *boq*. As can be seen from this example, the surface form in both cases is *boq*, but the underlying structure of the form in Proto-Lezgian and in Udi is different, and this corresponds to the definition of reanalysis given above. Not surprisingly, Schulze argues that this development is connected to the fact that Udi is losing the old system of gender-class agreement.

The case of Udi is an example of the loss of morpheme boundaries, but the merging of multiple morphemes into a single unit is not the only way in which reanalysis is manifested. The opposite development is also attested, and speakers can create new boundaries, as happened in the history of the English word *pea*. In the case of *pea* the original singular form was *pease*, and its final *-s* later became interpreted as a plural ending *-s* in analogy with other English plurals ending in *-s* (Lehmann 1992: 223). Thus, reanalysis took place from *pease* > *pea-s* and a new

[1] Harris and Campbell note that the vowel was not originally part of the verb (Harris and Campbell 1995: 66-67).

morpheme boundary was created where there wasn't one before. As third possibility they may shift a boundary to a different position in the morpheme, often motivated by factors such as analogy or iconicity. An example of boundary shift is the French *argent-ier* [silver-AG] 'treasurer' which served as a model for other words such as *bijou-tier* [jewel-AG] 'jeweler', where the *-t* originally belonged to the lexical stem of the noun *argent* and became incorporated in the suffix, leading to a suffix *-tier* (ibid.)

As mentioned before, reanalysis can occur when an alternative analysis of a morpheme (lexical or inflectional) becomes possible, and more plausible, to speakers for various reasons. When both analyses are still available, this results in allomorphy, but eventually it can lead to a permanent change, when the initial allomorph becomes unacceptable to speakers. This process can proceed via the following pathway. Allomorphs can emerge through analogy with other forms in the language (as in the example for *pea*), or through a change in other domains of the language system, in particular phonology. If a phonological change leads to new allomorphs that are ambiguous with respect to their underlying morphological structure, there is the potential for reanalysis to take place (Koch 1996: 237). Whether or not it happens depends on various factors, including economy of processing, frequency of occurrence of the new allomorph (and potentially other forms of the stem) in paradigms as well as in texts, as well as cognitive processes relating to iconicity and markedness.

5.2.3. SAKHA NOMINAL DECLENSION

5.2.3.1. RELATIONAL NOUNS AND REFERENTIAL NOUNS

According to traditional grammatical description, there are two types of nouns in Sakha: independent referential nouns and relational nouns. Although relational nouns do not differ from referential nouns in their inflectional paradigm, the two types do differ from each other in function and context of use, which in turn has consequences for the frequency of occurrence of particular formal properties such as case and possessive marking. To make this more concrete, referential nouns can occupy all main grammatical slots, such as subject and object, and fulfil all basic semantic functions, such as agent, patient and recipient. They denote 'an object or

an objectivised notion'[2], which can occur as an independent unit in the sentence. Relational nouns, on the other hand, are nominal stems that were historically referential nouns, but are used in present day Sakha with grammatical functions and a more figurative meaning. Unlike referential nouns, relational nouns do not occur in basic grammatical functions such as subject and object. They cannot occur as an independent constituent in the sentence and only appear in a dependency relation with other nouns, in particular to specify location, as is illustrated in example 5.6.

(5.6) *die* *ürdü-te*
house top-POSS.3SG
'top of the house'

Since phrases of this type have a schematic structure of NOUN + NOUN-(CASE.)POSS, in which the first noun is the referential noun, and the second one the relational noun, relational nouns almost always occur with possessive and/or case marking (see 5.2.3.2 for more details). As a result, relational nouns hardly ever occur in the unmarked nominative form, which makes it hard for a hearer to determine the underlying stem, particularly if the noun belongs to the category of unstable stems referred to in 5.2.1. This variation in surface form makes the unstable stems more prone to reanalysis than referential nouns, which appear more regularly as a nominative.

5.2.3.2 NOUNS IN PHRASES OF LOCATION

As in many other Turkic languages, dependency relations in Sakha are often expressed by means of the so-called *izafet* construction. This construction, which was copied into Turkic from Persian, expresses a dependency relation between a head noun and a modifier noun by means of an agreement feature (possessive marking) on the head noun. This applies to possessive relations with a literal (ex. 5.7), as well as with a figurative possessive meaning (ex. 5.8), the latter merely establishing a connection between the two nouns, as in phrases of location.

[2] "Имя существительное – часть речи (особый лексико-грамматический разряд слов), обозначающая предмет или опредмеченное понятие [...]." (Убрятова 1982: 108).

In Sakha, *izafet* constructions consist of a modifier noun in the nominative case, and a possessive marked head noun, which can be marked for case as well. As Sakha is a head final language, the possessive marked head noun occurs in final position in the phrase, and the unmarked modifier noun in initial position, resulting in the schematic structure referred to in the previous section: NOUN+NOUN-(CASE.)POSS.

SAKHA

(5.7) *elbeχ* *aŋar-bït* *Ha:skïla:χ-χa* *ba:l-lar,* ***iỹe-m***
many half-1PL Saaskylaax–DAT EXIST-PL mother-POSS.1SG
tördü-ler-e
ancestor-PL-POSS.3SG
'More than half (his relatives) are in Saaskylaakh, the ancestors of my mother.' (PNL: 13)

(5.8) *onton* *bal:an* *ïj-ïn* ***otto-tugar*** *köt-ü-t-en*
then yurt month-GEN middle-DAT.3SG fly-EP-CAUS-SEQ.CV
bardïlar
go-PST-PL
'Then in the middle of September they sent him off by plane.' (XLE: 236)

Phrases of location may express location in space or time, and in order to specify the nature of the relationship between the modifier noun and the head noun, the head noun is specified for case, in particular dative (location), ablative (direction from), and instrumental (mode). Thus, in a locational *izafet* construction, the head noun is always marked for possession to establish the dependency relationship between the head and the modifier, and for case to specify the nature of this relationship. This is illustrated in examples 5.9 and 5.10.

SAKHA

(5.9) *dʒie* ***kenniger*** *χoton* *ba:r*
house back.part.DAT.3SG cowshed EXIST
'Behind the house is the cowshed.' (elicited)

(5.10) *ïp-pït* *bes* ***kennitten*** *(...)* *kör-ö* *tur-ar*
dog-1Pl pine back.part.ABL.3SG (...) see-SIM.CV stand-PRS.PTC
'Our dog is looking out from behind the pine tree.' (elicited)

This fact, in combination with the above-mentioned property that relational nouns cannot occur independently in the functions of grammatical subject or object, virtually rules out the possibility of encountering them in non-possessive marked nominative case, and thus as a bare stem. As it will turn out, this has consequences for the likelihood that these nouns will become reanalysed.

If the underlying stem were never encountered, one may ask why we assume that there is one at all. While the overwhelming majority of the relational noun stems is inflected, in Sakha the bare noun stem of certain relational nouns obtains in adverbial phrases. The nominative form of the previously referential noun has become lexicalised as an adverb, as is shown in example (5.11) for *kelin*.

(5.11) *Anï* *taŋas-tarïn* *hu:j-uoχ-χun* *na:da,* ***kelin***
now clothes-ACC.3PL wash-FUT.PTC-ACC.2SG have.to.R later
taŋas *hu:j-bat* *buol-but-tara* *χata.*
clothes wash -PRS.PTC.NEG AUX-PST.PTC-POSS.3PL MOD
'Now you have to wash their clothes, but later they stopped doing that, fortunately.' (XKM: 36)

This noun stem has lost its referential meaning of 'back part', but has acquired the adverbial meaning 'later'.

5.2.3.3. REGULAR STEMS AND UNSTABLE STEMS

In principle regular as well as unstable stems appear in both the referential noun class and the relational noun class. However, as pointed out above, a relatively high proportion of unstable stems occurs in the class of relational nouns, in particular in phrases of location. Before turning to the data for Dolgan, I will discuss in more depth some of the morphophonological rules in Sakha and how they affect the shape of regular and unstable stems.

5.2.3.3.1. REGULAR STEMS

As mentioned above, a noun in Sakha consists of a stem, potentially followed by suffixes for number, case, possession, or predication when the noun is used as a

nominal predicate. Consonant assimilation is a very widespread feature of the language, especially at morpheme boundaries (Stachowski & Menz 1998: 419), and may involve: a) the spread of some phonological features of consonant A to consonant B, leading to more similarity between them but retaining an acoustic boundary; or b) transfer of phonological features across consonants A and B such that these consonants come to share the same set of features, leading to gemination, or doubling, of the consonant. Scenario a) is exemplified in 5.12, where the underlying *-T* in the partitive case suffix *-TA* has become voiced under the influence of the preceding *r* in the stem *iŋir* (progressive assimilation).

(5.12) *iŋirde*
iŋir-TA
thread-PART
'(some) yarn'

In scenario b) two different consonants merge into a single long consonant, whereby the resulting geminated consonant takes the phonological features of the first consonant (progressive assimilation), the second consonant (regressive assimilation), or a subset of features from both (mutual assimilation).

Table 5.2: Assimilation processes in Sakha

Assimilation	Stem	Suffix	Assimilated form	Translation	Assimilation process
Progressive	*at*	*-LAr*	*at-tar*	'horses'	t →l = tt
Regressive	*ba:r*	*-LAr*	*ba:l-lar*	'they exist'	r ← l = ll
Mutual	*at*	*-Ga*	*ak-ka*	'to the horse'	t ⇆ g = kk

Some scholars propose that every geminated consonant in Sakha are eventually reduced to an assimilation process (Ubryatova 1982: 66). In this view assimilation is indisputable when double consonants appear at morpheme boundaries, and when they occur in the middle of a stem, they must be the result of assimilation between a stem and a suffix, or between two stems, in an earlier stage in the development of the language. In the course of time, they argue, the assimilated form has been reanalysed and become the new stem of the noun e.g. *oloppos* < *oloχ* + *mas* ['seat' + 'wood'] 'chair'. The consonants that can be geminated in Sakha are *p*, *t*, *k*, *l*, *m*, *n*, *ŋ*, *s*, *χ*, *č*. In theory, a geminated consonant can be ambiguous with

respect to the underlying combination of consonants it represents, due to the variety of assimilation processes that occur (e.g *attar* in Table 5.2 could theoretically be the result of *at+-lar* as well as of *at+-tar*). However, the high token frequency of the assimilated forms, the regularity of their formation, in combination with peoples' exposure to non-assimilated nominative forms, make recognition of the underlying form in most cases an unambiguous task.

