

Preparing the files

Before starting ELAN, create separate .wav and .eaf files for FW (sometimes coded as “S” in Homeland data), Long questionnaire (IV in Heritage data) and any other conversation (“C”, “G” in Homeland data) or elicitation (“E”, “T” in Homeland data). You can easily cut up or paste together files using Audacity (<http://www.audacityteam.org/download/>). Either select the relevant portion and choose “Export selection as .wav” from the File menu *or* copy and paste relevant sections, one following another, into a new Audacity file and then be sure to “Export as .wav” as the final step for any new file you create. In the interview_catalog.xlsx, record the starting time-point in the original long .wav for FW and EOQ, so we can find them again if necessary.

If the digital recording of a particular interview is divided up on lots of different .wav files you should **paste them together before linking to ELAN**. Use Audacity.

Match .wav and .eaf filenames.

Using ELAN

- 1) Start ELAN.
- 2) Choose “New” from the “File” menu.
- 3) Select the appropriate **.wav** (audio) file as “Media” and select **Faetar_template.etf** or **Faetar_2speakers.etf** as your “Template.” Click OK. (These are in FAETAR (Heritage)/ELAN templates on iMac3)
- 4) Choose “AUTO BACKUP > 1 minute” from the File menu.
- 5) **Edit>Preferences>Edit Preferences > Editing > Enter key commits change in the inline box.** (This lets you save an annotation just by hitting Enter before you leave the annotation box.)
- 6) Immediately save your file with the appropriate filename for the speaker you are starting to annotate. I’ll call it **F2F35A_IV.eaf**. Save it in a memorable place on your computer (or your folder on a lab iMac). **Match .wav and .eaf** filenames. That means that the ELAN file should have the same name (Speaker ID) as the .wav file you are working with.)
- 7) You should now see a soundwave in the center of the **F2F35A_IV.eaf** window. If you don’t, Command-Click (**⌘**+Mouse click) on the place where the soundwave should be – right below the Play/Pause buttons (it may be showing as a flat horizontal line) and then choose a number to magnify/Vertical Zoom in by. Keep trying until you see the soundwave enough to be able to distinguish speech from silence. [If it is still too quiet to hear quickly, use the “Amplify” function to edit the .wav in Audacity.]

- 8) Test volume. (☰ menu > System Preferences > Sound > Output or the 🔊 icon on the upper right of your monitor.)

It saves lots of time to **divide up a file into phrases before transcribing**. To do that in ELAN:

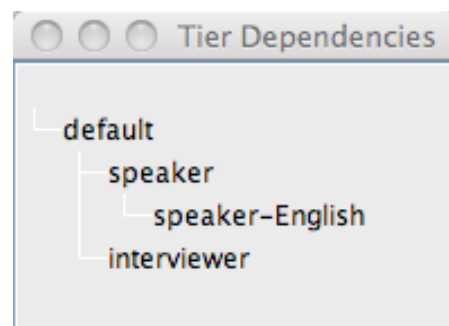
Divide the interview into phrases/clauses/speaker turns

- 9) Click on "Segmentation..." in the Tier menu.
- 10) Select the Default tier.
- 11) Selection "One keystroke per annotation" (my preference) or "two keystrokes per annotation (non-adjacent)."
- 12) Click the Play button in that new window. (You *might* want to first change the playback speed in the "Controls" view.)
- 13) Each time a new clause/phrase/sentence/turn starts (that you will want to transcribe in a separate Annotation), click the Return key. (If you've selected "2 keystrokes", then hit it at the start and stop of each annotation – this would be useful for a slow speaker with lots of pauses.) *Note: Nothing shows up when you click!*
- 14) After you've listened and clicked for a while, click Pause and then the "Apply" button.
- 15) After you do that, all the divisions you selected will show up in the default tier.

Create a "daughter" tier for each Speaker

The next step is to get these annotation blocks correctly associated with each speaker. For this, create a new tier for each speaker and make it a daughter of the default tier.

