

Guidelines for Transcribing Arabic Dialects

Bushra Zawaydeh, Dave Stallard, John Makhoul
bzawaydeh@hotmail.com, stallard@bbn.com, makhoul@bbn.com
BBN Technologies

<i>Author</i>	<i>Version number</i>	<i>Date of modification</i>	<i>Modification</i>
Bushra Zawaydeh	Initial release	6/10/02	
Bushra Zawaydeh	2	7/1/2002	Revised the codes, and the transcription guidelines
Bushra Zawaydeh	3	7/4/02	Revised the codes, and the transcription guidelines
Bushra Zawaydeh	4	9/6/02	added more examples. Reviewed session on glottal stop
Bushra Zawaydeh	5	9/12/02	revised document after receiving comments from John Makhoul
Bushra Zawaydeh	6	10/21/02	revised document after discussing it with John Makhoul, Mohammad Mamouri, and Dave Stallard
Bushra Zawaydeh	7	11/11/02	revised document after meeting with John Makhoul, Dave Stallard, and Razi Diab
Bushra Zawaydeh	8	1/8/03	added common mistakes section
Bushra Zawaydeh	9	12/18/03	reviewed the document after reviewing the transcriptions which were totally finished.

1	INTRODUCTION.....	3
2	TRANSCRIPTION CODES	3
2.1	TRANSCRIBING NON-SPEECH SOUNDS	3
2.2	TRANSCRIBING LAUGHTER	3
2.3	TRANSCRIBING COUGHING	3
2.4	TRANSCRIBING MOUTH NOISES	4
2.5	TRANSCRIBING FILLER WORDS.....	4
2.6	TRANSCRIBING YES SOUND	4
2.7	TRANSCRIBING ASIDES	4
2.8	THE CODE DELETE.....	4
2.9	OTHER ISSUES	4
3	TRANSCRIBING ARABIC SPEECH.....	5
3.1	SOUND SUBSTITUTIONS	5
3.2	TRANSCRIBING HAMZA (GLOTTAL STOP).....	6
3.3	DIACRITICS, DOTS, PUNCTUATION MARKS	10
3.4	USING THE UNDERSCORE IN MULTIWORDS.....	11
3.5	TRANSCRIBING COLLOQUIAL WORDS AS CLOSE AS POSSIBLE TO MSA WORDS	12
3.6	TRANSCRIBING WORDS THAT END WITH A DOUBLED LETTERS	12
3.7	TRANSCRIBING PREFIXES.....	13
3.8	TRANSCRIBING NUMBERS.....	13
3.9	TRANSCRIBING DIALECT SPECIFIC WORDS	14
3.10	TRANSCRIBING VOWELS/ BLOOD TYPE.....	14
3.11	TRANSCRIBING FOREIGN WORDS.....	15
4	MORE EXAMPLES.....	15
5	COMMON MISTAKES.....	16

1 Introduction

In this document we review the guidelines that we have decided upon so far, for transcribing Arabic speech. The dialects that we will be transcribing are the Levantine dialects: Jordanian, Syrian, Palestinian, and Lebanese Arabic.

The transcription will be done in Arabic. We will be using codes in English to mark non-speech and certain speech sounds. The codes that will be used are discussed in section 2. Any Arabic word that occurs in Modern Standard Arabic (henceforth, MSA) will get the transcription of MSA. The transcription guidelines can be found in section 3.

The transcription is done using Dongxin Xu's transcription tool Dongscriber. Currently the tool shows the waveform, and it has a window in which the transcriber can write in Arabic and English what was heard in the waveform. One of the features of the tool is that it has a categorizer, which allows the user to categorize words into the following:

- MSA words, MSA names
- Colloquial Arabic words.
- Foreign words, Foreign names

The tool also allows the transcriber to segment words that start with certain prefixes. This is discussed in section 4.

2 Transcription Codes

In this section we will review the codes that we use in the transcription. All of the transcription is done in Arabic. Only the codes in this section will be written in English. **Make sure that you put the non-speech codes exactly where you hear them. If you don't hear any noise or stuttering, or any breathing, do not put anything.**

2.1 Transcribing Non-speech Sounds

The following are transcribed by giving them a NONSPEECH code. Nonspeech will be used for noise in general.