5.2.3.3.2. UNSTABLE STEMS

The unstable stems change more significantly under the influence of the suffixes that are attached to them. More specifically, the category of unstable stems discussed here contains bisyllabic nouns with phonological structure (C)V -CV$_{high}$C. That is, the first syllable has an optional onset, a nucleus that is unspecified for frontness, backness or length and it has no coda. The second syllable of these nouns always has an onset consonant, a nucleus consisting of a high vowel and a coda of one consonant. Examples are the aforementioned stem *kelin* 'back part', as well as *tumus* 'beak', *ürüt* 'top side' and *alïn* 'bottom side'. When a suffix is attached to certain noun stems of this type, the high vowel in the final syllable is dropped and the consonants that are consequently adjacent undergo the same assimilation processes as discussed for the regular stems above[3].

In cases where the stem ends in a consonant and the added suffix begins with a vowel, it is attached to the formatted stem (which now ends in a consonant cluster), without further modification. For example, *kenne* [*kelin-(t)A*, back.part-POSS.3SG] 'its back part' consists of a stem *kelin* and a third person possessive suffix *-(t)A*. Since *kelin* is an unstable stem, the high vowel in the final syllable is dropped and the adjacent *l* and *n* undergo assimilation, resulting in a new stem *kenn*. The *t* in the third person posseessive suffix is optional and is only inserted if the preceding stem ends in a vowel. Since this is not the case here, only the low vowel *e* (represented by capital *A* in the underlying form according to Turkic tradition) is added to the stem, resulting in a surface form *kenne*.

[3] This rule also applies to certain verb forms with this phonological structure (e.g. *hïrït* 'to travel', or the passive on *-lIn*), but they will not be discussed here since with respect to these forms Dolgan does not behave differently from Sakha. This could be due to the fact that for verbs, the stem is identical to the imperative form, and thus occurs in discourse quite frequently.

On the other hand, in cases where the suffix begins with a consonant, an epenthetic high vowel (represented as capital *I*) is inserted between the formatted stem and the suffix. To continue with the same stem as before, the third person plural possessive form of *kelin* is formed by adding the suffix *-LArA* to the stem *kenn*. However, in this case an epenthetic vowel is inserted between the stem and the suffix, possibly to avoid too large consonant clusters and make the inflected form easier to pronounce and to parse. This results in the surface form *kennilere* [*kelin-I-LArA*, back.part-EP-POSS.3PL] 'their back part' consisting of *kelin*, an epenthetic vowel *I*, and a third person possessive suffix *-LArA*.

While this may seem many words spent on a small morphophonological detail, it will become clear that this epenthetic vowel has had important consequences for the current shape of unstable stems in Dolgan. Since most suffixes for nominal inflection begin with a consonant, the occurrence of epenthetic vowels with unstable stems in discourse is very high, and as will be shown in Section 5.2.4.3 this may explain for a certain group of reanalysed stems in Dolgan why they end in a high vowel. Since the assimilation process in unstable stems affects the consonant in the middle of the stem as well as those at the stem-suffix boundary, its consequences are more dramatic than in regular stems and determination of the phonological form of the underlying stem becomes less straightforward. However, it is important to note that only a subset of approximately 20 words with this phonological structure behave as unstable stems (see Tables 5.4 and 5.10). Other lexical items such as *kulun* 'foal' or *huruk* 'letter' have a regular stem and although every account of Sakha mentions this phenomenon as a fact (e.g. Stachowski and Menz 1998: 420), it is not quite clear which factors determine whether a stem is stable or not.

Table 5.3 illustrates the different stages of stem modification in unstable stems for the three situations referred to above: for an unstable stem followed by a suffix starting with a vowel, an unstable stem followed by a suffix starting with a consonant, and a stable stem of the same phonological structure, which does not undergo any change. The first column shows the underlying form of the stem and the suffix, for which optional consonants (as in *-(t)IgAr*) are omitted for the purpose of clarity. The second column shows the form of the word that we would expect to find if the stem were regular and assimilation processes applied as they normally do. Column three displays the crucial characteristic of unstable stems and shows the form of the word after the short high vowel in the last syllable has

been dropped. In column five the word is displayed in its actual shape, after it has undergone the assimilation process mentioned in column four.

Table 5.3: Assimilation processes in unstable stems

1 Components	2 Expected form	3 Vowel drop	4 Ass. process	5 Ass. form	6 Translation
ürüt + IgAr above + DAT.3SG	**ürütüger*	**ürtüger*	progressive	*ürdüger*	'over him'
murun + m nose + POSS.1SG	**murun-u-m*	**murn-u-m*	regressive	*munn-u-m*	'my nose'
kulun + IgAr foal + DAT.3SG	*kulunugar*	-	-	-	'to the foal'

Armed with this knowledge about formal properties of unstable stems in Sakha, 5.2.4 explores the differences in form and in use between Dolgan and Sakha. 5.2.4.1 focuses on relational nouns, and 5.2.4.2 does the same for referential nouns.

5.2.4. DOLGAN EQUIVALENTS TO THE SAKHA FORMS

5.2.4.1 RELATIONAL NOUNS

As was foreshadowed in Section 5.2.1, in Dolgan the unstable stems have been reinterpreted in such a way that the assimilated Sakha stem now serves as the root, thus eliminating irregularities due to stem change from the inflectional paradigm. For easy comparison of the forms and their use in Dolgan and Sakha, an additional set of examples is given in 5.13 and 5.14. Here, the form *ürdütünen* [*ürdü-(t)InAn*, upper.part-INST.3SG] 'on top of' is the Dolgan alternative to what in Sakha is *ürdünen* [*ürüt-(t)InAn*, upper.part-INST.3SG], clearly displaying the difference in underlying stem.

DOLGAN

(5.13) *on-tu-ŋ* ***ürdü-tünen*** *ot-tor-u*
that-DER-POSS.2SG upper.part-INST.3SG grass-PL-ACC
bïrag-al-lar, *ulaχan* *buruo* *kel-ien*
throw-PRS.PTC-PRED.3PL big smoke come-FUT.PTC.ACC.3SG
'On top of that they throw grass, so that there will be much smoke' (ESB: 71)

SAKHA

(5.14) *mannïk* *üs-tü:-nen* *oɣoruot* ***ürdünen***
in.this.way three-DISTR-INST fence upper.part.INST.3SG
kötö *hïdʒdʒar* *uonna* *ahïlïgar* *kele* *turar*
fly-SIM.CV go-PRS.PTC and food come-SIM.CV stand-PRS.PTC
'It jumped over a fence of three bars in this way and came to eat.'
(PYaI: 109)

While the difference between the inflected forms is obvious enough, the most compelling evidence that the assimilated stem has become the new root in Dolgan is provided by the fact that these stems occur as such in the unmarked nominative case, as in example 5.15. In this example, it is clear that *kenni* (as opposed to the Sakha stem *kelin*) fulfils the function of a referential noun instead of a relational noun, because it is followed itself by a relational noun *diek* 'side'. *Diek* 'side' has been claimed by some scholars to have grammaticalised into an enclitic particle of direction, or even a case form (Ubryatova 1985: 125), while others say it has the status of a noun that can be used independently or as a postposition. The comparative example (5.16) shows that *diek* (or its allomorph *diet*) normally combines with referential nouns like *mas* 'wood, forest' in the unmarked nominative case, in other words, with the bare stem of a noun. According to such an analogy, *kenni* must also be analysed as the unmarked nominative case, and thus as the underlying stem of inflected forms such as *kennitiger* (dative) and *kennititten* (ablative).

DOLGAN

(5.15) ***kenni*** *diek* *kör-dök-püne* *možet* *er-bin*
back.part side look-COND-COND.1SG can.R man-ACC.1SG
gïtta *araχ-s-an* *kel-iem*
with leave-RECP-SQ.CV come -FUT.1SG
'If I look back I might get divorced from my husband and return.' (APC: 95)

(5.16) *Didipte* *di-en* *üreχ* *ïra:χ,* ***mas*** *diet* *ba:r*
Dudypta call -SQ.CV river far wood side EXIST
tüörduon *kilometer*
four.ten kilometer.R
'The river Dudypta is far, forty kilometers to the south (lit.: in the direction of the forest).' (ANS: 12)

This kind of stem change is not restricted to just the few stems mentioned so far. Table 5.4 gives an overview of other relational nouns that have undergone reanalysis. The first column in the table specifies the language, the second column presents for both languages an inflected form (in this case the third person possessive form of the dative) to illustrate how the unstable stems occur most frequently in actual discourse, the third column shows in bold the underlying stems for both Sakha and Dolgan, and the translation is given in the fourth column.

Table 5.4: Relational nouns in Sakha and their Dolgan equivalents

Language	Dative	Stem	Meaning
Sakha	*ürdüger* *ürüt -(t)IgAr*	***ürüt***	'top side'
Dolgan:	*ürdütüger* *ürdü -(t)IgAr*	***ürdü***	
Sakha:	*annïgar* *alïn -(t)IgAr*	***alïn***	'bottom side'
Dolgan:	*annïtïgar* *annï -(t)IgAr*	***annï***	
Sakha:	*inniger* *ilin -(t)IgAr*	***ilin***	'front side'
Dolgan:	*innitiger* *inni -(t)IgAr*	***inni***	
Sakha:	*kenniger* *kelin -(t)IgAr*	***kelin***	'back part'
Dolgan:	*kennitiger* *kenni -(t)IgAr*	***kenni***	
Sakha:	*onnugar* *orun -(t)IgAr*	***orun***	'place'
Dolgan:	*onnutugar* *onnu -(t)IgAr*	***onnu***	
Sakha:	*ardïgar* *arït -(t)IgAr*	***arït***	'space between'
Dolgan:	*ardïtïgar* *ardï -(t)IgAr*	***ardï***	
Sakha:	*attïgar* *atïn -(t)IgAr*	***atïn***	'place next to'
Dolgan:	*attïtïgar* *attï -(t)IgAr*	***attï***	

Sakha:	*ördüger* *örüt -(t)IgAr*	***örüt***	'side'
Dolgan:	*öttütüger* *öttü -(t)IgAr*	***öttü***	

In all these cases, the Dolgan stem differs from the Sakha stem in a similar way, suggesting that the change in all these items was brought about by a single underlying process. A possible pathway for this development, in which surface ambiguity of the inflected Sakha form plays a central role, is illustrated in Table 5.5. It shows how the aforementioned Sakha form *ürdünen* [*ürüt-(t)InAn*, top.side-POSS.3SG] may have come to correspond to Dolgan *ürdütünen* [*ürdü-(t)InAn*, upper.part-INST.3SG].

