# **OUDB Glossing Conventions FLEx8 - Preliminaries**

#### 1. Start a new project

#### 1.1. Set-Up: Writing system and alphabetic order

In the task bar – Go to:

File > Project Management > Fieldworks Project Properties > Writing System > Modify > Sorting > Custom ICU rules

Copy & Paste into the box:

#### 2. Language

#### 2.1. The meta language for glossing is English,

therefore the glosses are in English as well as the translations in the lexicon entries. Translations into other languages (German, Russian, Hungarian) are optional.

# 2.2. In the lexicon, every translation has to be put correctly into the corresponding translation field,

i.e. English translations are to be written into the field for English translations, Russian into the Russian field, etc.

Eng	1SG	
Rus	1SG	
Ger	1SG	
Hun		
Fin		
oth		
NotePu	du	
NoteInt	t	
	Rus Ger Hun Fin oth NotePu	Rus 1SG Ger 1SG Hun Fin

Figure 1 Sample entry with several translations

# 2.3. It is important that the language of the translations matches with the language preset in the corresponding fields.

# 2.4. In texts & words, the language preset in the taskbar on top of the screen must match the language of the translation line in the interlinear layers, too.

If only one language has been chosen for translation, the taskbar field remains inactive and has not to be edited. If there is more than one language of translation, this has to be checked thoroughly as there will be major errors in the export of the texts if there is any mismatch.

	ŋkəli pɛ:nə o:pi ittle bird and his						
nfo Baseline Gloss	Analyze Taggi	ng Print Vi	iew Text Chart				
I.1 Word Morphemes Lex. Entries Lex. Gloss	pi:t <sup>j</sup> əŋkəli pi:t <sup>j</sup> əŋkəli pi:t <sup>j</sup> əŋkəli little bird	pemə pemə pemə and	o:pi o:pi o:pi older sister	βałłə βał βałı live	yən -ł -łı PRS	-əyən -əyən 3DU	#.#
Lex. Gram. Info.		cconj	subs	v	v:Any	v:Any	

Figure 2 Matching languages in taskbar and translation line

To alter the language in the taskbar field if there is a mismatch, mark the whole line first and then adapt he language in the taskbar; sometimes it is necessary to do this with all lines of translations, even if only one line has a mismatch.

# 2.5. Words of foreign origin in corpus texts (code-switching)

If there is a word of foreign origin in the text (e.g. a Russian term is used instead of the correct Khanty or Mansi term), it will be glossed and translated, too. This applies mainly to conjunctions or similar fillers and is often found in recorded texts. Adapted foreign words or loanwords (e.g. words which are found in Khanty or Mansi dictionaries) are glossed as usual.

2.4	Word		unio.	toye	mən	#,#	poł	t <sup>j</sup> t <sup>j</sup>	i	ŧi:β		#,#	βo:je	t <sup>j</sup> i	ŧi:β		#,#	ŧi:β	#,#	pen
	Morphem	les	unie	toye	mən		po₽	t <sup>j</sup> t <sup>j</sup>	i	ŧi:β			βo:je	t <sup>j</sup> i	ŧi:β			ŧi:β		pen
	Lex. Entri Lex. Glos Lex. Gran	s	uriə down adv	-	mən <sub>1</sub> +[PST.3SG] go v		po# fat sub:	s	D	H:+[PST eat v	.3SG]		βo:j fat subs	t <sup>j</sup> i <sub>2</sub> so ptcl	ii:+[₽ eat v	ST.3SG]		ii:+[PST.3SG] eat v		pe:nə+fr. var. and cconj
əj m	ete fe:tne	ke:mən	##	ke:k	jeßo	#	##	meŋk	i:k	i lu	mp	seŋki	tə		stej	#,#				
əj m	ete <del>l</del> e:tne	ke:mən		ke:k	jeβo			meŋk	i:k	i lu	ump	seŋkił	-te		suj					
	e day	ke:mən outside adv		kæ:k [Ru. how] sconj	jeβo [Ru. 3SG.ACC] ppron			meŋk maŋk subs	i:k Sir su	s)	ump ai 1bs	seŋkił hit v	-tə PTCP. v>ptcj		suj sound subs					

