

UNIVERSITY OF MARYLAND



CASL

CENTER FOR ADVANCED
STUDY OF LANGUAGE

Final Technical Report E.3.4 User Manual for AVIA Level 1 (v1.2)

Date of report: November 20, 2008

Information current as of: November 20, 2008

Contact Information:

Timothy Buckwalter

301-226-8984

tbuckwalter@casl.umd.edu

www.casl.umd.edu

Authors:

Timothy Buckwalter, Nathanael Lynn, Jeff Carnes, Bridget Hirsch, Jason White, Pamela

Kling

CDRL: A021

DID: DI-MISC 80508A

Contract No. MDA904-03-C0543

Table of Contents

Table of Contents	2
List of Figures	4
1. Introduction	5
1.1 How can AVIA Level 1 help me?.....	5
1.2 How does the AVIA define “dialect?”	5
2. Getting Started	7
2.1 Installing the AVIA Level 1	7
2.2 Web Server Installation	7
2.3 Desktop Installation.....	7
3. Navigating the AVIA	8
3.1 Home page	8
3.2 Parts of the AVIA	8
4. How to Effectively Use the AVIA	10
4.1 Introduction to Using the AVIA.....	10
4.2 The Word List.....	11
4.3 Morphological discriminators.....	13
4.4 Phonological discriminators	14
5. AVIA Level 1 Workflow	16
5.1 Suggested Workflow	16
5.2 Searching for a word (lexical discriminator)	16
5.3 Searching for a sound (phonological discriminator)	17
5.4 Searching for a morphological discriminator	17
5.5 Searching for word classes, dialects.....	18
5.6 Example of a full AVIA session	18

5.7	Potential Traps.....	20
	Appendix A. Troubleshooting Guide.....	21
	Appendix B. Glossary of Terms.....	22
	Appendix C. Dialect ID Worksheet.....	23
C.1	Worksheet.....	23
C.2	Worksheet Example.....	23

List of Figures

Figure 1. Map of the major cities for each of the 17 dialects in AVIA Level 1.	6
Figure 2. AVIA Level 1 welcome page.....	8
Figure 3. Parts of AVIA Level 1.	9
Figure 4. AVIA Level 1 word list.	11
Figure 5. Dialects that contain /ana/.	12
Figure 6. Sorting the Word List by the transcription column.	13
Figure 6. AVIA Level 1 morphological discriminator chart.	14
Figure 7. AVIA Level 1 phonological discriminator chart.	15
Figure 8. Workflow using the AVIA.....	16
Figure 9. Finding the word انا using Internet Explorer.	17
Figure 10. Finding the word انا using Firefox.	17
Figure 11. AVIA Level 1 list of words that contain /ž/.	18
Figure 12. AVIA Level 1 dialects that contain /kiif/ <i>how?</i>	19
Figure 13. AVIA Level 1 dialects that contain /kiifaš/.	19
Figure 14. AVIA Level 1 worksheet.	23
Figure 15. Example of using the AVIA Level 1 worksheet.	23

1. Introduction

1.1 How can AVIA Level 1 help me?

The AVIA Level 1, which will be from this point referred to simply as the AVIA, is designed to broadly identify Arabic **dialects**¹. It does this by providing a list of basic high frequency words and expressions which are given both in their written and auditory forms. You can then match what you hear in the audio cut with the words in the AVIA to determine what dialect you are hearing.

1.2 How does the AVIA define “dialect?”

The AVIA uses “dialect” in a very broad sense. Arabic dialects are notoriously complex. One country, or sometimes even a single city, may have a number of quite different dialects. You should therefore be aware that at this point the AVIA provides only a broad orientation for each dialect.

1.3 What methodology is the AVIA using to distinguish dialects?

The AVIA will help you use the *successive narrowing method* to identify dialects, which we will discuss in Section 5.2.

For example, when you hear the sound /ž/ (as in Section 4.1), there is a high likelihood that the dialect is Moroccan or Lebanese but that it is not Egyptian or Emirati since these two dialects do not have this sound. As a next step, you would then return to your audio cut and search for further clues, this time to help decide between Moroccan and Lebanese without worrying about Egyptian or Emirati.

1.4 How were the dialects chosen?

Dialects associated with major cities are usually the most dominant, widely spoken and understood in a country. Therefore, the dialects chosen for the AVIA are most likely linked to the largest cities in a country or region, often the capital city. The dialects and their correlating cities in the AVIA are shown in the map below:

Dialects in AVIA Level 1:

Moroccan (Casablanca)
Tunisian (Tunis)
East Libyan (Benghazi)
Egyptian (Cairo)
Sudanese (Khartoum)
Palestinian (Jerusalem)
Lebanese (Beirut)
Syrian (Damascus)
Jordanian (Irbid)
Hijazi (Jedda)
Yemeni (Sana^ca)
North Iraqi (Mosul)
Iraqi (Baghdad)
Kuwaiti (al-Kuwayt)
Emirati (al-^cAin)
Qatari (Doña)
Bahraini (Manama)

¹ Words in **bold** appear in the glossary in Appendix B.