Table 5.5 Possible analyses of ürdünen in Sakha and Dolgan

SAKHA		Possible analyses	DOLGAN	
Stem	INST.3SG		INST.3SG	Stem
ürüt	*ürdünen*	*ürüt -(t)InAn*		
'top side'	top.side .INST.3SG	top.side -INST.3SG		
		ürdü -(I)nAn	*ürdü-tünen*	*ürdü*
		top.side -INST	top.side - INST.3SG	'top side'

In this table the crucial column is headed 'possible analyses' as it shows that the Sakha surface form *ürdünen* is ambiguous with respect to its underlying structure: on the one hand, a hearer could understand this form as consisting of a stem *ürüt* and an instrumental case suffix, which in theory could be the possessive marked form *-(t)InAn*, as well as the non-possessive instrumental case suffix *-(I)nAn.* However, for a native Sakha speaker, the primary understanding of this form would probably be a possessive marked form for the following reason. Since *ürüt* is a relational noun, it occurs primarily in locational *izafet* constructions, in which the head noun is always marked for possession. Although in this particular form the surface structure is ambiguous with respect to the presence of possessive marking, analogy with other (regular) stems, in which the possessive marking is clearly audible, renders this interpretation for native speakers most likely. This is why the possessive marked underlying form is presented as the first option in the table.

However, from a purely structural, point of view, *ürdünen* is more straightforwardly analysed as a stem *ürdü* and a non-possessive

instrumental *-(I)nAn*. While this interpretation may be less likely for native adult speakers who have in-depth knowledge of the entire linguistic system and its irregularities, such an analysis is easy to imagine for second language learners of Sakha, as well as for young children, who are trying to parse new language forms. After all, the second analysis is much more transparent (since there would be no assimilation involved), economical (since no rules are needed for the inflection of irregular stems) and thus more plausible than the first. Therefore, it is easy to imagine that after having concluded that *ürdü* is a noun stem in Sakha, L2 speakers as well as infants store this form in the mental lexicon. At a later stage, when speakers comprehend that phrases of location involve possessive marked nouns, and in analogy with other possessive marked forms, the suffix *-(t)InAn* is added to the stem *ürdü*, resulting in the innovative form *ürdütünen* that is found in Dolgan today. In small children, such deviant interpretations are typically overruled by the standard usage in the Saha-speaking community. Children will adjust their analysis simply by being exposed to the every-day input of standard Sakha forms or they may be corrected. However, this is not necessarily so for adult second language speakers, and if their number is large enough, or their input of standard Sakha too low, there is a possibility that the deviant interpretation takes root in the L2-speaking community and may even spread among L1-speakers too (see Section 3.1.4 for a detailed review of this scenario).

5.2.4.1.1 VARIATION BETWEEN DOLGAN AND SAKHA STEMS

While the data presented above show that a difference between Dolgan and Sakha in the form of these unstable stems is undeniable, the story would not be complete without mentioning the fact that occasionally the Sakha stems are encountered in Dolgan as well. This variation holds only one way, however: while the Sakha stems are sometimes found in Dolgan, the Dolgan stems are never found in standard Sakha. This is illustrated by the Dolgan examples 5.17 and 5.18, which suggest that the two forms can be used interchangeably. Although the relational noun in 5.17 refers to location in space and location in time in 5.18, other examples from the corpus show that this criterion does not play a role in the choice between *kelin* or *kenni*.

DOLGAN

(5.17) *onton* *ol* *dʒaχtar-ïŋ* *bolog-un* ***kennitten***
then that woman-POSS.2SG balok-ACC.3SG back.part.ABL.3SG
buo *gü:le* *boloχ* *ba:r*
PRT hall balok EXIST
'Then behind the women's balok there is the hall balok' (IMA: 34)

(5.18) *(...) ol* *tühe:-bit-im* ***kenni-titten*** *ke* *kim-iŋ,*
(...) that dream-PST.PTC-POSS.1SG back.part-ABL.3SG CONTR who-POSS.2SG
ol *ös* *iste-bin* *buo*
that story hear.SIM.CV-PRED.1SG PRT
'Well after my dream, ehm, I hear that story' (TJP: 126)

Thus, both stems are acceptable in Dolgan, but they certainly do not occur with the same frequency. The picture is dominated by frequent use of Dolgan stems, supplemented by occasional Sakha stems for a small set of relational nouns, such as *kelin* and *ürüt*. This statement is based on a frequency analysis of relational noun stems in my Dolgan corpus, in which I determined for each noun its overall frequency as well as the number of underlying Sakha stems and Dolgan stems. In this context it is important to note that the underlying stem can only be determined with certainty for a limited number of forms. More precisely, these are the unmarked nominative case (i.e. the stem), and for other cases the third person possessive form. Case forms marked for other persons as well as non-possessive case forms are ambiguous with respect to their underlying stem. As was explained in Section 5.2.3.3.2 it is impossible to determine on the basis of these forms alone whether the high vowel in the middle of the word belongs to the stem (as would be the case when the Dolgan stem is used) or whether it is the epenthetic vowel that is added in Sakha stems between the stem and suffixes that start with a consonant. This is visualised in Table 5.6, in which the possessive paradigm is shown for the dative of *kelin*/*kenni*, in addition to the nominative form. The forms for which the stem is unambiguous are put in a box. Since the unmarked nominative and the third person singular possessive forms are the only unambiguous forms, the main focus in the discussion of the data will be on those, and the inflectional nouns marked for other persons will play only a marginal role.

Table 5.6: Nominative case and possessive paradigm of dative case kelin/kenni

	Surface form		Underlying form	Underlying form
	Sakha	Dolgan	Sakha	Dolgan
NOM.	*kelin*	*kenni*		
DAT.1SG	*kenniber*		*kelin -**I**- BAr*	*kenn**i** -BAr*
DAT.2SG	*kenniger*		*kelin -**I**- GAr*	*kenn**i** -GAr*
DAT.3SG	*kenniger*	*kennitiger*	*kelin -(t)IgAr*	*kenn**i** -(t)IgAr*
DAT.1PL	*kennibitiger*		*kelin -**I**- BItIgAr*	*kenn**i** -BItIgAr*
DAT.2PL	*kennigitier*		*kelin -**I**- GItIgAr*	*kenn**i** -GItIgAr*
DAT.3PL	*kennileriger*		*kelin -**I**- LArIgAr*	*kenn**i** -LArIgAr*

The question addressed here only concerns the proportion of Sakha stems with respect to Dolgan stems, therefore spontaneous data as well as elicited data are included in the analysis. The combination of these two sources provides more specific data than would spontaneous speech alone, and does not reduce the reliability of the results, since it is unlikely that text genre influences the choice of noun stem. An overview of both stems in Dolgan is given below.

Table 5.7: Proportion of Sakha stems and Dolgan stems in Dolgan relational nouns

Meaning	Sakha stem	Dolgan stem	Ambiguous	Total
'back part'	***kelin***	*kenni*		
	47.3% (32)	40.4% (23)	12.3% (7)	62
'top side'	***ürüt***	*ürdü*		
	30% (6)	60% (12)	10% (2)	20
'space between'	***örüt***	*öttü*		
	28.5% (2)	57.1% (4)	14.3% (1)	7
'bottom side'	***alïn***	*annï*		
	13.3% (2)	80% (12)	6.7% (1)	15
'front side'	*ilin*	*inni*		
	0%	88.9% (8)	11.1% (1)	9
'place next to'	*atïn*	*attï*		
	0%	100% (8)	0	8
'place'	*orun*	*onnu*		
	0%	100% (2)	0	2
Total	42	69	12	123

In Table 5.7 the relational nouns are ranked by occurrence of Sakha stems in decreasing frequency. The most striking observation from this table is that those relational nouns, for which a considerable number of Dolgan and Sakha stems are used, are also the most frequent relational nouns in the Dolgan corpus, with the exception of *örüt*, for which the total number is only seven. In other words, there is a correlation between the presence of a Sakha stem for a relational noun in Dolgan and its overall frequency of occurrence. This claim holds less so for *örüt*, for which the overall number is relatively low but the proportion of Sakha stems quite high, and *alïn*, for which the total number is not greatly different from *ürüt* but the proportion of Sakha stems is much lower. However, the correlation for *ürüt* and even more so for *kelin* is so blatant that it is unlikely to be due to chance and therefore requires further investigation. The numbers in this table are based on the frequency of use in Dolgan only, but they do raise the question what the frequency of these relational nouns is in Sakha discourse, and whether the current selection of Sakha stems in Dolgan may be explained by a high frequency of their equivalents in Sakha.

The reason for this hypothesis is the idea that highly frequent items are less likely to undergo a permanent change, since speakers have regular exposure to these forms (Bybee 1991: 72-73). Within this context, frequently used items in Sakha (in this case relational nouns) are more likely to retain their Sakha shape in Dolgan than infrequently used items. The high exposure to these items makes them more likely to be conceived of, and stored in memory as, unanalysable units (like proper postpositions) instead of nouns consisting of a stem and a variable case suffix. Frequency here includes text frequency (the number of times a certain stem occurs in the corpus) as well as paradigmatic frequency (the number of slots a certain stem fulfils in the inflectional paradigm). The data from Sakha that will be presented below suggest that both text frequency and paradigmatic frequency influence the distribution of Sakha stems in Dolgan.