1. Unintelligible speech.
2. Background speech such as people talking.
3. Background noise such as music.
4. Microphone noise.
5. breathing/ sniffing
6. Clearing throat.

2.2 Transcribing Laughter

1. We use the code LAUGH

2.3 Transcribing Coughing

2. We use the code COUGH

p.s. coughing rarely occurs in our data. I encountered it only once so far.

2.4 Transcribing Mouth Noises

The following are given the code MOUTH:

1. Stuttering, if the stuttering is segmental, for example if the person is stuck on a certain segment. However, if the person repeats whole words or phrases, transcribe the phrases. Do not use the code MOUTH. Use the code MOUTH only when it is either not speech, or it is segmental, unintelligible, and hard to decipher stuttering.
2. If the person stutters by saying “bi EH”, just use the code MOUTH. Don’t transcribe “bi EH”.

2.5 Transcribing Filler Words

Filler words are considered speech. We will give them the following codes: AH, EH, UM, EM. If somebody says “AAAA” you can transcribe it as AH. We limit the filler utterances to the four codes AH, EH, UM, EM only.

2.6 Transcribing Yes Sound

Use UHUM when the person uses that sound to mean yes.

2.7 Transcribing Asides

An aside is when the speaker talks to somebody else, rather than answering the question. In this case we use the code AS. We will use the back slash \ when it is the left part of the phrase that is the aside. We will use the forward slash / when it is the right part of the phrase that is the aside. For example:

I don’t know what this means\AS five years ago
Five years ago AS/what does this mean

2.8 The code DELETE

We will use the code DELETE if the wave file should be deleted. In this case the operator (interviewer) failed to skip the question.

2.9 Other issues

Here are some issues that came up in our discussions, and how we deal with them.

- Silence: do not transcribe silence.
- Speech cut off. We do not transcribe it.
- Slips of the tongue. If somebody said “bog” and they meant to say “dog”. **Transcribe what they actually said**, not what they intended to say. The difference will be dealt with in the translation module.
- If the speaker makes a grammatical error, do not transcribe the correct form. Transcribe exactly what they said.
- Words in a foreign language like English: For example, if the speaker uses the word "like" while talking in Arabic. Just transcribe it in Arabic. Hence, write it as “لايك”. The foreign words will be categorized as foreign by using the categorizer.

3 Transcribing Arabic Speech

In transcribing Arabic, we decided that we would try to minimize the number of differences between the dialects by transcribing the allophones as their underlying phonemes. We also try to transcribe by using the MSA spelling of words as much as possible. In this section you will find the decisions that we have agreed upon so far.

3.1 Sound Substitutions

We will try to minimize the number of differences between the dialects by transcribing the allophones as their underlying phonemes. We do that especially in cases where the phonological rule that is applying is a sound substitution rule. In this section you will find a table that shows the sound substitution rules and how we transcribe words that have them.

Here is a table that shows the dialectal vs. MSA sounds. In the transcriptions, we will transcribe by using the MSA sound, i.e. we will use the sound in the leftmost column. In the examples column, slashes are used to indicate the underlying phonemic representation of the word. Brackets are used to illustrate the dialectal pronunciation of the word (allophonic). The word that we use in the transcription is the Arabic word that corresponds to the English word that appears in the SLASHES, not the brackets.

MSA Sound (underlying phoneme)	Corresponding dialectal sounds (allophones)	Example
/q/ ق	<ul style="list-style-type: none"> [g] Bedouins in Jordan [ʔ] Ammani-Jordanian [k] in Ramalla Palestinian 	“before” /qabel/ قبل [gabel] [ʔabel] أبل [kabel] كبل
/k/ ك	<ul style="list-style-type: none"> [tʃ] Rural Jordanian 	“all of it” /kullo/ كلو [tʃullo] تشلو
/θ/ as in three ث	<ul style="list-style-type: none"> [t] 	“three” /θalaθa/ ثلاثة [talate] تلاتة
/ð/ as in the ذ	<ul style="list-style-type: none"> [z] in Jerusalem and Jaffa 	“the atom” /ðarra/ ذرة [zarra] زرة
	<ul style="list-style-type: none"> [ð] (emphatic thaa) ظ in the rural areas in Jordan 	“this” /haða/ هذا [haða] هاظا
	<ul style="list-style-type: none"> [d] 	/ðan/ دان [dan] دان
		/haða/ هذا