- 16) Choose "Add new linguistic type" from the Type menu.
- 17) Type name: call it "transcription"
- 18) For Stereotype choose "Symbolic Association"
- 19) Click "Add" then "Close"
- 20) Choose "Add New Tier" from the Tier menu.
- 21) Name your new tier "Interviewer"
- 22) For Participant, put in the interviewer 's initials
- 23) For Annotator, put in your initials.
- 24) For Parent Tier, select "default"
- 25) For "Linguistic Type" choose "transcription" (as you named it in Step 17)



- 26) Select the language/alphabet that you will transcribe in for "Default language." Use IPA-96 SAMPA for Faetar.
- 27) Click "Add."
- 28) ("More options..." lets you do things like color-code your tiers. Play if you like.)
- 29) Follow these steps to create a tier for each other speaker in the recording, naming the tiers with the Speaker Codes.
- 30) Click "Close."

You will need to create a daughter tier for the English translation. To do this:

- 31) Choose "Add new linguistic type" from the Type menu.
- 32) For type name: call it something useful like "speaker-translation"
- 33) For Stereotype choose "Symbolic Association"
- 34) Click "Add" then "Close."
- 35) Choose "Add new Tier" from the Tier menu.
- 36) Name the tier "Speaker-English."
- 37) For Parent Tier, select "Speaker" (or the Speaker's codename).
- 38) For "Linguistic Type" choose "speaker-translation" (as you named it in Step 32)
- 39) Click "Add."
- 40) Repeat these steps for any other Faetar speakers. (You don't need to translate the Interviewer's speech.)
- 41) Click "Close."

Associate the phrases/clauses/speaker turns to the right Speaker

- 42) Go back to the first annotation in the Default tier, **play it (highlight, then shift+space**, optionally first selecting **Loop Mode**) to hear which speaker it is, then double-click on that speaker's tier (with the segment still highlighted). This will create a new Annotation field in the appropriate tier. ***It will always be the same length as you created in the Default tier. To change its length/location, change it in the Default tier.***
- 43) Click and play the next Annotation segment in the Default tier and repeat.
- 44) Once you have Annotation fields in the right tier for a good chunk of the interview (or all of it), go back to the beginning and start to transcribe in each field.

Transcribe and Translate

- 45) Transcribe the first field in the Speaker tier, and then move to the next field. Repeat. Use IPA symbols.

After transcribing for a while in Faetar, go back and fill in the English translations in the "Speaker-English" tier. To do this:

- 46) Click the check-box at the left of the ELAN window (next to the pull-down menu showing all of your tiers). This will make all the tiers show up in a table-format. (To see this, you need to have selected "Grid" from the view options across the top of the window.)
- 47) You can type in the translations in the "speaker-English" column right next to each Faetar annotation. (You can select annotations and use the Play/Pause buttons, etc.)
- 48) If time allows, transcribe other speakers in the interview as well.

Tips:

- **Click shift+space**, optionally first selecting Loop Mode, to hear a highlighted segment.
- **Click Control+space** to start playing wherever the cursor is.
- You might find that you need to slightly **move some of the Annotation boundaries**, especially if you play the file fast to divide it up. You can do that by clicking the annotation and, when it turns blue, option+mouse click the edge of the Annotation that you want to move, and drag it to the right place.
- You can also **delete an annotation**: option+D
- You can **Merge an Annotation** with the next one (if you clicked and made a division that you later decide you don't want): Choose "Merge w/ next Annotation" in the Annotation menu.
- **Move to the next field by clicking return, then option+right arrow.**
- **Double-click to open an Annotation box** to type in.

Of course, you can use the mouse or the menus, but these key-combos will be quicker.

General rules for transcribing for this project.

Break up the recording into Annotation segments. Normally make 1 Annotation per sentence or turn, but that can be modified later, so don't worry about it a lot.

You may find the Shortcut Keys (section 7.2. of the ELAN manual – online and downloadable PDF) helpful.

Transcribe word for word what is said. Use International Phonetic Alphabet symbols by selecting "IPA-SAMPA" as the Default Language for the Annotation tiers in which you type Faetar. Select "English" for the tiers in which you translate to English.