Figure 1. Map of the major cities for each of the 17 dialects in AVIA Level 1.

Next Steps for the AVIA Tool

We hope in the future to add more detail, so that multiple dialects within a country can be differentiated. A step in this direction has already been taken in the case of Iraqi Arabic, where North Iraqi is distinguished from the Arabic of Baghdad, which is the most widely-used variety of Iraqi Arabic. You will notice that there are many differences between these two dialects alone, for instance:

	<i>North Iraqi (Mosul)</i>	<i>Iraqi (Baghdad)</i>
“children”	/ulaad/	/jahhaal/
“men”	/yjaal/	/rijjaal/
“how much”	/bašqad/	/šgadd/
“now”	/hassa ^C ta/	/hassa/

2. Getting Started

This section contains information on installing the AVIA. If you've already installed the AVIA, please go to *Section 3: Navigating the AVIA Level 1*.

2.1 Installing the AVIA Level 1

There are two ways that you may install the AVIA:

1. On a web server accessible to authorized users
2. On a desktop computer, by accessing the files from a:
 - CD-ROM
 - local hard drive
 - network file share

Please note: *Installing the AVIA Level 1 may trigger warning messages from your browser as explained below in section 2.3 Desktop Installation.*

2.2 Web Server Installation

There are no application server or database requirements for the AVIA. It can be installed on any standard HTTP web server such as Apache or IIS. The files and links are relative and should not need any special configuration except to insure that the directory index is set to allow *index.html* as a valid directory index file.

Please note: *As there is no application server or database component currently with the AVIA Level 1, if access control is desired, it must be configured at the web server level via HTTP Digest authentication or some other preferred method to control access.*

2.3 Desktop Installation

To view/use the AVIA from a drive or file share, insert the CD or copy the files to the desired location. Begin by opening the *index.html* file. You may navigate to any of the desired files from there.

Please note: *The JavaScript and embedded audio player employed in the AVIA Level 1 may trigger security warnings from the browser. These warnings are generic warnings that any JavaScript run on a web page in this fashion (not from a web server) will cause. In Internet Explorer, you may be prompted or warned about 'active content'. Allowing active content to run will allow the sound files to play. Depending on the group policies and browser settings of the machine, the audio links may be prohibited from working. If you have any questions or problems, it is always best to check with your local system administrator.*

3. Navigating the AVIA

3.1 Home page

When you load the AVIA, the first screen you will encounter is the home page. The home page is the homepage for AVIA Level 1 and can be accessed at any time by choosing “Home” on the toolbar as shown by the red arrow in Figure 2.

UNIVERSITY OF MARYLAND
CASL
CENTER FOR ADVANCED
STUDY OF LANGUAGE

Arabic Variant Identification Aid

Welcome to the Arabic Variant Identification Aid!

Home
About the AVIA
Transcription Guide
Phonological Discriminators
Morphological Discriminators
Level 1 Word List
By Dialect
By Word Class
User Guide

© University of Maryland
Center for Advanced
Study of Language

We have designed this Level One tool to help listeners quickly identify the dialects of Arabic that they hear. The tool consists of four parts: a list of our transcription symbols for Arabic; a table of phonological discriminators, a list of distinctive words in 17 Arabic dialects, and a table of morphological discriminators. We recommend that you start with the phonological discriminators; they should enable you to eliminate some possible identifications (ex. "If they're saying /L/ then they can't be Egyptians!") and make some guesses about what your speaker's dialect might be. Next, try the list of distinctive words; this should allow you to make a tentative identification. Finally, look at the morphological discriminators; they might provide you with extra evidence that your identification was correct. If you have any trouble understanding our transcription symbols, just look at our list of transcription symbols. Note, also, that our word list is searchable, like any Word document. To search for a term in either English or Arabic letters, Select "Edit" from your browser's menu bar. Then select "Find (on this page)", type in what you want to search for and press "Enter".

This tool was developed by the Center for Advanced Study of Language (CASL) under contract to the Laboratory for Language Sciences

© University of Maryland Center for Advanced Study of Language

Figure 2. AVIA Level 1 welcome page.

3.2 Parts of the AVIA

The AVIA consists of four basic parts:

1. **phonological discriminators:** A chart of three major phonological discriminators with a breakdown of the dialects.
2. **morphological discriminators:** The meaningful elements inside of a word. In AVIA, these “elements” are affixes added onto verbs..
3. **word list:** 56 common words and phrases, categorized by dialect and word group.
4. **transcription guide:** An explanation of the transcription system to assist with finding words in the list.