Figure 3 Words of foreign origin in Surgut text sample

In consequence, the foreign word gets an entry in the lexicon, too. Every translation is put into square brackets including the information of the foreign language. Part of speech information is according to the language of origin.

bo:tʃkɐ	bo:t∫ke	[Ru. barrel]	Substantive	[Ru. бочка]	[Ru. Fass]	[Ru. hordó]
jə∬o	jə∬o	[Ru. still]	Particle	[Ru. emë]	[Ru. noch]	[Ru. még]
ne:	ne:	[Ru. take]	Particle	[Ru. на]	[Ru. nimm]	[Ru. tessék]
∫tobu::	∫tobu::	[Ru. so that]	Subordinating conjunction	[Ru. чтобы]	[Ru. um zu]	[Ru. azért, hogy]
t <sup>j</sup> e:be	t <sup>j</sup> e:be	[Ru. 2SG.DAT]	Personal pronoun	[Ru. 2SG.DAT]	[Ru. 2SG.DAT]	[Ru. 2SG.DAT]
βo:t	βo:t	[Ru. here]	Particle	[Ru. вот]	[Ru. nun]	[Ru. nos]
βзјо	βsjo	[Ru. that's all]	Particle	[Ru. всё]	[Ru. das ist alles]	[Ru. ez minden]

#### ANALYSED TEXT CORPORA AND DICTIONARIES FOR LESS DESCRIBED OB-UGRIC DIALECTS

#### Figure 4 Lexicon entries of words of foreign origin, SK

The same applies for other registers of speech that are notable, e.g. child language: put a note behind the word in square brackets, e.g.

Eng fish [child lang.] Ger Fisch [Kindersprache]

N.B.: we do not mark Code-Switching in the translations, since it is clearly marked in the glossing. If there is whole sentences in a foreign language within the text, however, we mark them in the translation.

# 2.6. Unknown senses

If we cannot find out the sense of a word, it will be glossed with [n.n.] in the lexicon, part of speech is an assumption derived from its role in the sentence. In the translation, we mark the unknown word with [n.n.], too.

Entries							
Headword	Δ	Lexeme Form	Glosses	Grammatical Info	Glosses (Rus)	Glosses (Ger)	Glosses (Hun)
Show All	•	Show All 🔹	n. 🔻	Show All 🔹 🔻	Show All 🔹	Show All	Show All
t <sup>j</sup> i:nt <del>l</del> it		t <sup>j</sup> i:nt <del>l</del> it	[n.n.]	Adjective	[n.n.]	[n.n.]	[n.n.]

Figure 5 Lexicon entry with unknown sense, PA

# 3. Corpus-based working

In order to work as corpus-based as possible, it must be guaranteed, that only those forms are part of the lexicon that occur in the texts. Therefore please pay attention to the following:

- Only note those senses in the lexicon that are used in the texts. If more senses for one lemma can be found in the dictionary, do not enter them in our lexicon if they are not used in the texts.
- The same applies to words that are not used in the texts (so-called empty head entries). We only create empty head entries if we e.g. deal with complex word forms (such as prefix verbs). In these cases it might be necessary to enter words which do not occur in the texts individually in order to state all components of a complex word form. It might also be necessary to create empty head entries for variants. If a verb only occurs e.g. in the third person singular past in the texts, and this form is a variant, we have to create an entry for the basic verb form, too, even it is not used in this form in the texts. Try to keep the amount of empty head entries small and use them only if necessary.