Investigation of the Sakha corpus shows that of all relational nouns *kelin* 'back part' stands out as the most frequent one in Sakha oral speech. *Kelin* and its inflected forms with assimilated stems make up for 25.8% of all relational nouns in the Sakha corpus (42 out of 163). To compare, the second most frequent relative noun is *örüt* 'space between' with 14.7% (24 instances). The high text frequency of *kelin* in Sakha is mirrored in Dolgan speech, where its equivalent occupies 51.6% of all tokens of relational nouns (33 out of 64). This number is based on only the spontaneous texts for both languages, and includes all relational nouns. Two

inflected forms are particularly frequent in Sakha, namely the ablative form *kennitten* [*kelin-(t)IttAn,* back.part-ABL.3SG] 'from behind' and the possessive marked third person singular *kenne* [*kelin-(t)A*, back.part-POSS.3SG] 'after'. Each of these forms constitutes 38.1% (or 16 in absolute numbers) of all occasions of *kelin* in Sakha. Interestingly, it is exactly these forms, which are found in this shape (i.e. their Sakha shape) in Dolgan.

To support this claim, a comparison of frequencies in Sakha and Dolgan is given below. Table 5.8 presents the frequency distribution in Sakha and Dolgan of the forms with an unambiguous Sakha or Dolgan stem, i.e. the third person possessive forms and the unmarked nominative. Since Dolgan has the option of using both stems, separate columns are created for forms with an underlying Dolgan stem and those with an underlying Sakha stem.

Table 5.8: Frequency distribution of inflectional forms of kelin/kenni *'back part' in Dolgan and Sakha*

Infl. cat.	SAKHA	No.	DOLGAN Dolgan stem	No.	Sakha stem	No.
Nom.	*kelin* back.part	6	*kenni* back.part	2		
Nom. 3SG	*kenne* *kelin-(t)A* back.part-POSS.3SG	16			*kenne* *kelin-(t)A* back.part-POSS.3SG	9
Dat. 3SG	*kenniger* *kelin-(t)IgAr* back.part-DAT.3SG	1	*kennitiger* *kenni-(t)IgAr* back.part-DAT.3SG	2		
Abl. 3SG	*kennitten* *kelin-(t)IttAn* back.part-ABL.3SG	16	*kennititten* *kenni-(t)IttAn* back.part-ABL.3SG	3	*kennitten* *kelin-(t)IttAn* back.part-ABL.3SG	5
Adj.	*keliŋŋi* *kelin-GI* back.part-ADJZR	2				

The table strikingly confirms the idea that Dolgan speakers have only preserved the Sakha version of the forms that occur most frequently in Sakha. This is most clearly illustrated by *kenne*, which is one of the two most frequently occurring forms in Sakha. The table shows that *kenne* occurs in Dolgan rather frequently as well, and most importantly, it exists only in this form. The expected innovative

form **kennite* [*kenni-(t)A*, back.part-POSS.3SG] is not attested in the corpus at all, despite the fact that the nominative form, and thus the underlying stem in Dolgan is *kenni*. This supports the idea that frequently used forms may be stored in the brain as a single unit and thus less prone to change, as is argued for example in Bybee (1991: 77).

Additional evidence, though slightly less stringent, comes from the third person singular ablative form *kennitten*. As for *kenne*, this form with an underlying stem *kelin* is very common in Sakha, and appears in Dolgan in exactly this form as well. However, parallel to this Sakha-based form, Dolgan speakers also use *kennititten*, based on the Dolgan stem *kenni*. This suggests that for some forms two stems are available, which may create confusion regarding the ultimate underlying form for the inflectional paradigm. However, this confusion is unnecessary, if one adheres to the idea that highly frequent forms can become stored as unanalysable units in the mental lexicon. In that case the speaker does not conceive of items such as *kenne* and *kennitten* as consisting of a stem (*kelin*) and suffix (*-(t)A* or *-(t)tAn*), but they would exist as fossilised forms in the mental lexicon. Consequently, it is not necessary to assume an underlying stem *kelin* for these forms, and their existence does not clash with the existence of forms like *kennitiger* and *kennititten*, which are clearly based on an underlying stem *kenni*. These less frequent forms are constructed with the assimilated stem through a productive process of stem + case suffix. This is exemplified by the much less frequent form *kenniger*, of which there is only one instance in Sakha, and which in Dolgan consistently occurs as *kennitiger*, based on the Dolgan stem *kenni* + *-(t)IgAr* [back.part + -POSS.3SG]. Needless to say, it remains impossible to look into the speakers head and leaf through their mental lexicon, but these data suggest that reanalysis of the Sakha stem *kelin* has been completed in Dolgan, resulting in the employment of forms based on the Dolgan stem for the infrequent case forms, while forms based on the Sakha stem (*kenne* and *kennitten*) are lexicalised Sakha 'islands', which show the remnants of an earlier stage in the development of the language.

A similar trend, although less pronounced, and less reliable due to the lower number of occurrences, applies to *ürüt* 'top side'. The left half of Table 5.9 shows the frequency distribution of the relevant inflectional forms of *ürüt* in Sakha, which is clearly dominated by dative and instrumental case forms. The right half of the table displays the occurrence of this relational noun in Dolgan and one can see once again that Sakha stems in Dolgan correspond to the case forms that occur

most frequently in Sakha (dative and instrumental). For *ürdüger* one may argue that it is ambigous with respect to its underlying structure. After all, *ürdüger* could be analysed as *ürüt-(t)IgAr* [top.side-DAT.3SG] or as *ürdü-GAr* [top.side-DAT.2SG]. However, the discourse context is sufficient to disambiguate this form unequivocally as a dative in the third person possessive and not as a second person, and therefore it can be confidently listed under Sakha stems. Although overall numbers are small (13 for Sakha and 10 for Dolgan), and there is not all that much variation in the Sakha forms either, the results in Table 5.8 at least do not contradict the claim made for *kelin* above, and the more frequently used forms in Sakha are also preserved in Dolgan.

Table 5.9: Frequency distribution of inflectional forms of ürüt/ürdü *in Dolgan and Sakha*

Infl. cat.	SAKHA	No.	DOLGAN Dolgan stem	No	Sakha stem	No.
Nom.			*ürdü* *top.side*	3		
Dat. 3SG	*ürdüger* *ürüt-(t)IgAr* top.side-DAT.3SG	6	*ürdütüger* *ürdü-(t)IgAr* top.side-DAT.3SG	1	*ürdüger* *ürüt-(t)IgAr* top.side-DAT.3SG	4
Inst. 3SG	*ürdünen* *ürüt-(t)InAn* top.side-INST.3SG	7	*ürdütünen* *ürdü-(t)InAn* top.side-INST.3SG	1	*ürdünen* *ürüt-(t)InAn* top.side-INST.3SG	1

The data for *örüt* and *alïn* are even sparser and therefore they cannot be discussed in great detail. Nevertheless, it is worth mentioning that the Sakha stem *örüt* does occur in Dolgan, despite its relatively low overall occurrence in Dolgan discourse. Importantly, the presence of this Sakha stem needs to be viewed agains the knowledge that *örüt* is the second most frequent relational noun in Sakha spontaneous speech (see Table 5.7). In addition, a closer look reveals that the Sakha stem used in Dolgan is a possessive marked dative form *öttüger*, a form which constitutes one third of all the occurrences of this stem in Sakha (29%), and is thus encountered regularly. Although one instance in Dolgan is clearly no basis from which to draw any definite conclusions, it does provide additional support to the general idea that common forms in Sakha are kept in their original form in Dolgan.

Thus, the data from my spoken Dolgan corpus suggest that only the relational nouns *kenni*, *ürdü* and *öttü*, *annï* can occur with both Sakha and Dolgan stems, but

this corpus is not necessarily exhaustive. The overall number of the other relational nouns is rather small and so the absence of Sakha stems may be due to chance. Although a dictionary is not the most reliable source for solving this issue, it may provide supplementary evidence; and indeed, Stachowski's work confirms the presence of both the Dolgan and Sakha stem for the three relational nouns *kenni*, *ürdü* and *annï*, including an additional Sakha stem *ilin* for Dolgan *inni*, which does not occur in my corpus. Stachowski's dictionary and my corpus data agree that for *attï* 'place next to', *onnu* 'place' and *öttü* 'space between' (with one exception), only the Dolgan stem is used. The current variation in use of at least a subset of the relational noun stems suggests that we are witnessing a process of on-going change, which (still) allows both stems to be used, rather than a completed change in the language.

Other factors besides frequency that may condition or restrict the variation in use of noun stems in Dolgan are speaker age or geographical location. Age could affect the distribution of stems if one assumes that an on-going change is most likely to be promoted by the younger generation. In that case one would expect a skewed distribution, with the Dolgan stems occurring more frequently in the younger age groups than in the older generations. However, investigation of the corpus shows that age does not play any role in the distribution of the stem variants. Both stems are used by speakers of all age categories, and without any significant differences in frequency of use.

With respect to geographical location one would expect the people in villages closer to the Sakha border (Syndassko) to use more Sakha stems than the villages further away as a result of regular interaction with Sakha speakers. However, the current data do not explicitly support this expectation. Of the 42 Sakha stems, 18 were produced in Syndassko, the village closest to the Sakha border, 18 in Kheta further to the west and 6 in Volochanka, even more distant from the Sakha border. Thus there is no indication that Sakha stems cluster in the border areas where contact with Sakha is most frequent.

While conclusions with respect to the role of frequency must be drawn with care due to the relatively low number of occurrences in the text corpus, these data provide evidence for a significant role of discourse and paradigmatic frequency in the retention of certain Sakha stems in Dolgan. The skewed distribution of forms in Sakha (i.e. many ablatives and third person possessive forms in the case of *kelin*) may be part of the explanation for the observed correlation between stem type and case form in Dolgan. In other words, it may motivate why exactly these Sakha

forms, and not others, have been preserved in Dolgan, whereas for other forms only the Dolgan stem is used. This retention of Sakha forms does not seem to correlate with other factors such as age or geographical location, and thus stresses the importance of discourse frequency in language variation and change.

5.2.4.2 REFERENTIAL NOUNS

While relational nouns constitute a considerable proportion of unstable stems in Sakha, the observed phenomenon is not restricted to this category of nouns alone. Table 5.10 displays a number of unstable referential noun stems in alphabetical order that have undergone a similar type of morphophonological change as that described for relational nouns. It shows that the noun stems, which have the same phonological structure as the stems discussed above ((C)V-CV$_{high}$C), have lost their final high vowel and have undergone consonant assimilation just like the relational nouns.