They are found on the scroll bar on the left hand side of the page as shown in Figure 3 below:

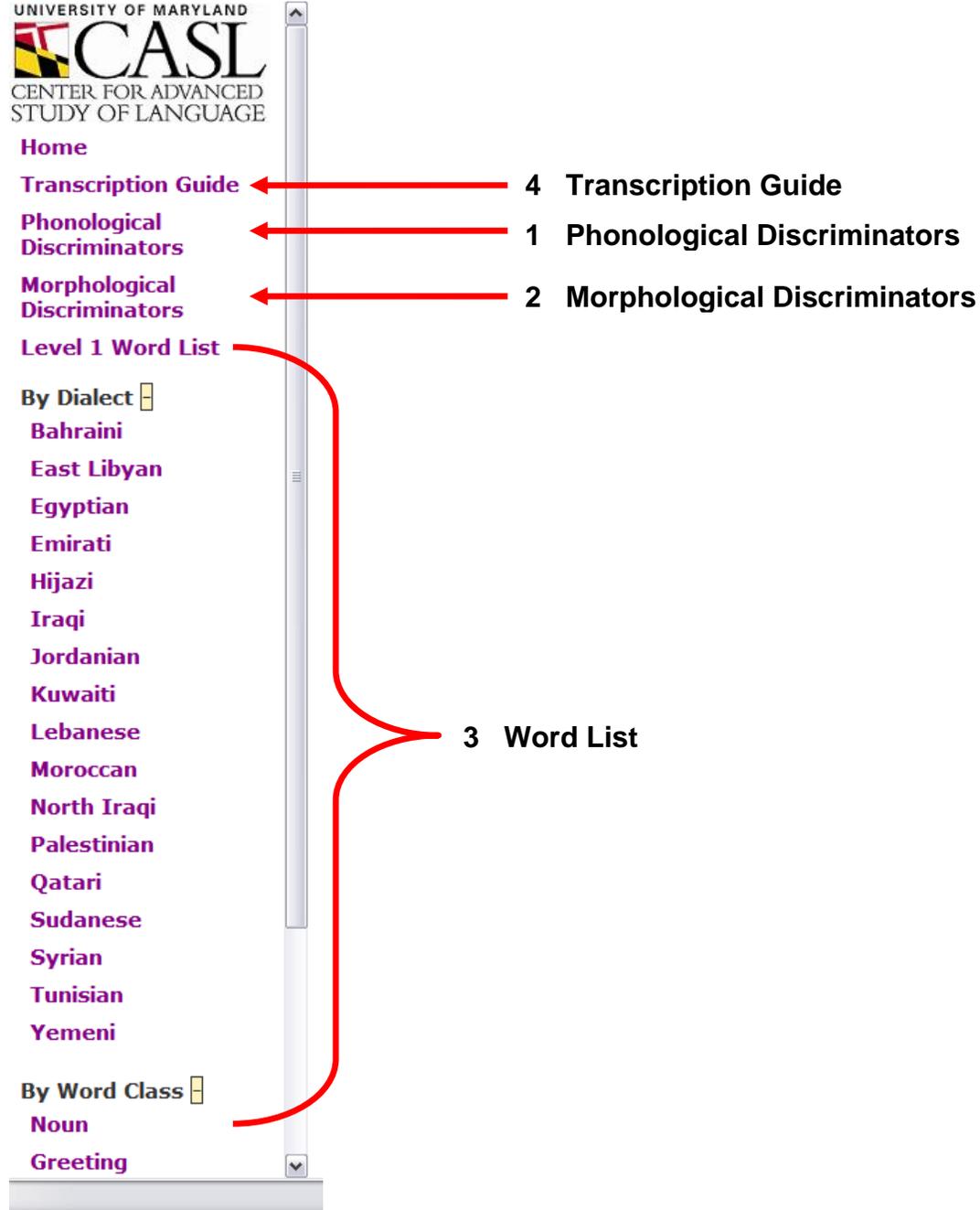


Figure 3. Parts of AVIA Level 1.

4. How to Effectively Use the AVIA

4.1 Introduction to Using the AVIA

The basic idea behind the AVIA is that you can determine the identity of the dialect you are listening to by linking individual elements that you hear in your audio cut to the distinctive features found in the AVIA and progressively narrowing down the dialects that fit those criteria. For instance, as soon as you hear the sound /ž/, there is a high likelihood that the dialect is Moroccan or Lebanese, but that it is not Egyptian or Emirati, since these two dialects do not have this sound. As a next step, you would then return to your audio cut and search for further clues, this time to help decide between Moroccan and Lebanese, without worrying about Emirati.

Here are general steps for using the AVIA:

1. Listen to your audio cut.
2. Write down any sounds, words, or morphological elements that you think might be distinctive. You can write these down in **SATTS**, in Arabic script, or in the modified **International Phonetic Alphabet (IPA)** system.
3. Go to the word, phonological discriminators, and morphological discriminators list to determine if these elements match a dialect or set of dialects.
4. Listen to the sound clips in the AVIA to compare what you heard in your audio cut to the AVIA.
5. Determine whether there is a unique dialect that contains this set of dialect discriminators.
6. Continue listening to your audio cut until you are able to narrow down the dialect to which you are listening from a region or group of dialects to one unique dialect.