# 4. There are several layers of language analysis in FLEx 8:

- Word (automatically filled with the input on the baseline)
- Morphemes (segmentation of words into morphemes)
- Lex. Entries (the lexemes according to their entry in the lexicon, e.g. indication of spelling variants)
- Lex. Gloss (the actual glossing, generally rendered by abbreviated grammatical category labels)
- Lex. Gram. Info. (information on part of speech)
- Free (Translations)

1.1	Word	e:k <sup>w</sup> a piyris <sup>i</sup>	a:kwei	ntəl		o:lea	Y		#.#
	Morphemes	e:k <sup>w</sup> a piyris <sup>i</sup>	a:k <sup>w</sup>	-en	-təl	o:1	-e	-әү	
	Lex. Entries	e:k <sup>w</sup> a piyris <sup>i</sup>	a:k <sup>w</sup>	-en <sub>2</sub>	-əl	o:l2	-c:y	-Y	
	Lex. Gloss Lex. Gram. Info.	E:k <sup>w</sup> a Piyris <sup>j</sup> nprop		SG<3DU n:Any	INST n:Any		PRS v:Any	3DU v:Any	

Free Eng E.P.'s been spending life with his aunt.

Ger Es lebten E:kwa piyrisi und seine Tante.

#### Figure 6 Example FLEx 8 layers of analysis (NM)

# 5. The glossing rules are based on the Leipzig Glossing Rules

(version 08-02-05 http://www.eva.mpg.de/lingua/pdf/LGR08.02.05.pdf) which are modified corresponding to the needs of Ob-Ugric languages and the data handling in FLEx 8.

# 6. For the glossing abbreviations capital letters are used (SG, DLAT ...).

A list of abbreviations is given at the end of this document.

If the preset abbreviations do not suit our abbreviations are missing, please set up new ones and discuss them with the coordinators of all dialects; it is always possible, to add abbreviations so you don't have to set up all before you start to work. This keeps our glossings as corpus-based as possible.

# 7. All in all, glossing in FLEx follows a word - by -word alignment

with a morpheme-by-morpheme correspondence (there must be exactly the same number of segments in the example and in the gloss).

# 8. Glossing Characters and formal conventions

Generally, we should analyze and separate by hyphens as many morphemes as possible. Ideally, there should be a 1:1 correspondence of segmented morpheme and grammatical information.

- 8.1. The break characters for the analysis which are integrated in FLEx are
- the hyphen [Unicode 002D] for all kind of affixes (prefix, suffix, circumfix, infix) and
- the equal sign = [Unicode 003D] for all kind of clitics (enclitic, proclitic)

The break character is always grouped with the bound morpheme/clitic, not with the stem. Please apply this rule in dependence from the status as prefix/proclitic resp. as suffix/enclitic:

Insert Mor	nsert Morpheme Breaks					
		e morphemes morpheme typ				
Examples	blackbird	→ black bird	1			
Word:	0:1 -е -әұ				[aa ] Stem [a- ] Prefix	
Break Cha	aracters				[-a] Suffix	
	stem prefix suffix ] infix	[*aa ] [a= ] [ =a] [ =a= ] [ ~a~ ]	bound stem proclitic enclitic simulfix suprafix		[-a-] Infix [*aa] Bound Stem [a=] Proclitic [=a] Enclitic [=a=] Simulfix	
	ОК	Cancel	Help	_ ר	[~a~] Suprafix Morpheme Break Help	

Figure 7 Morpheme break box FLEx 8 with possibilities for separating morphemes

8.2. Person and number are always separated by hyphens within the gloss if possible.

$\odot$		41
v	v:Any	v:Any
begin	PST	3PL
💌 pat <sub>1</sub>	💌 -S1	💌 -t <sub>3</sub>
💌 pat	-S	-ət
patset		

Figure 8 Person and number glossing, NM

8.3. If a single object-language element represents a combination of several meta language categories, these will be separated by a full stop.



Figure 9 Glossing of several meta language categories, NM

The order of the abbreviations corresponds to the general order of categories/suffixes within the dialects in cases when the affixes can be separated transparently. Participles, for instance, are glossed: PTCP.PST and PTCP.PRS

# 8.4. The symbol "<" is used in the glossing of verbal personal markers encoding simultaneously subject and direct object (objective conjugation) and of nominal personal markers encoding two referents (possessive suffixes).