/ḏ/ ظ	<ul style="list-style-type: none"> [Z] (emphatic [z]) 	[hada] هدا “ice cream” /buḏa/ بوظة
/D/ ض	<ul style="list-style-type: none"> [Z] (emphatic [z]) 	[buZa] بوزة “officer” /Dabet/ ضابط [Zabet] زابط
	<ul style="list-style-type: none"> ظ [ḏ] 	/Dabet/ ضابط [Thabet] ظابط
	<ul style="list-style-type: none"> [d] 	“hit” /Darab/ ضرب [ḏarab] ظرب
/ž/ ج	[ʃ] ش in Lebanon	“in front of” /bwiž/ بوج [bwiš] بوش
/f/ ف	[t] ت	“their mouth” /famhom/ فهمم [tumhom] تمهم
/S/ ص	[Z] (emphatic [z])	“little” /Sayira/ صغيرة [Zyire] زغيرة
/ʔ/	/y/	/miʔt dinar/ مئة دينار [mit dinar] ميت دينار

3.2 Transcribing Hamza (glottal stop)

Transcribing words with the Hamza appears to be very difficult. We tried to use this as general guideline: by and large we will not be transcribing the Hamza that is on an alef if it is not pronounced. If the speaker pronounces the glottal stop, we will transcribe it. However, after the transcriptions were completely finished, and I reviewed the dictionary, I found that there were many problems. People didn't follow this guideline. They tended either not to write the hamza at all or to include it. Hence, I decided to go back to MSA, and write the hamza as it is written in MSA.

Below are some guidelines.

We transcribe the glottal stop in the following cases:

1. If the speaker pronounces the hamza, and the word has a hamza, then we have to write it:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
أشخاص	people
أشياء	things
أشكال	forms
أعضاء	parts
أعصاب	nerves

2. If the word had an underlying glottal stop, and in the speaker's dialect it appears without a glottal stop, write the word WITH a glottal stop. A lot of such cases happen when the hamza is on a Nabra, or it is at the end of the word whether on the line or on the alef. Examples:

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
سمرا	سمراء	tan
دوا	دواء	medicine
هنا	هناء	Hana' (proper name)
هيفا	هيفاء	Hayfa' (proper name)
لميا	لمياء	Lamya' (proper name)
شي	شيء	something
زايد	زائد	extra
بييدا	بيبدأ	begins
خافين	خائفين	afraid
دايقة	دايقة	anxious/ troubled
عيلتي	عائلتي	my family
تهجى	تهجئ	spell

Notice that in the above words, the glottal stop is at the end or it is written on a "Nabra", as we say in Arabic.

** Notice that we write the word like شيء by typing the ي followed by a ء. We don't use the key where the letter Z is, because that is not for a ي. It is for hamza on a Nabra.

3. In words that could be pronounced in two ways, one where the hamza would be written on top, and the other where the hamza would be written on the bottom, transcribe such words by spelling the word as it would be in MSA. For example, أخواني is pronounced by some speakers as أخواني and by others as إخواني. In such words, transcribe the word as it would be in MSA. In this case it would be: إخواني. Other examples are:

<i>Pronunciation 1</i>	<i>Pronunciation 2</i>	<i>Transcribed</i>	<i>Gloss</i>
أما	إما	أما	either
أنه	إنه	أما	that
أخواني	إخواني	أخواني	my brothers
أعمامي	إعمامي	أعمامي	my uncles

If the word is colloquial such as أُجِتْ, let us agree on transcribing it with the hamza on top. The important thing is consistency, and to make sure that every word has on transcription only.