The list in Table 5.10 is not exhaustive, since it only shows the nouns that occur in my corpus. Voronkin (1999) lists a few more lexical items, but since he in fact describes variation in the dialects of Sakha, I do not want to presuppose the existence of all these words in Dolgan without having checked this explicitly. It is striking that more than half of these lexical items consists of body part or kinship terms. This is probably no coincidence, since members of both semantic fields are intrinsically linked to an owner, in a literal sense (body parts) or in a figurative sense (kinship) and therefore are more likely to occur with a possessive suffix than without one. Thus, an unmarked nominative form like Sakha *murun* 'nose' is rarely encountered in spontaneous Sakha texts. More common are possessive marked forms like *munnum* [*murun-(I)m*, nose-POSS.1SG], 'my nose', *munnuŋ* [*murun-(I)ŋ*, nose-POSS.2SG], 'your nose', or a possessive marked case form such as *munnubar* [*murun-I-BAr*, nose-EP-POSS.1SG] 'on my nose', which are ambiguous with respect to their underlying stem in a similar fashion as described for relational nouns: *munnum* can be divided up as represented above, where the high vowel in the final syllable is an epenthetic vowel, but from the surface this form could equally well have an underlying structure *munnu-(I)m*, where the high vowel belongs to the stem. Given the higher frequency of the ambiguous possessive marked forms when compared to the non-possessive and non-assimilated forms, and given the opacity of the relation between the surface form and the underlying stem due to the

morphophonological rules, Dolgan speakers may have developed different assumptions with respect to morpheme boundaries in these forms. A possible pathway of this reanalysis, similar to the principle illustrated in Table 5.5 is given in Table 5.10.

The left half displays forms derived from the Sakha stem, the right half shows forms derived from the Dolgan stem and the column 'possible analyses' reveals the two underlying morpheme structures that a hearer could infer from the Sakha input. The stem and third person possessive form are displayed for both languages because these are the only forms in which the stem can be unambiguously determined, and therefore they show most clearly the difference between Dolgan and Sakha. The first person possessive form has been included as an example of an ambiguous form, which could have triggered the reanalysis.

Table 5.10: Potential pathway of reanalysis of referential nouns in Sakha and Dolgan

SAKHA				DOLGAN			
Stem	Poss.3SG on *-(t)A*	Poss.1SG on *-(I)m*	**Possible analyses**	Poss.1SG on *-(I)m*	Poss.3SG on *-(t)A*	Stem	Translation
harïn	*hanna*	*hannïm*	*harïn-(I)m*				
			hannï-(I)m	*hannïm*	*hannïta*	*hannï*	shoulder
hürün	*hünne*	*hünnüm*	*hürün-(I)m*				
			hünnü-(I)m	*hünńüm*	*hünńüte*	*hünńü*	spinal cord
kïlïn	*kïnna*	*kïnnïm*	*kïlïn-(I)m*				
			kïnnï-(I)m	*kïnnïm*	*kïnnïta*	*kïnnï*	father in law
köγüs	*köχsö*	*köχsüm*	*köγüs-(I)m*				
			köksü-(I)m	*köksüm*	*köksüte*	*köksü*	back
murun	*munna*	*munnum*	*murun-(I)m*				
			munnu-(I)m	*munnum*	*munnuta*	*munnu*	nose
törüt	*tördö*	*tördüm*	*törüt-(I)m*	*tördum*	*tördö*	*törüt*	ancestor, root
			tördü-(I)m	*tördüm*	*tördüte*	*tördü*	clan, root
tumus	*tumsa*	*tumsum*	*tumus-(I)m*	*tumsum*	*tumsa*	*tumus*	cape, island
			tumsu-(I)m	*tumsum*	*tumsuta*	*tumsu*	protruding object, top

Since the unmarked nominative of body parts and kinship terms occurs so rarely in discourse, evidence from the spoken corpus is mainly based third person singular possessive forms on *-(t)A,* which yields *munna* [*murun-(t)A*, nose-POSS.3SG] 'his nose' in Sakha, but *munnu-ta* [*munnu-(t)A*, nose-POSS.3SG] 'his nose' in Dolgan.

DOLGAN:

(5.19)	*oh*	***munnu-ta***	*ńaltajan*	*kel-ie*	*di-en*
	PRT	nose-POSS.3SG	bring.near?-SQ.CV	come -FUT.3SG	say-SQ.CV

'Oh, he came very close with his nose.' (TJP: 14)

As can be seen from Table 5.10, for some nouns, such as *törüt* and *tumus*, both the Dolgan and the Sakha stems are used. However, according to the dictionary they have different semantic connotations. For example *tumus*, which has the meaning 'beak' in Sakha, is found in Dolgan as *tumus*, with the meaning of 'cape, island', and as *tumsu*, with the meaning 'top' or 'protruding object' more generally. While for *törüt* I have no evidence other than the dictionary, the semantic difference between *tumus* and *tumsu* are confirmed by data from my own corpus as well. However, the details of this semantic specialisation of cognate stems require further research.

5.2.4.3 EARLIER EXPLANATIONS IN THE LITERATURE

The difference in unstable stems between Dolgan and Sakha for relational as well as for referential nouns has been recognised by other scholars. Voronkin observes in his overview of Sakha dialects that "[i]n the northern dialects (more regularly in the northwestern dialects), a particular formation of the possessive forms is observed (about 20 nouns)" [4]. Note that in his account, Dolgan is viewed as a dialect of Sakha, but this point of discussion is not relevant for the treatment of the data here. He continues that there are different interpretations of the phenomenon.

Some authors (Voronkin 1980 in Ubryatova 1985) have analysed these forms as double possessive marking. Others (Ubryatova 1985: 187 and Voronkin 1999) explain the change as phonological metathesis. According to this account, a noun

[4] "В северных говорах (более регулярно в северо-западных) отмечается своеобразное оформление притяжательной формы (около 20 имён)." (Voronkin 1999: 140).

like Sakha *orun* 'place' would have developed into Dolgan *onnu* in the following way: *orun* (→ metathesis) *ornu* (→ assimilation) *onnu*. Voronkin argues that this account is more plausible because there is no need to include the mysterious double possessive marking, and because metathesis is rather common in other related languages, e.g. Tatar *boron* → *borno* 'nose'. While the process of metathesis itself may be plausible, it leaves unexplained why it would have happened in the first place. The account based on reanalysis that has been hinted at throughout the chapter, and that will be further explored in Section 5.4, builds on cognitive principles instead and gives insight into this development on a deeper level.

Stachowski (e.g. 1993: 144) also refers to the link between the Dolgan and the Sakha stems when he describes the etymology of Dolgan stems in his dictionary. However, he assumes that the Dolgan stems of this type are derived from the third person possessive form in Sakha, for which he assumes a fossilised ('erstarrt') possessive suffix of the form *-(t)I*. The high vowel in this suffix should account for the final high vowel we find in modern Dolgan stems. More precisely, Stachowski analyses the Dolgan stems as coming from a Sakha form which looks like NOUN-*(t)I*. This suffix, which he sometimes calls a 'third person marker' and sometimes a third person possessive marker', was attached to the noun stem, and after this suffixation consonant assimilation took place, as in e.g. *murun-u* > *munnu* 'nose', *sarïn-ï* > *hannï* 'shoulder' and *kelin-i* > *kenni* 'back part'. The fact that Stachowski assumes *-(t)I* as the third person singular possessive suffix is somewhat puzzling, because the current Sakha suffix for this category is unequivocally recognised in grammars as *-(t)A*. In contrast to the form proposed by Stachowski, this suffix contains a low vowel, and so do the forms that carry it, e.g.: *munna*, *hanna*.

There are two possible explanations for why Stachowski takes *-(t)I* as the underlying form, but as I will show, they are both unsatisfying to explain the observed difference between Dolgan and Sakha. On the one hand, Stachowski may have based his analysis on an idea he expressed elsewhere that *-(t)I* can be extracted as a third person singular marker from oblique case forms of the possessive declension in Sakha (Stachowski and Menz 1998: 422). According to this argumentation, possessive marked case suffixes in Sakha can be subdivided into a part that encodes possession and a part that encodes case. However, the form of these parts does not exactly match the forms the possessive suffixes have in their isolated form. For example, the third person possessive dative suffix *-(t)IgAr* is analysed as a third person possessive marker *-(t)I* and a dative case ending *-gAr,*

and an instrumental case as consisting of *-(t)I* and *-nAn.* To illustrate the motivation for this idea, an overview of all the third person singular case forms is given in Table 5.11. For comparison the case forms for the first person are also included, as well as the non-possessive case forms.

Table 5.11: Possessive and non-possessive case forms for third and first person singular (Stachowski and Menz 1998: 422).

CASE	NON-POSS.	POSSESSIVE					
		POSS. 1SG	Poss. suff.	Case suff.	POSS. 3SG	Poss suff.	Case suff.
Nom	-	*-Im*	*-Im*	-	*-(t)A*	*-(t)A*	-
Dat	*-GA*	*-BAr*	*-BA*	*-r*	*-(t)IgAr*	*-(t)I*	*-gAr*
Acc	*-(n)I*	*-BIn*	*-BI*	*-n*	*-(t)In*	*-(t)I*	*-n*
Part	*-TA*	*-BIna*	*-BI*	*-nA*	*-(t)Ina*	*-(t)I*	*-nA*
Abl	*-(t)tAn*	*-BIttan*	*-BI*	*-ttAn*	*-(t)IttAn*	*-(t)I*	*-ttAn*
Inst	*-(I)nAn*	*-BInAn*	*-BI*	*-nan*	*-(t)InAn*	*-(t)I*	*-nan*
Comit	*-LI:n*	*-BInA:n*	*-BI*	*-nA:n*	*-(t)InA:n*	*-(t)I*	*-nA:n*
Comp	*-TA:γAr*	*-BInA:γAr*	*-BI*	*-nA:γar*	*-(t)InA:γar*	*-(t)I*	*-nA:γar*

On the basis of this overview, the argumentation for *-(t)I* as a third person marker seems quite plausible, since it is a constant factor in all cases except the nominative. However, it is questionable whether this analysis is sufficient for a satisfactory analysis of *murun-(t)u > munnu*. One may ask whether speakers actually analyse complex morphemes such as *-(t)IgAr* or *-(t)InAn* as consisting of a person and a case component. More likely, these forms are processed, produced and stored as a single unit, which is supported by the phonological reduction of some possessive-marked case forms (e.g. Sakha *-(t)IgAr* → *-Ar*). The high level of analytical skill on the part of the speaker and consciousness of the internal morpheme structures make this explanation unattractive and implausible.