Do not change the word order, etc., of what was said. Do not make corrections of what the speaker said.

Use periods, question marks, and exclamation points to mark the end of a sentence.

Hyphenate multi-word proper nouns like "UNIVERSITY-OF-TORONTO."

Do not type in any last names of personal acquaintances or relatives that might get mentioned during the interview. Just put an initial, e.g. "Maria G." For any other proper nouns that might identify a participant, type "[REDACTED]".

Use ALL CAPITAL LETTERS to transcribe ENGLISH WORDS.

Spell out any numbers the speaker says (like years).

Don't worry about transcribing pauses, etc. They are easy to find automatically with ELAN.

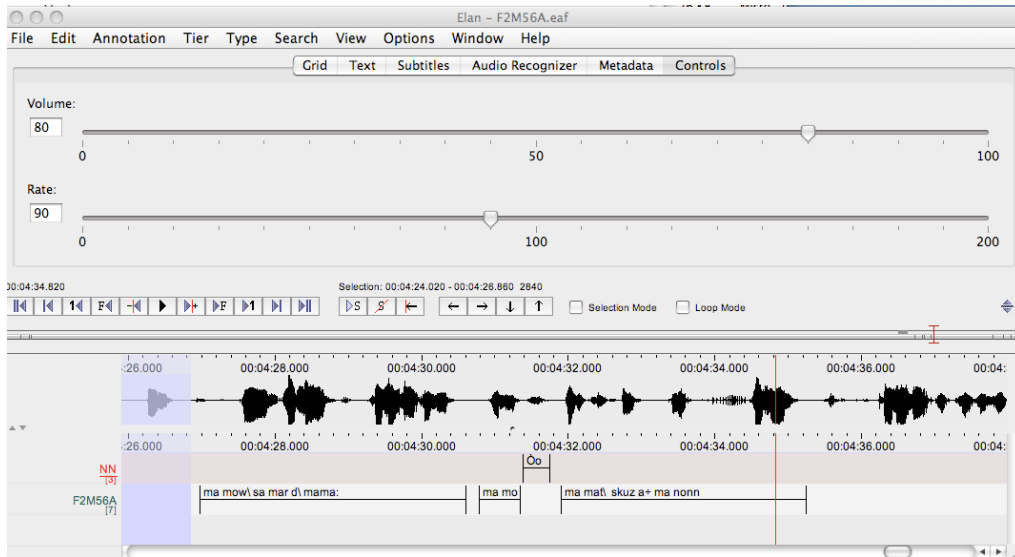
Wear headphones to transcribe. You'll hear better and get distracted less (and distract others less). There are some in the lab.

Take a short break every 10 minutes or so so you don't burn out.

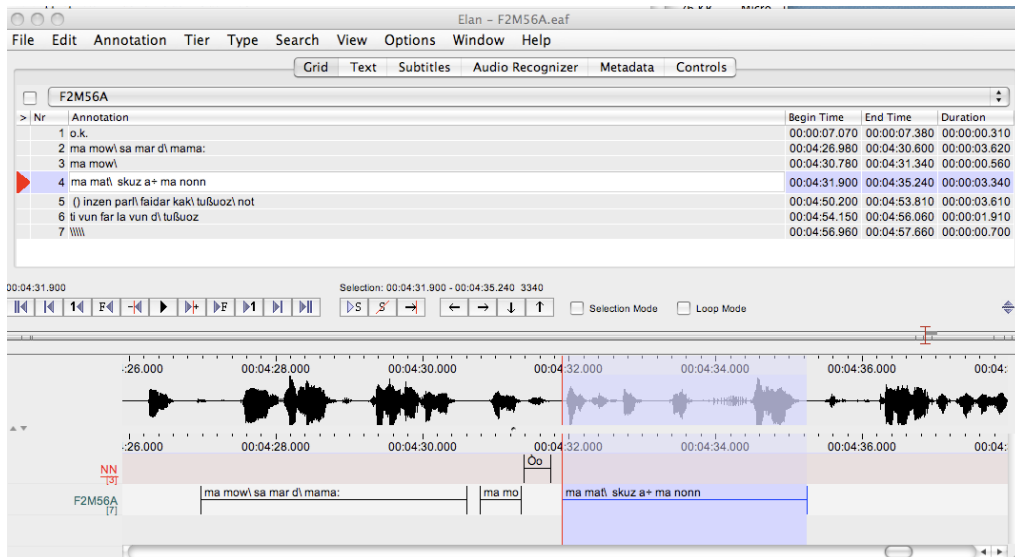
Back-up everything to iMac1 at the end of the workday (since it gets backed up). Put it in the appropriate LANGUAGE folder inside HLVC.