أُجِتْ أُجِتْ أُجِتْ she came

4. For colloquial words that sound like they have a hamza at the beginning, spell them without the hamza. All the below have the alef at the beginning followed by an “n”. For these, we do transcribe the hamza because they rhyme with اِنْفَعَلَ and because there are MSA words that have that template.

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
اِنوَلدت	I was born
اِنْدَفنت	she was buried
اِنْجرح	he was injured
اِنْصابوا	they were hurt

Below are some MSA words that rhyme with the colloquial words above. Of course, the glottal stop has to be written because it is pronounced.

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
<i>MSA word</i>	
اِنْتظمت	I became a member
اِنْتفاخ	swelling
اِنْتحرت	she committed suicide
اِنْتهى	it ended

We do not transcribe the glottal stop in the following cases:

1. If the word is colloquial, and the glottal stop is not pronounced, but a long alef is pronounced, we transcribe the word without the glottal stop. We keep the long alef.

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
باخذ	I take
باكل	I eat
باكلش	I do not eat
باكلها	I eat it

If the word is MSA, and it has a glottal stop underlyingly, put the glottal stop. Examples:

مأدبا	Madaba (city in Jordan)
رأس	head
مرأة	female

2. Words that are pronounced with a glottal stop in the dialect, but don't have a glottal stop in MSA should be typed as they would be in MSA

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
إلي	لي	for me
إلو	له	for him
إلها	لها	for her

3. The below had to be revised after the transcriptions were completely finished. This was not totally abided by the transcribers. They had a problem with understanding how to use the Silent Alef. Here is what we told them to do:

- Transcribing the Silent Alef: Create an alef when the word looks awkward (ungrammatical) without it. Don't put hamza on the alef. Notice that in some of the examples below, if we don't put the alef, there will be ambiguity between the pronounced word with out the alef and another word, as in ريج which means "wind" and in the dialectal meaning it would mean "I get comfortable". Examples:

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
خمستيام	خمسة ايام	5 days
تلاتيام	ثلاثة ايام	3 days
أربعتيام	أربعة ايام	4 days
أربعتشهر	أربعة اشهر	4 months
تلاتتشهر	ثلاثة اشهر	3 months
ستتشهر	سنة اشهر	6 months
صحابه	اصحابه	his friends
صحابها	اصحابها	her friends
صبت	ا صبت	I got (a disease)
صوات	ا صوات	sounds
دخن	ا دخن	I smoke
ريج	ا ريج	I get comfortable
خوالي	ا خوالي	my uncles
خوان	ا خوان	brothers

Remember that if the speaker says the glottal stop, we would need to transcribe it as such. Example:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
ثلاث أيام	3 days

People didn't follow the above. They tended either to remove the alef if it was not pronounced, or put a hamza. Also, the problem why the above doesn't work is that while the Lebanese people might say "بدي ريج حالي" the Jordanians will say "بدي أريج حالي". Hence, it seems that it is the Lebanese who remove the alef, and the Jordanian tend to keep it.

Therefore, since we do not want multiple transcriptions for the same word, it is better to write the word with the Hamza. Hence, we would transcribe the words that have alef above as:

صحابه	أصحابه	his friends
صحابها	أصحابها	her friends
صبت	أصبت	I got (a disease)
صوات	أصوات	sounds
دخن	أدخن	I smoke
ريح	أريح	I get comfortable
خوالي	أخوالي	my uncles
خوان	إخوان	brothers

similarly, أيام would be written with a hamza.

4. Transcribing Yes and No utterances

- yes utterances that are pronounced in the colloquial way will be transcribed as آ.
- no utterances that are pronounced in the colloquial way, or the MSA way will be transcribed as لا. We do not distinguish between لا and لآ. It is also important to add here that we type the لا by typing ل and then typing آ to create a لا. We do not use the single character that appears on the keyboard where the English B letter is.
- if the person says “bala” to mean yes, transcribe it as بلى. If they say “bala” to mean without, transcribe it as: بلا.