Sections 4.2 to 4.4 will show you how to use the AVIA word list, morphological discriminators, and phonological discriminators in order to make more confident dialect identifications.

How did we choose our lexical discriminators?

Fifty-three words and three phrases were chosen for inclusion, based on their high frequency of use and occurrence in discourse and for their ability to distinguish dialects. For instance, the word “what?” is both frequent and occurs in a number of different forms (at least 12) in our sample of 17 dialects. Its different forms can thus serve to distinguish a number of dialects. On the other hand, we did not include a word like “see” since in most dialects this word is /šaaf/, so it would not serve as a good lexical discriminator. Note that we use the term “**lexical discriminator**” in accordance with widespread linguistic practice to mean “word discriminator.” In addition, three common greetings which might be heard at the beginning of conversations are included.

4.2 The Word List

You will see that the distinctive word list is arranged in six columns. In the first column, the words are organized alphabetically according to CASL’s modified IPA representation (see Section 4.5 below for a description of IPA). As you can see, the word list is arranged roughly in alphabetic order with respect to the English alphabet (not in order according to the Arabic alphabet). If you hear /kiif/, look up the word under “k.” The second and third columns give the transcription in SATTS and Arabic script respectively. The fourth column gives the English gloss while the fifth specifies the dialect. Finally, the sixth column gives the grammatical class of the word or “word class.” In the first column there is a  icon to click on and hear the pronunciation of the word.



TRANSCRIPTION	SATTS	ARABIC	ENGLISH	DIALECT	WORD CLASS
aana 	ANA	أنا	I	Moroccan	noun
aana 	ANA	أنا	I	Tunisian	noun
AAAna 	ANA	أنا	I	Bahraini	noun
AAAna	ANA	أنا	I	Hijazi	noun
AAAna	ANA	أنا	I	Kuwaiti	noun
AAAna 	ANA	أنا	I	Qatari	noun
aane 	ANA	أنا	I	East Libyan	noun

Figure 4. AVIA Level 1 word list.

The word list, or lexical discriminator chart, allows you to filter down what you are viewing according to a number of categories. You can filter according to:

1. **Transcription:** If you want to find all of the dialects which pronounce the word for “I” as /ana/, click on /ana/ in the transcription column (the first column), and you will pull up eight dialects, all of which have the same pronunciation for this word. These are the dialects in our data characterized by the pronunciation /ana/, and you can compare how they sound by clicking on the  icon next to /ana/.

TRANSCRIPTION	SATTS	ARABIC	ENGLISH	DIALECT	WORD CLASS
ana 	ANA	أنا	I	Egyptian	noun
ana 	ANA	أنا	I	Emirati	noun
ana 	ANA	أنا	I	Hijazi	noun
ana 	ANA	أنا	I	Jordanian	noun
ana 	ANA	أنا	I	Kuwaiti	noun
ana 	ANA	أنا	I	Lebanese	noun
ana 	ANA	أنا	I	North Iraqi	noun
ana 	ANA	أنا	I	Palestinian	noun
ana 	ANA	أنا	I	Qatari	noun
ana 	ANA	أنا	I	Sudanese	noun
ana 	ANA	أنا	I	Syrian	noun
ana 	ANA	أنا	I	Yemeni	noun

Figure 5. Dialects that contain /ana/.

2. **Dialect:** If you want to see only words from one dialect, such as Egyptian, just click on the word “Egyptian” in the “Dialect” column, and you will find all of the words from this dialect. Finally, if you want to find only the words of a single grammatical class, only question words, for instance, click on “question word” in the “word class” column, and you will find all of the question words from all of the dialects.
3. **Word class:** You can filter the word list by clicking on the relevant word within the Level 1 word list itself as shown above. You can also filter by clicking on the dialect in the scroll bar in the left-hand column. If you click directly on “Bahraini,” for instance, you will call up only words from this dialect.

Remember, if you want to see all of the words on the list, click on “Level 1 Word List” in the left-hand scroll bar.

The Sort Feature

The AVIA Word List (see Section 4.2) also contains a sort feature that allows you to search each column (Transcription, SATTs, Arabic, English, Dialect, Word Class) in alphabetical order. Click on the header of each column to sort:

UNIVERSITY OF MARYLAND CASL CENTER FOR ADVANCED STUDY OF LANGUAGE

Home
Transcription Guide
Phonological Discriminators
Morphological Discriminators
Level 1 Word List

By Dialect
Bahraini
East Libyan
Egyptian
Emirati
Hijazi
Iraqi

view normal version

Ignore Me

Dialect Word List - Egyptian

Click on a column header to sort by that column (or toggle the sort order)

Transcription	SATTs	Arabic	English	Dialect	Word Class
aani waahid		أني واحد	which one	Egyptian	question word
ahlan	A?LA	أهلاً	hello	Egyptian	greeting
ana	ANA	أنا	I	Egyptian	noun
awlaad	AWLAD	أولاد	children	Egyptian	noun
axbaar-ak eeh	AOBARK AI?	أخبارك آيه	what's up	Egyptian	greeting
b-surfa	BSR??	بسرعة	quickly	Egyptian	adverb/adjective
barDu	BRV??	برضة	also	Egyptian	adverb/adjective

Figure 6. Sorting the Word List by the transcription column.