A more realistic explanation of the difference between Sakha *murun* and Dolgan *munnu* is provided by the fact that these forms are typically used in their possessive form in discourse whereby an epenthetic high vowel is inserted between the stem and the possessive (case) suffix. As for the relational nouns, the morphological structure of the resulting surface form is ambiguous for the hearer, and the epenthetic vowel can be analysed as part of the inflection, as is the case in Sakha, or it can be analysed as part of the stem, as is the case in Dolgan. Since this epenthetic vowel appears in every possessive marked case form except for the

third person, the assimilated stem followed by a high vowel has a high text frequency, which makes this form a suitable candidate for reanalysis. To illustrate this, the possessive paradigm for the stem *murun* is displayed in Table 5.12, showing the nominative, accusative and dative case. From this it is clear that only in the third person no epenthetic vowel is inserted. And even of those third person forms, only in the nominative case this results in a low final vowel.

Table 5.12: Possessive inflection for murun *'nose' (NOM., ACC., DAT.)*

	NOM	ACC	DAT
1SG	*munnum*	*munnubun*	*munnubar*
	murun-I-m	*murun-I-BIn*	*murun-I-BAr*
2SG	*munnun*	*munnugun*	*munnugar*
	murun-I-n	*murun-I-GIn*	*murun-I-GAr*
3SG	*munna*	*munnun*	*munnugar*
	murun-(t)A	*murun-(t)In*	*murun-(t)IgAr*
1PL	*munnubut*	*munnubutun*	*munnubitigar*
	murun-I-BIt	*murun-I-BItIn*	*murun-I-BItIgAr*
2PL	*munnugut*	*munnugutun*	*munnugutugar*
	murun-I-BIt	*murun-I-GItIn*	*murun-I-GItIgAr*
3PL	*munnnulara*	*munnularin*	*munnularigar*
	murun-I-LArA	*murun-I-LArIn*	*murun-I-LArIgAr*

Thus, the high paradigmatic and text frequency of the assimilated stem followed by a high vowel may explain why this sequence has been interpreted as a new stem in Dolgan.

On the other hand, Stachowski may have assumed that the vowel in the third person singular possessive suffix was high at some stage in the history of Sakha, and has changed into a low vowel over time. This is not unimaginable, since the third person singular possessive in other Turkic languages often contains a high vowel, e.g. *-(s)I(n)* in Old Turkic and Turkish (Erdal 1998: 142, Csató and Johanson 1998: 212), *-(s)I* in Kirghiz, Middle Kipchak, Azerbaijanian and Turkmen (Kirchner 1998: 347, Berta 1998: 160, Schönig 1998: 252, 264). However, for this to be a satisfactory explanation, the vowel change must have taken place in Sakha only after the Dolgan people diverged from the Sakha and their characteristic Dolgan speech had been established, since the Dolgan stems all contain a high vowel in the final syllable. In other words, the change in Sakha must have happened after the 17th century. Although there is very little data on Sakha that go back to the 17th

or 18th century (the earliest mention of the Dolgan people), old word lists show that the POSS.3SG suffix in Sakha already contained a low vowel in the beginning of the 19th century. Sauer's word list, which was compiled in 1803, contains a number of items, which testify that the vowel in the third person possessive was already low at that time. For example, there is the form *a:t-a*[5] [*a:t -(t)A*, name-POSS.3SG] 'his name' (Sauer 1803: 318) and *tï:n-a*[6] [*tï:n -(t)A*, breath-POSS.3SG] 'his breath' (ibid.: 7). An additional problem of such an analysis is the fact that every other third person possessive form in Dolgan ends in a low vowel, just as in modern-day Sakha. If the above scenario were true and the third person possessive vowel in Sakha was lowered only after Dolgan diverged from Sakha, it is hard to explain why Dolgan has not retained a high vowel in other third person singular possessive forms, in particular since Dolgan is assumed to have preserved archaic aspects of Sakha in other domains. These two arguments render an account of the Dolgan stems that is based on a third singular possessive form with a high vowel highly unlikely, even if this final vowel were high in Sakha at an earlier stage.

To summarise, Section 5.2 has shown that relational nouns as well as referential nouns which in Sakha have the phonological structure (C)V-CV$_{high}$C, often have a different shape in Dolgan, namely (C)VC-CV$_{high}$. For both relational and referential nouns we have seen that the Sakha stems can also sometimes be used in Dolgan, although the Dolgan stems occur in the majority of cases. In the case of relational nouns it was postulated that this is connected to frequency of use in discourse, where particularly common forms in Sakha (such as *kenne* and *kennitten*) may have become stored as unanalysable units in the speaker's mental lexicon as a result of high input frequency. Some referential noun stems can also occur in two versions (such as *tumus* 'cape'). In this case each stem variant has come to emphasise different aspects of the Sakha meaning, and nowadays they occur in different contexts. The extent to which the development of different meanings influenced the retention of the two stems needs to be investigated in more detail, however.

This difference was explained through the process of reanalysis, whereby Sakha forms that from a hearer's perspective have an ambiguous morphological structure, were divided up in a different way by speakers of Dolgan. While this cannot determined with absolute certainty, for relational nouns it is most likely

[5] Original transcription: *aatta* (Sauer 1803: 318).

[6] Original transcription: *tina*, translated as 'ghost, soul' (lit.: Geist, Seele) (Sauer 1803: 320).

that the high final vowel in the Dolgan stem corresponds to the high vowel that appears in the possessive case suffix of the third person singular. Since relational nouns as a default figure in the *izafet* construction of locational phrases, and often have third person referents, this form occurs most frequently in discourse and is therefore most likely to have served as a basis for this reanalysis. For the referential nous, the most plausible origin of the final high vowel was argued to be the epenthetic vowel that is inserted between stem and suffix in all possessive forms but the third person singular[7].

5.3. REGULARISATION OF THE PARADIGM OF AUXILIARY VERB *E-* 'BE'

5.3.1 DESCRIPTION

Another instance of paradigm regularisation in Dolgan is the declension of the defective auxiliary verb *e-* 'be'. Like in Sakha, this verb is only used in the past tense in Dolgan, and is employed as: a) an auxiliary verb to form analytical past tense forms (resultative past, the imperfective past and the remote past); and b) as a copula with nominal predicates. Both uses are illustrated in example 5.20.

(5.20) *hoγoto:χ* *kï:s* ***e-ti-m*** *buo* *kergetter-ber,*
single girl be-PST-POSS.1SG PRT family.PL-DAT.1SG
χahan *da* *giniler-ten* *araχ-pataχ* ***e-ti-m***
when NEG 3PL-ABL leave-PST.PTC.NEG be-PST-POSS.1SG
'Well, I was a single child for my parents, I had never been separated from them' (LKS: 24)

In normal verbs such as *bar* 'go', most person-number forms of the recent past are formed according to the scheme: STEM+-*TI*+POSS, whereby -*TI* is the recent past suffix, and POSS the possessive person-marking suffix which agrees with the subject. However, the third person deviates from this scheme in the singular and plural. A regular formation of -*TI*+POSS for the verb *bar* 'go' would look like **bardïta* [*bar-TI-(t)A*, go-PST-POSS.3SG.] for the singular, and **bardïlara* [*bar-TI-LArA*, go-PST-POSS.3PL] for the plural. However, Table 5.13 shows that the attested

[7] As was shown in Table 5.12, only in the nominative case this leads to a low final vowel. In all other cases the vowel following the assimilated stem is also high, but from a morphological point of view this high vowel belongs to the suffix, and is no epenthetic vowel, as is the case in other persons.

forms are *barda* and *bardïlar*. For the third person singular it could still be argued that it follows the regular scheme if the past tense suffix is analysed as *-T(I)* instead of *-TI,* that is when the high final vowel is seen as an epenthetic vowel, which only appears when the following suffix begins with a consonant, as in *bardïbït* [*bar-T(I)-BIt*, go-PST-1PL]. In this case, the analysis of *barda* would be *bar-T(I)-(t)A*, go-PST-POSS.3SG, and would be regular despite the absence of a high vowel in the past tense suffix.

However, the third person plural form is not so easy to explain within the regular paradigm. Even an analogical division of the suffix *-TIlAr* into past tense suffix *-T(I)* and plural suffix *-LAr* does not make it compatible with the regular scheme of stem+PST+POSS, according to which we would expect a third person plural form of *bardïlara* [*bar-T(I)-LArA*, go-PST-POSS.3SG]. An example of the entire glossed paradigm for the recent past is presented in Table 5.13, alongside with the inflectional paradigm of a possessive marked noun for comparison. The difference in person marking is highlighted in bold.

Table 5.13: Inflectional paradigm for recent past and nominative possessive declension

RECENT PAST OF *BAR* 'GO'		POSSESSIVE DECLENSION OF *ɑγɑ* 'FATHER'	
bar-dï-m	'I went'	*ɑγɑ -m*	'my father'
bar-T(I)-(I)m		*ɑγɑ -(I)m*	
go -PST -POSS.1SG		father -POSS.1SG	
bar-dï-ŋ	'you went'	*ɑγɑ-ŋ*	'your father'
bar -T(I)-(I)ŋ		*ɑγɑ -(I)ŋ*	
go -PST -POSS.2SG		father-POSS.2SG	
bar-d-a	'he went'	*ɑγɑ-tɑ*	'his father'
bar -T(I)-(t)A		*ɑγɑ -(t)A*	
go -PST.3SG		father POSS.3SG	
bar-dï-bït	'we went'	*ɑγɑ-bït*	'our father'
bar -T(I)-BIt		*ɑγɑ -BIt*	
go -PST -1PL		father-1PL	
bar-dï-gït	'you went'	*ɑγɑ-gït*	'your father'
bar -T(I)-GIt		*ɑγɑ -GIt*	
go -PST -2PL		father-2PL	
bar-dï-lar	'they went'	***ɑγɑ-lɑrɑ***	'their father'
bar -T(I)-LAr		***ɑγɑ-LArA***	
go -PST -3PL		father-POSS.3PL	

In Sakha, the past tense of the auxiliary verb *e-* 'to be' is inflected in exactly the same way as *bar* or any other verb, with the third person singular and plural following the 'regular' irregular pattern:

SAKHA

(5.21) *onon* *bu* *χara**ɣ-ïm*** *uruk-ka-ttan* *möltöχ* *e-te*
therefore this eye-POSS.1SG in.past-ADJZR-ABL weak be-PST.3SG
'And this eye was weak even before.' (ESY: 98)

(5.22) *Hür-deːχ* *üčügej,* *aktïːbïnaj* *oɣo-lor* ***e-ti-ler*** *[...]*
very.much -PROP good active child -PL be -PST -PL [...]
'They were very good, active kids [...]' (AGM: 177)

In Dolgan, however, the form of the third person plural is changing. Instead of using the typical suffix *-T(I)-LAr*, the overwhelming majority of third person plural forms in the spoken corpus appear as *etilere* [*e-T(I)-LArA*, be-PST-POSS.3PL] 'they are'.