3.3 Diacritics, Dots, Punctuation Marks

- Make sure that all words that have an underlying word final taʔ marbuta have the two dots written on top of it. The taʔ means that the word is feminine. For example:

<i>Transcribed</i>	<i>Gloss</i>
شاطرة	good (feminine)
شامية	Syrian (feminine)
شعرة	one hair (feminine)

If the word ends with the prefix that is a pronoun which means “his”, the word would be ending with a “haʔ” not a “taʔ” marbuta. Hence it doesn’t need two dots. Examples are:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
شعره	his hair
شعراته	his hair
عمره	his age

- Make sure that all words that have a word final yaʔ have the two dots underneath the yaʔ, as opposed to just the alef maqsura. For example:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
عندي	I have
عمري	my age
عيوني	my eyes
ضواحي	suburbs
العمى	blindness
الحمى	fever

- **Diacritics:** the only diacritics that we will need to use are the tanwin, and the shaddeh. We do not transcribe the short vowels such as the fatha, dameh, kasra, or sukuun.
- With regards to the tanween, if somebody pronounces a word with the tanween, always transcribe it. Examples:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
جداً	a lot
حالياً	currently
طبعاً	of course
دائماً	always
فعادةً	so sometimes

- Use the shaddeh only to disambiguate words, for example:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
درس	he studied
درّس	he taught
بدرس	I study
بدرّس	I teach

- No punctuation marks will be used in the Arabic transcription at all.

3.4 Using the Underscore in Multiwords

Names of cities, last names, first names, and foreign abbreviations that are multiwords could be joined by using an underscore. Examples:

<i>Transcribed</i>	<i>Gloss</i>
برج_البراجنة	Burj Il-Baraaajneh
صاقيّة_الجنزير	Saaqyet Il-Janziir
خان_دنون	Khaan Danuun
بيت_لحم	Bethlehem
أبو_منصور	Abu Mansuur
عبد_الله	Abdullah
عبد_النبي	Abd Il-Nabi
أم_آي_تي	MIT
بي_آتش_دي	Ph.D.

Notice that for the foreign abbreviations, such as Ph.D., we do not follow the English spelling in this case. Hence, we do not put a period after the “h”. We just transliterate the letters into Arabic and write the. Hence, بي_أتش_دي .

However, for spelling out the letters of a word, don’t use an underscore. Just spell the separate letters as words. Example, the word عالم is spelled as:

عين ألف لام ميم

3.5 Transcribing Colloquial Words as Close as Possible to MSA Words

- If the speaker says a word or a phrase which corresponds to an MSA word or phrase, but in the dialect the word is shorter, try to spell the word or phrase as close as possible to MSA. Here are some examples:

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
رح	راح	I will
صرلي	صار لي	I have been
صرلا	صار لها	she has been
ماني قاعدة	ما إني قاعدة	I am not staying

- If a sound is there underlyingly, write it. For example:

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
لما بحكوا	لما بـيـحـكـوا	when they talk

- Transcribe the word final third person singular pronoun as in MSA. Hence, the Lebanese or Syrian colloquial “عندو” would be spelled as “عنده”. Moreover, the colloquial “عندا” would be transcribed as “عندها”.

- Transcribe the definite article “al” as such, even if the speaker did not pronounce.

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
إتخت	التخت	the bed
إسحر	السحر	the magic
إسيارة	السيارة	the car
الباب	الباب	the door

3.6 Transcribing words that end with a doubled letters

- If the doubling refers to the possessive, as in “my” or “me” then transcribe the doubled “y”. Here are some examples:

Words where we transcribe the doubled” y”:

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
بيي	my father
خبي	my brother
عليي	on me

- If the doubling appears in pronouns, transcribe the pronoun with the doubled “y” letter. Function words such *هَي* should not be converted to the MSA form. We can transcribe it in the colloquial way, because the function words are a small set of words, so we can just add them to the lexicon. It is worth mentioning here that we will transcribe *هو* in the colloquial way as *هوي*.

- Words that end with doubled letters will be transcribed with the doubling of letters.