In addition, you can sort in reverse alphabetical order by clicking twice on the column header or using the toggle:

Note that the English alphabet and the SATTs transcription do sort in alphabetical order but put the “special” characters at either the end or beginning of the alphabet. For the English alphabet, “z” is followed by the following characters: ُ, č, đ, ħ, š, ž, ə, ʃ, ɣ, θ. In SATTs, the “, :, ? appear before the letter “A.”

4.3 Morphological discriminators

Morphological discriminators are meaningful elements inside a word. Egyptian Arabic, for instance, as the phrase /ma šuft-úu-š/ “I didn’t see him.” The suffix /-uu/ “him” and /-š/ “negative” are elements added after the verb, which makes the whole word mean “I didn’t see him.” Some dialects have this negative /-š/, while others do not. In this version of the AVIA we have included a comparison of five morphological elements, which are all part of the verb:

- **Habitual action**—Some dialects mark this with the prefix /b-/, others do not mark it at all.
- **Ongoing action**—Some dialects mark this with a prefix like /^Cam-/.
- **Future tense**—Some dialects mark this with prefixes like /raħ/, /raaħ/ or /ħa-/.
- **Negation**—All dialects have the negative word /ma/. Some dialects add the suffix /-š/ at the end of the verb (see example above).
- **First person plural, imperfect verb**—Some dialects add the suffix /-u/ to the end of the verb, /nimš-u/ “We go.”

Morphological Discriminators Across 17 Arabic Dialects

	Habitual action marker (“I always see him.”)	Ongoing action marker (“I see him now.”)	Negation (“I didn’t see him.”)	Future tense marker (“I’ll see him tonight.”)	1st Person pl. in imperfect
Moroccan	/ki-/ /diima ki-nšuuŋ-u/ ديما كنشوفه	/ki-/ /ki-nšuuŋ-u fi daabe/ كنشوفه في دابه	/ma...-š/ /ma-šuuftüu-š/ ما شفتوش	/ʔaad/ /ʔaad nšuuŋ-u/ غاد نشوفه	/n...-u/ /nšuuftüu-ha/ نشوفوها

Figure 7. AVIA Level 1 morphological discriminator chart.

At this point the morphological discriminator chart is a static table and is not linked to audio files. We hope to change this in the future.

4.4 Phonological discriminators

The phonological discriminator chart gives three examples of major phonological differences among dialects. Each of these sounds—/g/, /ž/, and /č/—are not in Modern Standard Arabic (MSA) and therefore easier to pick out while listening. The phonological discriminator chart tells you, for instance, what dialects use the sound /g/. If you hear this sound, then there is a good chance that it is one of 17 dialects listed in the phonological discriminator chart.

<u>Dialects with /g/</u> <u>[Words with /g/]</u>	<u>Dialects with /ž/</u> <u>[Words with /ž/]</u>	<u>Dialects with /č/</u> <u>[Words with /č/]</u>
Bahraini		Bahraini
Egyptian		
Emirati		Emirati
Hijazi	Hijazi	
Iraqi		Iraqi
Jordanian		
Kuwaiti		Kuwaiti
	Lebanese	
Libyan		
	Moroccan	
Omani		Omani
	Palestinian	
Qatari		Qatari
Sudanese		
	Syrian	
	Tunisian	
Yemeni		

Figure 8. AVIA Level 1 phonological discriminator chart.

4.5 Transcription system

Words are represented in three different transcription systems: native Arabic script, SATTs, and a modified version of the IPA. For a comparison of Arabic script to the modified IPA system, see the transcription guide in the AVIA. Native Arabic script and SATTs represent consonants only. The modified IPA system, on the other hand, gives both consonants and vowels, and thus offers a more detailed representation of how a word is actually pronounced. Note that the sound clip associated with the word in the AVIA is linked to the modified IPA transcription.

5. AVIA Level 1 Workflow

5.1 Suggested Workflow

When using the AVIA, we suggest that you follow the steps in Section 4.1 above. First, listen to the audio cut on which you are working, identify a word or sound, and then look for a match in the AVIA. If a match is found, this narrows down the possible dialects according to the word, phonological discriminator, or morphological discriminator lists in the AVIA. In some cases, you might find a unique match with the first dialect discriminator you find in an audio cut. Looking for the word /itjaabar/ “chat”, for instance, will reveal that this is only Yemeni Arabic. In most cases, however, the first match will only narrow down the possible dialects of your audio cut.