What ELAN can look like

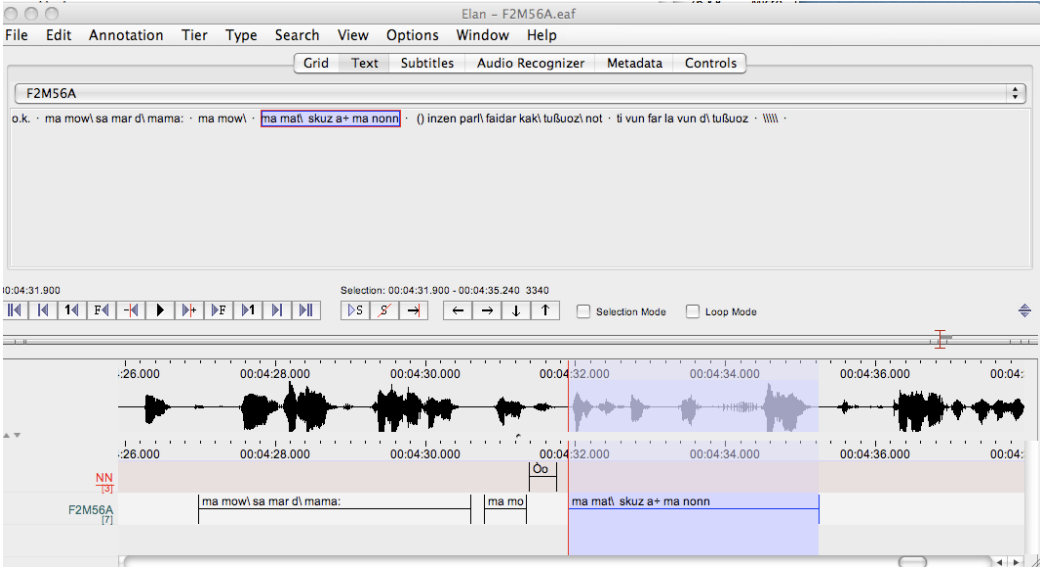
Top window with “Controls” selected:



Top window with “Grid” selected:



Top window with "Text" selected:



Transcribing Faetar in the SAMPA system

"Mini." = Vincenzo Minichelli's *Dizionario francoprovenzale Celli di San Vito e Faeto* system, similar to Italian spelling.)

"IPA" = the International Phonetic Alphabet symbols, which will show up in ELAN transcription.

"SAMPA" = what to type in ELAN to get each IPA symbol.

In this system, each symbol always represents exactly the same sound.

Each sound has its own symbol.

Red indicates ones that are tricky/different from Italian.

The Faetar and Italian columns show an example word for each sound.