<i>Pronounced & Transcribed</i>	<i>Gloss</i>
حاسس	I feel
حاطط	I put

3.7 Transcribing Prefixes

We will write the prefix “**fm**” which refers to the progressive aspect in colloquial Arabic as a single word, not as an attached prefix. For example:

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
عمبقرا	عم بقرا	I am reading

The prefix “ha” which means “this” will also be written separately.

<i>Pronounced</i>	<i>Transcribed</i>	<i>Gloss</i>
هالحرب	ها الحرب	this war
هاالشغلة	ها الشغلة	this thing

The transcription tool has a categorizer, which enables the transcriber to mark words that start with “bi” which is a prefix that means “in”, the definite article “wa”, and the colloquial prefix **fa**. By marking, we mean that we will add a plus sign + after the bi, **fa**, wa. Here are some examples:

<i>transcribed</i>	<i>marked</i>	<i>Gloss</i>
ببيروت	ب+بيروت	in Beirut
وبيروت	و+بيروت	and Beirut
عبيروت	ع+بيروت	to Beirut

Thus, you should not put a + sign in verbs where the “bi” is a colloquial progressive prefix, as in:

بياكل	he is eating
بيمشي	he is walking

Notice that when we type, we do not write the +. The + will be marked through the categorizer part of the tool. Hence, the + signs will appear in the local dictionary.

3.8 Transcribing Numbers

Numbers should be written out fully. For example, 1980 would be transcribed in the equivalent words “ألف وتسع مئة وثمانين”. Hence, no digits should be used in the transcription

at all. Numbers between 11-19 that are pronounced in the colloquial way, will be transcribed as the following¹:

<i>Trascribed</i>	<i>Gloss</i>
احد عشر	11
طنعشر	12
ثلاثعشر	13
أربععشر	14
خمسة عشر	15
ستة عشر	16
سبعة عشر	17
ثمانعشر	18
تسع عشر	19

We will use the above transcription whether the speaker actually pronounced the final “r” or not. In the dictionary, at some point we will add pronunciations for words that have two pronunciations. The above words will be examples of such words.

3.9 Transcribing Dialect Specific Words

Lexical words that are dialect specific would be transcribed as they are. We do not convert them to MSA. For example, in Lebanese, they use the word for “father” [bayyi], in Jordanian, they use [abuuy], in Syrian, they use [abi]. We transcribe all these variations of the word father as they are uttered. We do not merge them to the MSA word [abi].

Other examples

- Syrian: ma watita ما واتتها (i.e. it didn’t agree with her)
- Lebanese: mafasita siyyara معستها سيارة (she was run over by a car)

Notice that in the first two examples above, we used the third person singular rule which we specified previously. In Syrian and Lebanese, speakers delete the “ha” finally, but in the transcription, we spelled the word with the “ha”.

3.10 Transcribing vowels/ blood type

- For example, a Syrian speaker answered that her blood type is A+. In Arabic she said the sound [a] as in apple and the word “ijaabi” which means “positive”. We would transcribe that as أ إجابي. We do NOT transcribe it as any of the other possibilities below:

- أ إجابي this is not correct
- +أ this is not correct

For blood types, we do not use the + and - symbols. We type up the word fully, whether it was spoken in Arabic, English, or French.

¹ We might change our transcription convention for spelling numbers between 11-19 to make it similar to transcribing numbers that are followed by a word like “الام” as in خمسة ايام. Hence we might spell خمسة as خمسة عشر.

we would only transcribe the أ if the speaker said ألف

3.11 Transcribing Foreign Words

Foreign words would be written in Arabic script. Each foreign word will receive one and only one possible spelling. We will maintain a lexicon that contains all the words that have been transcribed.

- Transcribing English drug names such as: Motrin and Pramperram. All foreign drugs will receive a specific spelling in Arabic, and that will be used every time the word is transcribed. موترين - برميران. If the word contains a sound that does not exist in Arabic such as the [p] or [v], we will use the [b] and [f] instead.
- Transcribing French words. Lebanese people use French in their speech frequently. For example, “accidant” for accident. In this case, we will transcribe French words in Arabic, for example :

□ أكسيدان

Notice here that in the French word “accidant”, the [t] and [n] have been deleted word finally, and the vowel [a] is heavily nasalized. We can transcribe this by restoring the [n]. I could imagine somebody uttering the [n] finally by saying [aksidan].