The encoding of both referents is asymmetric:

- person and number of subject and number of direct object
- person and number of the referent generally referred to as *possessor* and number of the referent generally referred to as *possessum*;
- The order of the glossing components is direct object < subject, resp. possessum < possessor</li>

put	PST v:Any	SG<3PL v:Any
💌 pin	💌 -S1	▼ -anel1
💌 pin	-S	-anəl
🔻 pinsa	nəl	

Figure 10 Glossing of objective conjugation, NM

8.5. If an object-language element cannot be segmented formally or semantically, and the meta language lacks a single-word equivalent, use several words, separated by blanks.

a:ps <sup>j</sup> im	
💌 a:ps <sup>j</sup> i	-m
💌 a:ps <sup>j</sup> i	💌 -m1
younger brother	SG<1SG
subs	n:Any
•	4 1

Figure 11 Glossing of translations consisting of several lexemes, NM

8.6. Grammatical information which is not overtly expressed (zero morphemes), is displayed in square brackets [], i.e. information expressed by zero is bracketed; the bracket is attached to the preceding gloss.

minas	
💌 min	-85
▼ min <sub>1</sub> +sp. var.	▼ -S1
go	PST[3SG]
v	v:Any
•	4 1

Figure 12 Glossing of zero morphemes, NM

# ANALYSED TEXT CORPORA AND DICTIONARIES FOR LESS DESCRIBED OB-UGRIC DIALECTS

Punctuation character	Unicode (hex)	To use in following cases
	002E	Ad: if you need the triple point at the end of a sentence, please type this character triply
,	002C	
:	003A	WARNING: don't mix up with :
;	003B	
-	002D	WARNING: don't mix up with –
-	2013	please use it when text is interrupted
!	0021	
?	003F	
	2026	WARNING: use it only in the middle of a sentence, never at the end!
#	0023	
(	0028	Do not use brackets in texts, as FLEx cannot proceed
)	0029	- them
[	005B	
]	005D	
"	201C	please use it at the beginning of the passage
"	201D	please use it at the end of the passage
«	00AB	please use it only in case of text in text at the beginning of the passage
»	00BB	please use it only in case of text in text at the end of the passage
1	007C	If a poetry text is rather structured in lines than in sentences please put the following combination of symbols at the end of each line: # #, as a marker for line break.

8.7. Punctuation characters used in Baseline

#### 8.8. Phrases consisting of several independent words

For a phrase, first click on the chain symbol which is displayed on the upper left side of the box in order to connect the two elements:

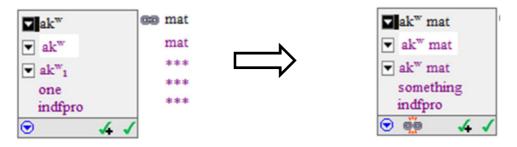


Figure 13 Connecting two words into one phrase, NM

Then insert the phrase as a lexicon entry and specify the morpheme type as "phrase".

work with the texts it turned out to be helpful to take notes about where to find an entry in one of the dictionaries we are working with. Since this is valuable information, we decided to create a custom field "dictionary" which will be part of our corpus-based dictionary on the website, too.

# 8.9. How to create a new field in FLEx Lexicon:

In the menue, click on *Tools*;

Choose Configure and then select Custom Fields:

	00												
Ob-Ugric_PelymMans	i-Flex8	- FieldV	Vorks La	nguage E	xplorer								
File Send/Receive	Edit	View	Data	Insert	Form	at	Tools	Parser	Window	/ Hel	р		
i 🔇 🔘 🗙 🗠 🗠	۵ 🕻	<b> </b>	⇒ 4	1 : 8	1 🤁	¢	C	onfigure		•		Dictionary	
Lexicon	Ent	ries					N	lerge with	entry		₽.	Columns	
Lexicon Edit	i 🛆 Lexeme Form G				Glo	alc Spelling +			•		Homograph Number	hs	
Browse Dictionary	.11		Show	v All	•	Sł	U	tilities				Restore Defaults	•
Collect Words			teæt				0	ptions		Se		Set up Writing Systems	
Classified Dictionary	25C OPI Descend eren		Custom Fields										
Bulk Edit Entries	ŝu		teæw	cặn			G.DLA	1	Persona				<u> </u>
Bulk Edit Reversal Entries			teæw	e		350	G.OBL		Persona	al prono	oun		
						200	TODI		D				