<u>Mini.</u>	<u>IPA</u>	<u>SAMPA</u>	<u>Faetar</u>	<u>Gloss</u>	<u>Italian</u>	<u>Gloss</u>
Le vocale						
a	a	a	lu man	<i>the hand</i>	la mano	
e, é	e	e	e	<i>is</i>	è	
è	ɛ	E	l ombrél	<i>umbrella</i>	l'ombrella	
i	i	i	lu win	<i>wine</i>	il vino	
i	ɪ	I	pɪʃərl:ə	<i>small</i>		I APERTA
o, ò	o	o	lu orə	<i>garden</i>	l'orto	
o, u	u	u	(v)un	<i>one</i>	uno	
ó	ɔ	O	la pórðə	<i>door</i>	la porta	
u	ʊ	U	la zúppə	<i>soup</i>	la zuppa	
--	ə	@ (shift+2) option+e (type before the vowel)	vəniə	<i>to come</i>		VOCALE INDISTINTA
`	´		vəniə	<i>to come</i>		INDICA LA SILLABA TONICA
Le consonante						
b	b	b	bun:	<i>good</i>	buona	
c	k	k	lu kəvəl	<i>horse</i>	lu cavallo	
c	tʃ	tʃ	tʃél:ə	<i>Celle</i>	Celle	
d	d	d	ditʃémbərə	<i>December</i>	dicembre	
f	f	f	fa	<i>do</i>	fa	
g	g	g	grétʃi	<i>Greci</i>	Greci	
g	ɣ	G	la yát:ə	<i>cat</i>		FRICATIVA VELARE
g	dʒ	dʒ	dʒónə	<i>young</i>	giovane	
gl	ʎ	L	ʎo	<i>yes</i>	agl ^l io	<i>garlic</i>
h	--		(muta)			
j	j	j	lu jók:ə	<i>snow</i>	ieri	<i>yesterday</i>
l	l	l	lu lívrə	<i>book</i>	il libro	
m	m	m	la mæliə	<i>wife</i>	la moglie	
n	n	n	nu	<i>no</i>	non	
n	ŋ	N	síŋkə	<i>five</i>		NASALE VELARE
nj	ɲ	J	dɪŋjé	<i>in</i>		NASALE PALATALE
p	p	p	pɪʃərl:ə	<i>small</i>	piccolo	
r	r	r	ruz	<i>red</i>	rosso	
s	s	s	sa səráw	<i>sister</i>	la sorella	<i>sister</i>
s	ʃ	S	lu je	<i>bird</i>		FRICATIVA PALATALE SORDA
t	t	t	la tét	<i>head</i>	la testa	
v	v	v	vəniə	<i>to come</i>	venire	
u	w	w	lu win	<i>wine</i>	quattro	<i>four</i>
z	dʒ	dʒ	na dzik	<i>little (n.)</i>	lo zaino	<i>backpack</i>
s	ʒ	Z	ruz	<i>red</i>		FRICATIVA PALATALE
written double	:	:	pɪʃərl:ə	<i>small</i>		CONSONANTE GEMINATE

Faetar_transcription.doc

In ELAN, there are 2 options for setting up to type Annotations in IPA. You can choose either of them by Control-Clicking on an Annotation box or in the **Tier menu > Change Tier Attributes > Default Language.**

If you choose **IPA-96 (rtr)**, you type a character that *looks like* the one you need. A menu pops up and you use arrow keys or the return key, BUT NOT the mouse to select the right character. You have to do this even for “normal” letters. (RTR=Roman Typographic Root)

IPA character	Phonetic Descrip.	typog. based on ...	RTR char.
ð	Voiced dental fricative	Latin capital d with stroke	D
ɣ	Voiced pal. lat. approx.	Turned Latin y	y

If you choose **IPA-96 SAMPA**, you use various key-combinations to type characters other than normal alphabet characters. These are shown on a pop-up keyboard (but one w/ British layout). For example, shift+2 = schwa, but they map shows schwa on the apostrophe key because that is where “@” is on the British keyboard, but on a USA keyboard it’s with “2”. MOST keys are where you expect them to be, however.



? See also <https://archive.mpi.nl/tla/elan/ipa-extension>

How to edit IPA

When you want to enter the International Phonetic Association (IPA) character set with a computer keyboard, you need to know how to enter each IPA character with a sequence of keyboard strokes. This document describes a number of techniques.

The complete SAMPA and RTR mapping can be found in the attached html documents. The main html document (ipa96.html) comes in a pdf-version (ipa96.pdf) too.