- Any foreign word that has the sound [g] should be transcribed with a jiim. Examples:
 - regularly ريجيولارلي
 - negative نيجاتيف

4 More Examples

Number	Dialect	What was actually spoken (phonetic transcription)	BBN transcription	Gloss
1	Lebanese	انا عزبا [ana ʕazba]	أنا عزباء [ana ʕazbaʔ]	I am single
2	Lebanese	بعيش بجبل حسين baʕiish bijabal hsen	بعيش بجبل حسين baʕiish bijabal hsen	I live in Jabal Hussein
3	Jordanian	الأم رأبتي وأصابع إجري وأصابع إيدي aaah raʔbti u ʔsabʕ ʔjrai u ʔsabʕ ʔday	[UM] رقبتي وأصابع إجري وأصابع إيدي [UM] raqbatu ua ʔsabeʕ ʔjrai wa ʔasaabeʕ iday	Ammm my neck, my toes, and my fingers
4	Lebanese	بيتي بوش مدرسة لفرير beti bwif madraset lifrer	بيتي بوجه مدرسة الفرير beiti bwij madrasat alfrer	My house is facing the Frer school
5	Jordanian	بيتي بوج مدرسة لفرير	بيتي بوجه مدرسة الفرير	My house is

		beti bwij madrasat lifrer	[bayti bwij madrasat alfrer]	facing the Frer school
6	Lebanese	جيت عام 1980 jiit ʕaam 1980]	جيت عام ألف وتسع مئة وثمانين ji?t ʕaam alf u tisi9miyye u Θamaaniin	I came in the year 1980
7	Jordanian	أجيت سننل 1980 ʔijiit sant al 1980	أجيت سنة الف و تسع مئة و+ثمانين ʔijiit sant alf u tsi9miyye u Θamaaniin	I came in the year 1980
8	Lebanese	بيي إسمو سمير bayyi ʔismo smir]	بيي إسمه سمير bayyi ʔismo smir	My father's name is Samir
9	Jordanian	أبوي إسمو سمير ʔabuuy ʔsmo samir	أبوي إسمه سمير ʔabuuy ʔsmo samir	My father's name is Samir
10	Syrian	أبي إسمو سمير ʔabi ʔismo samir	أبي إسمه سمير ʔabi ʔismo samir	My father's name is Samir
11	Syrian	أ إجابي æ ijaabi	أ إجابي æ ijaabi	A+
12	Lebanese	أ بوزيتيف e pozitiv	أ بوزيتيف e pozitiv	A+

Transcribing Specific Words:

- بكرة is transcribed as such as opposed to بكرة because one speaker said: بكرة الصبح which indicates that there is a ta? there.
- ذيني means “my ears”. I am transcribing all the derivations of the word “ear” with the interdental voiced fricative.
- علسوبرماركت should be transcribed: عالسوبرماركت. Do not transcribe any words as ...عل always convert them to their origin, which has the definite article.
- إخوة

5 Common Mistakes

Correct
ما يعرف
هذا
عالشغل

Incorrect
مايعرف
هادا
عالشغل

مرأتي	مرأتي
التقيؤ	التقيأ
الحمد لله	الحمد لله
عالظهر	عاضهر
بشمال صبرا	بشمالي صبرا
إخوة	أخوة
تيشيرت	تيشرت
أسبو عين	إسبو عين
أكون	كون
بظاني	بضلني
دغري	ضغري
خيات	خبيات
راسي	راسي
براسي	براسي
أنه	أنو
الظهر	الضهر
عالغذاء	علغذاء
إئذي	أذي
نيسكافية	نيسكافية
فش	فيش
بظل	بضل
ظله	ضله
هديك	هديك
كم	كام
و	should not occur by itself at all. Put MOUTH if people say it by itself with EH after it.
بأكل	بأكل
رأس	راس
بكرة	بكرا
بنادول	بانادول
قديه	أديه
هلق	هلاً

meaning now