A box Custom Fields opens:

ANALYSED TEXT CORPORA AND DICTIONARIES FOR LESS DESCRIBED OB-UGRIC DIALECTS

Custom Field	s:		Settings
Name	Location	Туре	Custom Field Name:
			Location: Entry -
			Description:
			Type: Single-line Text v
			List:
			Writing System(s):

Click *Add* on the bottom, a new entry appears on the left side of the box.

Now you can edit the new entry on the right side; fill in the name of the new field:

# Dictionary

Note: Make sure that *Entry* is selected in the field *Location*, otherwise change it with clicking on the arrow!

Custom Fields			Settings				
Name	Location	Туре	Custom Field Name:				
Dictionary	Entry	Single-line Text	Dictionary				
			Location: Entry				
			Description:				
			Type: Single-line Text   List:   Writing System(s):				
			First Analysis Writing System 🔻				

Click OK;

Every lemma entry should now show the new field:

ANALYSED TEXT CORPORA AND DICTIONARIES FOR LESS DESCRIBED OB-UGRIC DIALECTS

Entry		🔳 Show Hidden Field
oj n summer		
Lexeme Form	ObU_IPA	
Morph Type	toj stem	
Citation Form	ObU_IPA	
Components		
Note	Eng	
	Rus	
	Ge/Hu/Fi	
	Test	
Dictionary	M/K 478	
Messages	G	
Sense 1		
Gloss	Eng summer	
	Rus	
	Ge/Hu/Fi	
	Test	
Definition	Eng	
	Rus	
	Ge/Hu/Fi	
	Test	
Grammatical Info.	Noun	•

#### Format

Here are the dictionaries used for reference for Mansi, e.g.:

- K = Kannisto, Artturi (2013): Wogulisches Wörterbuch. Gesammelt und geordnet von Artturi Kannisto. Bearbeitet von Vuokko Eiras. Herausgegeben von Arto Moisio Lexica Societatis Fenno-Ugricae XXXV Kotimaisten kielten keskuksen julkaisuja 173. Helsinki: Suomalais-Ugrilainen Seura / Korimaisten Kielten Keskus.
- M/K = Munkácsi, Bernát Kálmán, Béla (1986): Wogulisches Wörterbuch Budapest: Akadémiai Kiadó
- **В/V** = Баландин, А. Н. Вахрушева, М.П. [Balandin, А. N. Vahruševa, М. Р.] (1957): *Мансийский язык. Учебное пособие для педагогических училищ* Ленинград: Учпедгиз
- Ká = Kálmán, Béla (1963): Chrestomathia Vogulica. Budapest: Tankönyvkiadó, 124
- Ká2 = Kálmán, Béla (1976): Wogulische Texte mit einem Glossar. Gesammelt und bearbeitet.
   Aus dem Ungarischen übersetzt von H. Krüger-Tokody und P. Kocsány. Budapest:
   Akadémiai Kiadó
- **R** = Riese, Timothy (2001): *Vogul.* Languages of the World/Materials, 158 . München New Castle: LINCOM EUROPA

If you wish to display the dictionary field in the columns on the left side, click on the tiny icon in the upper right edge of the columns.