DOLGAN

(5.23) *min* *haː-laːχ* *e-ti-m* *beje-m,* *onton*
1SG gun-PROP be-PST-POSS.1SG self-POSS.1SG then
doɣottor-um *ït-tar-daːχ* ***e-ti-lere***
friend.PL-POSS.1SG dog-PL-PROP be-PST-POSS.3PL
'I had a gun myself, and my friends had dogs' (SEK: 10)

Out of 55 occurrences of third person plural forms in the corpus, 47 are *etilere* (85.5%), and only 8 (14.5%) correspond to the Sakha form *etiler*. In the Sakha corpus, which consists of several Sakha dialects from a wide range of geographical regions including the Olenek region, which is relatively close to the Taimyr, not a single instance of *etilere* was recorded. In contrast to the third person plural, the third person singular is identical in both languages, which would support an analysis of *ete* as a regular form and thus for the past tense suffix to be of the form *-T(I)*. An overview of the paradigm for *e-* 'be' for both Dolgan and Sakha is given in Table 5.14.

Table 5.14: Recent past for auxiliary verb e- 'be'

SAKHA	DOLGAN	
e-ti-m	*e-ti-m*	'I was'
e -T(I) -(I)m	*e - T(I) -(I)m*	
e-ti-ŋ	*e-ti-ŋ*	'you were'
e - T(I) -(I)ŋ	*e - T(I) -(I)ŋ*	
e-t-e	*e-t-e*	'he was'
e - T(I)-(t)A	*e - T(I)-(t)A*	
e-ti-bit	*e-ti-bit*	'we were'
e - T(I) -BIt	*e - T(I) -BIt*	
e-ti-git	*e-ti-git*	'you were'
e - T(I)-GIt	*e - T(I) -GIt*	
e-ti-ler	*e-ti-**lere***	'they were'
e - T(I) -LAr	*e - T(I) -**LArA***	

5.3.2. DIRECTION OF CHANGE AND EARLIER ACCOUNTS IN THE LITERATURE

With respect to the direction of change, we can be confident in assuming that Dolgan is the innovative language. Comparative data from other Turkic languages show that a past tense category with a suffix related to *-T(I)-LAr* is very common within the Turkic language family (including Dolgan for all verbs except *e-* 'be'). To substantiate this, Old Turkic uses *-dIlAr* in the constative preterite (Erdal 1998: 145, 2004: 327), Turkish and Azerbaijanian use *-DI-lAr* for the simple past (Csató and Johanson 1998: 214, Schönig 1998: 254), and Tatar and Bashkir *-DĚ-lAr* (Berta 1998: 292).

As in the case of the unstable stems discussed above, grammatical descriptions of Dolgan provide contradictory information with respect to the inflectional paradigm of *e-*. According to Ubryatova (1985: 167), the paradigm is identical to the one in Sakha, i.e. the third person plural is *etiler*. As far as I am aware, she makes no comments on possible variation of this form in written or spoken discourse. Artemyev (2001: 196), on the other hand, gives only the Dolgan form *etilere*. However, in some of his examples elsewhere in the grammar the form *etiler* also appears (ibid.: 200). Despite the presence of these two forms in Artemyev's data, I have not been able to find a discussion of this variation.

There are several possible reasons for the divergence in these descriptions. First, to do both authors justice, it may be the case that the innovation in Dolgan is

relatively recent. Ubryatova's grammar is based on linguistic material collected in the 1930s, which leaves a significant time span of about 60 years before Artemyev published his grammar in 2001. The fact that Sakha *etiler* is still in use as well, despite the current dominance of *etilere*, could be interpreted as supportive evidence for the idea that this is a recent innovation in Dolgan, that is, both allomorphs are still used and the new form has not yet taken over.

A second explanation may lie in the fact that Ubryatova is a specialist on Sakha as well. Having a thorough knowledge of a very closely related language has many advantages, but could potentially lead to false assumptions about language Y (the new language) on the basis of language X (the language already known). In the case of *etiler* this is not unthinkable. First, only this defective auxiliary verb shows the allomorphy of *-T(I)-LAr* and *-T(I)-LArA*, therefore this variation would not show up while studying verbal morphology with other verbs. Second, it is certainly not the case that *etiler* in Dolgan is ungrammatical. Upon explicit asking whether there is a difference in use or meaning between the two forms, speakers refute that option and say they are fully equivalent and interchangeable. Only a corpus count shows that in practice *etilere* is evidently favoured in spoken discourse. Thus, depending on the way Ubryatova's language material was collected, she may not have come across the form, even if it were already present in Dolgan, because Dolgan speakers would never have rejected *etiler* as a form not belonging to their verbal paradigm.

5.4. DISCUSSION

This chapter has presented two instances of regularisation of paradigms. The first case was classified as an example of reanalysis, the second as regularisation on the basis of analogy with the paradigm of possessive person marking. In the beginning of this chapter, reanalysis was introduced as a process by which the underlying structure of a morphosyntactic sequence changes over time, while the surface structure remains largely unchanged. It was stated that this occurs as a result of potential structural ambiguity of a certain surface form through mechanisms of analogy. The data in this chapter have shown that this is what has happened in unstable stems in Dolgan and that reanalysis is an appropriate term to describe the differences between Dolgan and Sakha in this domain. But how common is this kind of change in languages, and how can it be explained? The next section gives

an overview of what has been said about this topic in terms of language-internal as well as language-external motivations for reanalysis.

5.4.1 PRINCIPLES UNDERLYING PARADIGM REGULARISATION AND REANALYSIS

Within the literature on language development and language change, reanalysis occupies a prominent place. This holds for syntactic as well as morphological change. According to Harris and Campbell (1995: 3) reanalysis has been "(...) the single most important mechanism for most attempts to explain syntactic change throughout the history of linguistics." Similarly, Joseph (1998: 358) describes the emergence of morphology as follows:

> The primary source of morphology is material that is already present in the language, through the mediation of processes of resegmentation and reinterpretation applied in a variety of ways, as well as by other processes of change – for example sound changes – that lead to grammaticalization.

An appealing account of the cognitive principles and mechanisms that may underlie the process of reanalysis is provided by Bybee (1991). In her word-based model of morphological organisation, with a focus on the organisation of paradigms, she stresses the crucial role of language use in shaping language structure. She postulates a strong link between the way language forms appear in discourse and the mental representations that underlie them.

While her account in the 1991 paper deals mainly with the acquisition of morphological paradigms, Bybee's theory suggests a direct link between the acquisition process and the restructuring of morphological paradigms, sometimes through reanalysis. Thus, she assumes that language learners (of L1 as well as L2) are directly involved in language change because they hold assumptions about paradigm structures which may be different from their predecessors or peers.

For Bybee, essential concepts in the dynamics of language are 'basic form', 'markedness' and 'frequency of use'. These three concepts are closely intertwined in the language learning process and in the determination of directionality in paradigm restructuring. 'Basicness' of a form is determined by two factors: a) high frequency in discourse; and b) semantic unmarkedness. Typically semantic unmarkedness correlates with morphological unmarkedness and the idea is that

the marked forms in the paradigm are derived from the basic unmarked form. Typical examples of basic grammatical categories are singular, nominative, first and third person, present tense and indicative mood.

However the correlation between semantic and morphological markedness does not always hold, and can be reversed depending on the frequency of use of the marked form. Sometimes semantically and morphologically marked forms can become 'basic' and thus serve as a basis for regularisation of a morphological paradigm. This phenomenon is called 'local markedness', which I will focus on in some detail because it is a relevant concept in the explanation for the paradigm differences between Dolgan and Sakha. Tiersma (1982: 833-834) provides examples of Frisian, where the plural stem of certain words has been generalised over the entire paradigm to form the basis of the singular too. By the same token, Mańczak (1980: 285) illustrates local markedness with the dominance of locative case in place names and the instrumental case of nouns designating tools. While 'local markedness' seems to go against the idea that basic forms correlate with the semantically and morphologically unmarked members of the paradigm (after all, in the examples referred to, it was nouns marked for plural, locative and instrumental that came to serve as the basis of the morphological paradigm), in all these cases the marked form has a much higher token frequency than the unmarked forms. The plural stems in Frisian were all nouns that normally occur as pairs or in groups in natural discourse, such as *arm*, *tooth*, or *tear*. Similarly, for obvious semantic reasons place names occur most frequently in the locative case and nouns denoting tools in the instrumental case. Therefore the 'marked' categories are for these words more 'basic' than the unmarked.

Bybee concludes that whatever is inherent in the meaning of a word (such as the plural for *tooth*, or the instrumental for *knife*) is treated cognitively as unitary, non-complex and 'basic'. What is marked and unmarked may therefore depend on the semantic properties of an individual lexical item. That is, the plural in *noses* is marked since an individual normally has only one of them, but the plural in *lice* is typically unmarked since they unfortunately tend to appear in large quantities. However, in all cases there is a strong correlation between semantic unmarkedness (singular for *nose* and plural for *louse*) and frequency of use. Thus, frequency of use has the potential to reverse the correlation between semantic and morphological unmarkedness. It strongly influences what is conceived of as 'marked' and 'basic' and is therefore a major and overruling factor in patterns of paradigm levelling.