1 How to use SAMPA for editing IPA

The Speech Assessment Method (SAM) Phonetic Alphabet has been developed by John Wells (<http://www.phon.ucl.ac.uk/home/sampa>). The goal was to map 176 IPA characters into the range of 7-bit ASCII, which is a set of 96 characters. The principle is to represent a single IPA character by a single ASCII character. This table is an example for five vowels:

Description	IPA	SAMPA
script a	ɑ	A
ae ligature	æ	{
turned a	ɐ	6
epsilon	ɛ	E
schwa	ə	@

A visual representation of a keyboard shows the mapping on screen.

The source for the SAMPA mapping used is

"Handbook of multimodal an spoken dialogue systems", D Gibbon, Kluwer Academic Publishers 2000.

2 How to use X-SAMPA for editing IPA

The multi-character extension to SAMPA has also been developed by John Wells (<http://www.phon.ucl.ac.uk/home/sampa/x-sampa.htm>). The basic principle used is to form chains of ASCII characters, that represent a single IPA character, e.g. This table lists some examples

Description	IPA	X-SAMPA
beta	β	B
small capital B	ɓ	B\
lower-case B	b	b
lower-case P	p	p
Phi	φ	p\

The X-SAMPA mapping is in preparation and will be included in the next release.

3 How to use RTR for editing IPA

IPA characters are *typographically* based on Latin and Greek characters. If an IPA character is derived from a Greek symbol, the RTR character is the equivalent Latin symbol of the Greek one. Therefore, all IPA characters can be typographically mapped to a Latin character. This method is called "Roman Typographic Root" (RTR).

Because the mapping is unambiguous, there is exactly one RTR-character for each IPA character, while more than one IPA character is allowed to map to the same RTR character. When you type the RTR-character, you will get a list from which you can select your IPA-character.

Note that the RTR-mapping is case-sensitive, and that there is no *phonetic* relation between the IPA- and its RTR character in general

The mapping from Greek to Latin letters used is

Greek letter	Latin letter
Capital Phi (ϕ)	F
Low case Beta (β)	b
Low case Gamma (γ)	g
Capital Omega (Ω)	O
Low case Theta (θ)	t

Most of the RTR mapping is obvious. This table lists the cases where some additional information is helpful.

IPA character	Phonetic Description	typographically based on ...	RTR character
δ	Voiced dental fricative	Latin capital d with stroke	D
λ	Voiced palatal lateral approximant	Turned Latin y	y
ɟ	Voiced palatal implosive	Dotless Latin j	j
ɔ	Bilabial click	Latin capital o	O
ɤ	Near-close, near-back rounded vowel	Turned Greek Capital Omega	O

Note that in general the phonetic value of the IPA character and its RTR-character is not related. With RTR, it is not intended to relate IPA characters to Latin characters by phonetic value. When you want to get to the RTR-character, you should take the viewpoint of a person which is not aware of the phonetic value of an IPA character, but which is strictly looking at the shape, the typography, of the IPA character.

3.1.1 Restriction to official IPA-96

The software restricts the input of IPA characters to the official IPA-96 character set. Withdrawn or superseded characters are not supported. This restriction is due to limited resources and may be lifted at request.

3.1.2 IPA Vowels

The ligatures (ae, oe) are related only to their first letter (a, o).

3.1.3 IPA Consonants

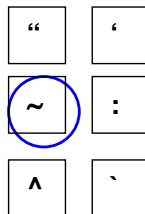
A number of IPA consonants map to the keyboard key question mark (?), the exclamation point (!) and the vertical bar (|), also called pipe.

3.1.4 Combining Diacritics

In the IPA alphabet, there are a large number of Combining Diacritics. Combining Diacritics are small signs that are appended or attached to an IPA character. The German umlaut is an example for a diacritic that can be attached to the vowel a o and u, leading to ä, ö and ü. It is very easy to enter combining diacritics.

Given an IPA character, which is symbolised below as a blue circle, there are five places where you can attach a combining diacritic. For each position, you press the keyboard symbol which is shown in one of the five boxes shown below. You will get a list of all diacritics that are available at this position.