ANALYSED TEXT CORPORA AND DICTIONARIES FOR LESS DESCRIBED OB-UGRIC DIALECTS

ticon	Entries							Entry	Show Hidden Fi	
wicon Edit	Headword	A Lexeme Form	Glosses	Glosses (Ger)		Variants	Dictionary	$\mathbf{m}$ <b>kui</b> <sub>2</sub> prs stem allomorph 3SG of , sp. var. of <b>kuj</b> <sub>2</sub>		
ctionary	Show All	<ul> <li>Show All</li> </ul>	Show All	Show All      √		Show All	✓ Non-bl ✓			
lect Words safled Dictionary	oſı	oſ	sheep	Schaf	Free Variant		K 105a	Lexeme Form	OBU_IFA kui	
Edit Entries ersal Indexes	jæt	jæt	middle	mittlerer	Free Variant		K 244b	Morph Type		
ersal Indexes Edit Reversal Entre	kal <sub>1</sub>	kal	middle	Mitte Zeit	Dialectal Variant		K 279ab.	Citation Form	stem ObU_IPA	
			gap	Zen			M/K 72b	Variant Type	Present Stem Allomorph 3SG Spelling Variant	
	ko:j	ko:j	come hit	ankomme n treffen	Spelling Variant	koj <sub>2</sub> koji	K 267a, M/K 95a	Variant of Components	kuj <sub>2</sub>	
		1	1-(ppc 2001	tremen	Present Stem		da bin	Note	Eng	
	kui <sub>2</sub>	kui	lie[PRS.3SG]		Allomorph 3SG, Spelling Variant		da bin ich!		Rus Ger	
	man;	man	1PL		Free Variant		K 475b		Hun	
	men	men	1DU		Spelling Variant	me:n	K 468b		NoteInt	
	tayl <sub>1</sub>	tayl	clothing dress	Kleidung	Free Variant		M/K 612a	Dictionary	NotePub da bin ich!	
	temo	temo	food	Speise	Spelling Variant		K 862b	Messages	Ga bin icn:	
	wutt	wutt	place build	setzen, stellen bauen	Stem merged with Present Marker		K 67a, M/K 697b	Sense 1 Gloss	Eng lie[PRS.3SG]	
exicon									Oer	
exts & Words									Hun	
									Fin	

A box *Configure Columns* opens. Choose *Dictionary* on the right side, click "ADD" in the center and it appears on the left side of the box (the selected columns). Click OK.

exicon	Entries							Entry	Show Hidden Fie
Lexicon Edit Browse Dictionary	Headword Show All	△ Lexeme Form     ✓ Show All	Glosses	Glosses (Ger) Variant Types		Variants Dictio Show All ~ Non		<b>kui</b> <sub>2</sub> prs stem alle	pmorph 3SG of , sp. var. of $kuj_2$
Collect Words Dassified Dictionary	of	oſ	sheep	Schaf Free Varian	t	K 1	105a	Lexeme Form	ObU_IPA
	jæt	jæt	middle	mittlerer Free Varian	t	K 244b			kui
	kal <sub>1</sub>	kal	middle	🛱 Configure Columns			_	Morph Type ×	obu_IPA
			time gap	Column Options		Current Columns			Present Stem Allomorph 3SG Spelling Variant
	ko:j	ko:j	come hit	Cross References A CV Patterns Date Created Date Modified Definition		Field Headword Lexeme Form Glosses	Writing S Default V Default V Default A	lemacular lemacular	kuj <sub>2</sub>
п	kui <sub>2</sub>	kui	lie[PRS.3	Dictionary     Discourse Notes     Encyclopedic Info     Environments     Environments (Allomorph)	Add >> Remove	Variant Types Variants	German Best Ana Default V	lemacular	Eng Rus Ger
	man <sub>1</sub>	man	1PL			Dictionary	Uefaut A	ralysis 🔶	Mun Fin
	men	men	1DU	Etymology - Comment Etymology - Form					NoteInt
	tayl <sub>1</sub>	tayl	clothing dress	Etymology - Gloss Etymology - Source Y					NotePub da bin ich!
	tema	tema	food	Writing System Default Analysis					A on ich:
	wutt	wutt	place build			ОК	Cance	l Help	Erg lie[PRS.3SG]
				bauen 697b					Rus
Lexicon									Ger Hun
Texts & Words									Fin
Grammar									Fin Noteint