After narrowing down the dialect to a set of dialects, go back to your original audio cut and continue looking for matches. Continue this process until you are able to identify a unique dialect. In order to keep track of what dialects you have narrowed the sound cut down to, use the AVIA Worksheet, which is described in Appendix C.

The following flow chart summarizes the main steps you should use to identify a dialect with the AVIA:

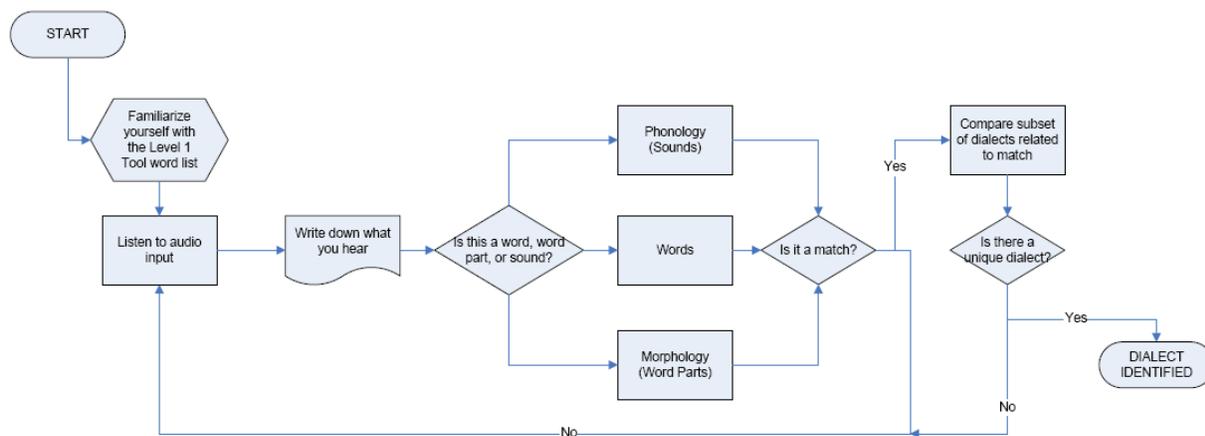


Figure 9. Workflow using the AVIA.

Note that there is no strict order that you should follow when using the AVIA. You will have to use the best strategy that works for you according to your skills in Arabic, the quality of the audio cuts, and so on.

5.2 Searching for a word (lexical discriminator)

If you have found a word from your sound clip that you want to search for in the AVIA, type the word using one of the three transcriptional systems (modified IPA, SATTs, or native Arabic script) in the “find” box in the “edit” menu of either Internet Explorer or

Firefox and then search for that word. You can also bring up the find box by using CTRL + F.

Note: The “find” feature functions differently in every browser. For example, in Internet Explorer, the “find” box appears in the middle of the screen, whereas in Firefox it appears at the bottom left. Both browsers will provide the same search results, just using a different interface. Screenshots from both have been provided in Figures 10 and 11, leading you to the word for “I” in the word list.

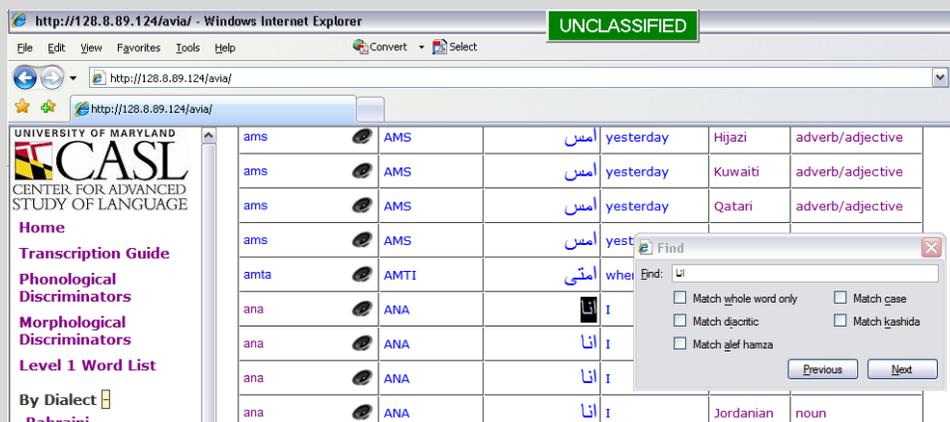


Figure 10. Finding the word انا using Internet Explorer.



Figure 11. Finding the word انا using Firefox.

5.3 Searching for a sound (phonological discriminator)

If you hear a sound that you think is distinctive (that is, a non-MSA sound), go to the phonological discriminator section. If the sound is there, click on the link at the top of the column. Only those dialects which typically have that sound will be displayed. This will help narrow down your search to those dialects.

5.4 Searching for a morphological discriminator

If you hear a particular morpheme (part of a word) that you think is distinctive, go to the morphological discriminator chart to help narrow down your search.