With respect to language acquisition and learning strategies, a higher token frequency means a higher proportion in the input and therefore a stronger representation in the mental lexicon. Therefore the most frequent forms are learned first, and during language acquisition the rest of the paradigm is interpreted as being derived from these forms. Normally this is the morphologically and semantically unmarked basic form, but in the case of 'local markedness' a morphologically marked form can serve this purpose equally well. Although Bybee does not provide support for this claim with experimental data, she assumes on the basis of the general cognitive nature of these learning principles that the same learning strategies and generalisations take place in second language learning.

5.4.2 REGULARISATION AND REANALYSIS IN DOLGAN EXPLAINED

The previous section has provided a definition of regularisation in morphological paradigms by reanalysis, and an understanding of the cognitive mechanisms that may underlie it, with a focus on language-internal development and L1 acquisition. In this section I will discuss how paradigm regularisation in Dolgan can be interpreted in terms of these mechanisms.

With respect to these changes there are two questions to be answered: a) how did this change happen in Dolgan; and b) why did it not happen in Sakha? After all, if reanalysis and regularisation are such common language-internal processes, there is no obvious reason why the same forms would not have developed in other Sakha-speaking regions. In this section I will deal with question a); question b) will be discussed in the next section.

For relational as well as referential nouns it was shown that the ambiguity of forms has led to a change where the oblique stem has become generalised over the entire paradigm, with a few exceptions of very frequent, possibly fossilised forms, where the basic Sakha stem is still used. As Koch explains it, in ambiguous situations the underlying form that will eventually be selected is 'the word form that appears most frequently for the particular lexeme' and has the highest 'paradigm frequency' (Koch 1996: 232), that is the number of slots in the paradigm in which it occurs. Typically, these conditions correlate with Bybee's prototypical basic form, in which high frequency of use, and semantic, and morphological unmarkedness conspire towards an ideal basic form. However, as was illustrated

with the example from Frisian, 'local markedness' provides an exception to this correlation, and an important cue to how the semantics of individual lexical items can shift the centre of gravity in the understanding of 'basicness'. The regularisation in Dolgan seems to be an illustration of exactly this.

The relational nouns illustrate this in an obvious way. Like postpositions, their inherent (read: unmarked) semantics represent locational, directional or instrumental relations, which in Sakha correlate with possessive marked dative, ablative and instrumental cases. Consequently, these are the forms of relational nouns that are expected to occur most frequently in discourse, which was unequivocally confirmed by the data from the Sakha and Dolgan corpus, in which only very few unmarked nominatives were found. Whether it was high frequency that triggered this interpretation of semantic unmarkedness or the other way round remains a chicken and egg type question, but it is clear that for relational nouns there is a strong correlation between the two. The high frequency and semantic unmarkedness qualify the morphologically marked forms to be interpreted as basic.

For the referential nouns the picture is very similar. Again, the oblique stem has become generalised over the whole paradigm, and judging by the high final vowel in Dolgan stems today, in this case forms taking an epenthetic vowel lie at the heart of this reanalysis, which is all possessive marked forms, except the third person singular. This nicely fits the idea of 'local markedness' discussed in 5.4.1. In many cases the reanalysed referential nouns concerned concepts (family members, body parts) that for semantic reasons typically require possessive marking. Therefore, the possessive-marked form is for these nouns the semantically unmarked, the most frequent and therefore most basic form. This makes it another illustration of 'local markedness' and explains how a morphologically marked form can become the basic form within a paradigm. More generally, it illustrates the importance of semantic properties in the frequency of occurrence of morphological features, and thus in the structuring of morphological paradigms.

Furthermore, it was demonstrated that some forms continue to exist in their Sakha shape, even though the oblique form in Dolgan clearly dominates the morphological paradigms in the category of unstable stems. For relational nouns, it was shown that this could be explained in terms of discourse frequency as well. As a reminder, *kenne, kennitten*, *ürdüger* and *ürdünen*, which are the Sakha forms that occur in Dolgan discourse, are the most frequently used forms in Sakha. For

these particular forms I argue that there are forms such as *kennitten* that, due to the high input frequency, have become stored in the mental lexicon as an unanalysable unit, and continue to exist alongside the reanalysed forms, where the relational nouns are taken apart into a stem and a case suffix. In the case of Dolgan, they seem to occur in free variation.

For referential nouns, the preservation of the Sakha forms may be explained along similar lines. However, an additional factor which may have increased the ambiguity in the first place is the apparent randomness of the application of the phonological rule that eliminates the final vowel in this word class (i.e. the difference between stable and unstable stems). The fact that there are some words for which this rule applies and others for which it does not, may have enhanced uncertainty with respect to the underlying form, which could have lead to the use of both stems, and the preservation of Sakha stems in some nouns. Over time the 'one form one meaning' principle could have lead to a differentiation in meaning between the two forms, as we see for *tumus*.

5.4.3 THE POTENTIAL ROLE OF L1 AND L2 LEARNERS IN REANALYSIS

The previous sections have shown that reanalysis and regularisation of paradigms is a common phenomenon in natural language change and can often be satisfactorily explained in terms of language-internal motivations. In other words, from a purely linguistic point of view there is no obvious need to include the influence of second language learners or language contact in the picture. Admittedly, distinguishing between L1 and L2 influence is a difficult task, since principles such as markedness and frequency are of a general cognitive nature, and presumably apply to learning mechanisms in L1 as well as L2 acquisition (Bybee 1991: 88). However, an exclusively language-internal account of the changes in Dolgan morphology leaves one issue unexplained, namely their geographical distribution. If paradigm regularisation is such a common language-internal development, why did this particular change happen only in Dolgan and not in any dialects of Sakha? All things being equal, one could expect a linguistic change to arise and gradually spread across the speech community if the social conditions are favourable. However, the changes described here are restricted to the Dolgan-speaking area only.

Since the cognitive conditions are obviously the same for speakers of Sakha on the Taimyr and, say, speakers of Sakha in Central Yakutia, more significance must be attributed to the equally important set of explanatory factors in language change, namely the sociolinguistic situation. It is widely accepted that the social situation in which a change occurs is crucial for its further spread and development within the speech community (see Section 3.1.4 for discussion). We know that the sociolinguistic situation was certainly not identical across Sakha-speaking communities, so therefore it may have conditioned the appearance and spread of certain linguistic variants within particular parts of the broader speech community.

An important sociolinguistic condition to consider within the context of this dissertation is the potential interaction with other populations. It appears that there is a correspondence between the area where paradigm regularisation has taken place (the Taimyr Peninsula) and where we know from historical records that different populations have been in contact (in particular Sakha and Evenks). Based on historical sources as well as on research in other linguistic domains, such as the lexicon (see Chapter 4) we may assume interaction between these populations and a certain degree of bilingualism. Although historical evidence of population contact alone is no guarantee for significant linguistic contact, the overlap between the area of population contact on the Taimyr and the spread of this particular change increases the potential relevance of second language learners in the rise of this change in Dolgan. Recognising this role, one could imagine that paradigm regularisation came about through L1 Evenki speakers, who were learning Sakha as a second language. During the learning process they may have reanalysed and regularised the paradigms of ambiguous noun stems, motivated by cognitive principles related to semantic and morphological markedness and frequency, as discussed above.

Thomason and Kaufman (1988: 145) describe paradigm regularisation, or simplification, as a phenomenon that is characteristic for L2 acquisition as much as it is for L1 acquisition, and associate it with a situation of language shift. Where in monolingual Sakha communities regularised variants in young children in all likelihood get corrected over time (presuming that they do occur in L1 acquisition as well), this has not happened in the Dolgan community. This difference may have several explanations. First, it may be due to the age of the bilingual speakers. It is commonly known that the ability to achieve native-like fluency in a second language decreases with age, and becomes impossible after the so-called critical

period (see Chapter 3, Section 3.1.3). While leaving the details as to the exact age limit unaddressed, the important point is that non-native variants are much more likely to remain uncorrected in adult bilinguals than in children. Since it is hard to imagine a shifting population that consisted only of children, it is plausible to assume a large number of bilingual adult speakers, leading to a stronger persistence of regularised variants.

Second, the regularised forms may have become established due to a large number of shifting speakers. When the group of shifting speakers is large, non-native variants are more likely to stay within the community than if only a few speakers were shifting. In a large group the bilingual speakers are more likely to hear each others' foreign version of the language they are shifting to than when the shifting group is small, and consequently their exposure to 'native' variants of Sakha would be lower than for small shifting groups.

A third important point is the fact that in the contact setting on the Taimyr Sakha was used as a lingua franca. This means that even without taking into account the shifting groups, there were many second language speakers of Sakha. Since regularisation is known to occur frequently in languages of wider communication, the function of Sakha on the Taimyr in this capacity is another plausible explanation for the geographical distribution of the change.

A final possible explanation may be found in current L1 attrition. Attrition often involves simplification due to the lack of language use, language input, or the lack of exposure to the language, which causes details and irregularities to be levelled out. Attrition can affect every linguistic subdomain, including the structure of a speakers' L1. As Sharwood Smith and Van Buren (1991: 20) have it,

> ...the attrition of competence may be triggered by changes in the learner's perception of the basic structure of his or her L1 grammar, and not by a tendency to ease the processing burden of an underused L1."

It is clear that it is impossible to categorically tease apart the influence from L1 and L2 acquisition because many of the cognitive principles apply to both. Thus it will never be possible to prove whether regularisation in Dolgan paradigms is the result of language-internal change of language shift, or of the use of Sakha as a lingua franca. However, if contact played a significant role in the development of these differences between Dolgan and Sakha, it would most likely be caused by considerably large groups of adult non-Sakha speakers, who shifted to Dolgan/Sakha and learned it as a second language. However, language-internal

development is probable too, since regularisation is common in internally motivated language change.

As unrealistic as it is to have the desire to completely separate language-internal and language-external motivations, so it is unrealistic that these factors were radically separated in reality. Rather I would argue that the linguistic outcome we see in present-day Dolgan is the result of the interplay between the two scenarios, which reinforced each other. In L1 and L2 acquisition, regularisation occurs as a consequence of general cognitive learning principles in a situation of plausible language change. If a regularisation was made by a second language learner, it may have become more easily accepted by native speakers because it is a plausible change for L1 speakers as well. Similarly, L1 speakers who regularise paradigms during the acquisition process may hear these forms from other people around them and thus the language-internal tendency would be reinforced. Thus it is important to take both factors into account as significant possibilities in the explanation of the divergence of Dolgan and Sakha in this respect, even if they are technically inseparable.