Positioning command		Position of the diacritic	Examples
“	quotation mark	above	ö õ
’	Apostrophe	in the upper right	o' oʷ oʰ
:	Colon	to the right	o:
`	Grave accent	in the lower right	
^	circumflex accent	under	
~	Tilde	on the character itself	



Example:

If you want to add an umlaut to the character "a", you type first the a, then the quotation mark and then select the umlaut from the list that pops up. The umlaut will be set above the character a, resulting in ä.

4 Restrictions, Limitations and odd behaviour

- Please send complaints to Markus.Kramer@mpi.nl
- On Windows 98, you cannot display both Chinese and IPA characters.
- You cannot use the mouse for the lookup window: it will disappear.
- You cannot stack diacritics.
- On Windows 2000, when you use the international keyboard, you have to type <space> after the quotation mark (") and the accents (', ` , and ^).
- The visual keyboard you get in Cyrillic and IP-96 (SAMPA) has the layout of a standard United Kingdom keyboard. The standard UK mapping is depicted in the lower left and upper left corner of each key.

5 Wishes, Improvements

Please send a note on what you wish to be improved to Markus.Kramer@mpi.nl

The current wishlist is:

- (1 wisher) option: lookup list pops up horizontally.

First Words
Faetar_transcription.doc

Put the items in Column 1 in the “Label” tier, and the Faetar transcription in the “Speaker” tier. (If you have a separate .wav file just for the FW task, you don’t need to put “FW-“ in the start of each label.)

For now, skip the picture descriptions.

Label	Common Faetar word	Other Faetar words
Page 2-3		
FW-Daddy	mun baij	
FW-Mommy	ma mar	
FW-boy	lu kwattra	
FW-girl	la fij	
FW-baby	lu kwadrakiot	lu bambola, lu bambelot, l infannə
FW-dog	lu tfin	
FW-cat	la attə	
Page 4-5		
FW-undershirt	la maʎettə	
FW-underwear	la mutandə	la mutandinə
FW-shoes	lu ʃkarpej	
FW-socks	la kauzzət	
FW-pants	lo kauzun	
FW-T-shirt	la maʎə	lə tifert
FW-sweater	la maʎə	
Page 6-7		
FW-bread	lu pan	
FW-milk	lu ley	
FW-eggs	loz Iy	
FW-apple	la pumələ	
FW-orange	l arantʃ	
FW-banana	la banan	
Page 8-9		
FW-table	lu tawolinə	
FW-chair	la sədʒ	la sedjə
FW-plate	lu kartolleh	lu piattə
FW-knife	lu kuttej	
FW-fork	la brokkələ	
FW-spoon	la kəʎijə	
FW-cup	la tatz	
Page 10-11		
FW-horse	la tfuwa	lu kavallə, la kavallə
FW-sheep	la peggurə	
FW-cow	la vatʃ	
FW-hen	la/lə dzəlinnə	
FW-pig	lo kaijunnə	
FW-train	lu tren(nə)	
FW-blocks	lo daddə	lo dʒa, lo dʒokatələ, lu blokkə

First Words, continued
Faetar_transcription.doc

Page 12-13		
FW-Grandma	ma mowə	mamanonnə
FW-Grandpa	sun biaran	papanon, piaran
FW-slippers	lo pantuofl	loz ambit
FW-dress	lu vestit	la vest
FW-coat	la dzakettə	lu kott, lu kapot
FW-hat	lu tfappej	lu zappeɣ
Page 16-17		
FW-car	la makkənə	
FW-bicycle	la bifiklet	
FW-truck	lu kamiontʃin	lu kamiən, lu kamiə
FW-bus	la korierrə	
FW-airplane	la parəkiə	l aeroplan
FW-house	la kas	la kaz
p. 18-19		
FW-ice cream	lu dzəlat	
FW-cake	la pitsa dauzə	la tortə, lu dauws
FW-balloon	lu pallun	lu palluntʃin
FW-clock	lu rəldz	
FW-fish	lu pəfunnə	
FW-cookies	lo bifkot	
FW-candy	lu karamellə	la zamandolə

The phonetic symbols shown here are typed on a Mac using the IPAPhon font.