5.5 Searching for word classes, dialects

Using the filter function (see Section 4.2 above), you can also find all words from a particular dialect, or all words of particular word classes.

5.6 Example of a full AVIA session

Using the methodology that we laid out in Sections 4.1 and 5.1, we can begin showing you an example of how to use the AVIA.

While you can start with either a word, a sound, or a morpheme, in practice the identification of words and sounds often is easier than identifying an individual morpheme (since these are only parts of much longer words in Arabic). We will start this example with the identification of sounds that are not in MSA.

1. **Search for a specific sound:** Let us say you hear the sound /ʒ/ in your audio cut. Click on the phonological discriminator chart and click on the sound /ʒ/. This will immediately bring up all of the dialects which typically use this sound. Among the 17 dialects in the AVIA, the ones that contain /ʒ/ are Tunisian, Moroccan, Eastern Libyan, Palestinian, Syrian, and Hijazi. (Algerian uses /ʒ/ as well, though this dialect is not among the 17 included in the current version of the AVIA.)



UNIVERSITY OF MARYLAND
CASL
CENTER FOR ADVANCED
STUDY OF LANGUAGE

Home
Transcription Guide
Phonological Discriminators
Morphological Discriminators
Level 1 Word List
By Dialect
Bahraini
East Libyan
Egyptian

view sortable version [beta]

TRANSCRIPTION	SATTS	ARABIC	ENGLISH	DIALECT	WORD CLASS
farʒa yfarʒi	FRJI IFRJI	فرجى يفرجى	to show	Lebanese	verb
riʒaala	RJAL?	رجالة	men	East Libyan	noun
Ruʒaala	RJAL?	رجالة	men	Moroccan	noun
rʒaal	RJAL	رجال	men	Lebanese	noun
rʒaal	RJAL	رجال	men	Palestinian	noun

Figure 12. AVIA Level 1 list of words that contain /ʒ/.

While this has narrowed down the search to eight possible dialects, you need more information to further narrow the dialect of your audio cut. Go back to the audio cut to look for more clues.

2. **Identify unique terms:** During a second listen through the audio cut, you identify the word /rʒaal/. Instead of looking for sounds, look at the word list and find the entries for “men.” It turns out that a number of dialects have this word using this pronunciation: Tunisian, Palestinian, Syrian and Lebanese. Moroccan and Eastern Libyan have a very similar word, /rʒaala/, so you have to be very sure of what they have heard before deciding further about the dialect on the basis of this word. In order to determine the difference, listen to the sound clips in the AVIA to determine whether you heard /rʒaal/ or /rʒaala/.

In order to verify that you are narrowing down the dialect of your audio cut correctly, find another word that is an additional discriminator in your audio cut. This time you hear the word /kiif/ “how.” Among the eight dialects with /ž/, only Syrian, Lebanese and Palestinian have this pronunciation of “how.” You can confirm this by clicking on the AVIA sound clips for all the words that are similar to /kiif/ to listen to how it is pronounced in these dialects.

Transcription Detail - kiif

[view sortable version](#) [beta]

TRANSCRIPTION	SATTS	ARABIC	ENGLISH	DIALECT	WORD CLASS
kiif	KIF	كيف	how	Syrian	question word
kiif	KIF	كيف	how	Lebanese	question word
kiif	KIF	كيف	how	Palestinian	question word

Figure 13. AVIA Level 1 dialects that contain /kiif/ *how*?

At this point, note that Tunisian also has the pronunciation of this word with a high vowel /ii/ (as opposed to the MSA pronunciation /kayf/). However, this dialect adds an /-aš/ at the end of the word, /kiifaš/. Although the vowel is like that of Syrian, Lebanese or Palestinian, the final /-aš/ is not, and you have further narrowed down the dialect to these three dialects that contain a /ž/, the word /rʒaal/, and the word /kiif/.

TRANSCRIPTION	SATTS	ARABIC	ENGLISH	DIALECT	WORD CLASS
kiifaš	KIF:	كيفش	how	Moroccan	question word

Figure 14. AVIA Level 1 dialects that contain /kiifaš/.

To confirm what you are hearing, click on variants of “how” in the AVIA that are transcribed with the long vowel /ee/, such as Jordanian, in order to determine that you heard /kiif/ and not /keef/. After confirming that what you are listening to is /kiif/ and not /keef/ by listening to the AVIA sound clips, you know that your audio cut is likely Syrian, Lebanese, or Palestinian.

Now that you have now narrowed down the choice to these three dialects, go back to you original audio cut.

3. **Consider morphological discriminators:** This time you hear the phrase /ma šuftuuš/. You recognize this to be a negative verb, and hear the final /-š/ at the end of it. This time, look at the morphological discriminator chart, since you know

that you are searching for a part of a word. It turns out that among the three dialects still in the running, only Palestinian has a negative with the morpheme /-š/. At this point you have, according to the parameters of the AVIA, identified the dialect of your audio cut as Palestinian with a high degree of confidence.

5.7 Potential Traps

Arabic dialects are vast and diverse and vary significantly both from MSA and each other. An Arabic dialect is the native language of all Arabic speakers, and there are between 200 – 250 million speakers worldwide. There are more than just the 17 dialects of Arabic described in the AVIA; however, by focusing on these 17 dialects, the AVIA provides maximum coverage of these 200 million speakers.

Some of the dialects are very similar to each other while others, such as North Iraqi and Tunisian, are very different. Also, some speakers may be familiar with multiple dialects; for example, a Palestinian may know Egyptian dialect very well.

So, watch out for the following pitfalls when using the AVIA:

- **False friends.** The same “word” might have different meanings in different dialects. In Bahraini, for instance, /dazz/ means “push,” but in Iraqi this *same* word means “send.” To avoid confusion, listen carefully to the context to know which alternative meaning is the correct one.
- **Dialect does not always equal speaker origin.** There is a strong possibility that if, for instance, you have identified an audio cut as being Egyptian, that the speaker does indeed come from Egypt. However, speakers can be bi-dialectal (know more than one dialect) and they sometimes adapt to other dialects. For instance, it is not infrequent that North Africans will adopt Egyptian or Syrian pronunciations when speaking with non-North Africans in order to be understood better because North African Arabic can be difficult for outsiders to understand.

Arabic is a large language, with many variants. The AVIA is not comprehensive for all Arabic dialects, so it is possible the dialect you are listening to is not represented in this aid. A match with a form in the AVIA is a step forward, but what you are listening to may still be compatible with another dialect which has not yet been incorporated into the AVIA

Appendix A. Troubleshooting Guide

Problem	Explanation	Solution
The word you identified is not in AVIA Level 1.		There is only one solution to this: go back and find a word that is.
The word is hard to hear in the original audio cut.	You may have found a word which you think will help you identify the dialect but for some reason it is not easy to understand.	Listen to the word carefully and compare it to possible matches in AVIA Level 1. Try several matches and fit the different ones into your original audio cut. If one of them makes sense, it might be the correct one and you have taken a step towards identifying the dialect.
The word is hard to hear in AVIA Level 1.	Some of the audio cuts in the AVIA were taken directly from authentic speech. Sometimes these are not as clear as words recorded specifically for AVIA Level 1.	In the future, the AVIA may re-record some of the harder-to-hear words.
The word you hear is in AVIA Level 1, but it does not have the same form as what you heard.	Arabic is complicated in a number of ways. One of them is the large inflectional system. “He saw” in Syrian is /warža/, but “I saw” is /waržeet/. The stress is different between these two words. Also, the final syllable is different. In the AVIA, only a limited number of forms could be recorded.	For the verbs, the third person masculine singular forms generally are used as they are the customary citation forms used in Arabic dictionaries. Thus, if you heard /waržeet/ in the audio cut, you will have to find /warža/ in the AVIA. Note that to identify Syrian or Palestinian, either form, /waržeet/ or /warža/ will work. There is no difference in the morphology (the suffix /-eet/) between these two dialects and many other dialects (e.g. Jordanian has /farž-eet/, with a different word for “show”, but the same suffix /-eet/ for “I”). At this point, there is no way around the problem posed by the added-on Arabic morphology. You yourself will simply have to add or change the “extra” morphology to find the word you are looking for.
The word is hard to hear in the audio cut (as in second point above), and what you think <i>might</i> be the same word on the AVIA does not appear to match what you hear well.	This is a problem you will have in dealing with real, live speech. Most of the audio prompts in the AVIA are carefully recorded segments, to enable you to hear them clearly. The AVIA recordings will thus usually be clearer than what you hear in your original audio cut.	Compare the original audio cut carefully to what you hear and read in the AVIA. Pronounce the words to yourself to get a better impression whether or not they match. Try to imagine how the original would have sounded if it had been said more slowly and carefully.

Appendix B. Glossary of Terms

Dialect. A language variant. Dialect is often a variant opposed to a standard language e.g. Modern Standard Arabic (MSA). Remember that an Arabic dialect is also the native language of Arabs while MSA is learned later while in school.

Dialect discriminator. A linguistic feature (sound, word, or morpheme) that serves to mark the identity of a dialect.

Habitual action. An action that occurs often, such as a daily action.

International Phonetic Alphabet (IPA). The internationally standardized phonetic representation of a language in which each distinctive sound—whether a consonant or vowel—has a distinctive letter representing it.

Lexical. Relating to words. A “lexical discriminator” is a word that serves to distinguish a dialect.

Lexical discriminator. A word that serves to distinguish one dialect from another.

Morphology. The internal structure of words in terms of their meaningful parts e.g. “bushes” = “bush” + “-es” = plural.

Ongoing action. An action that is occurring now, also called “progressive.”

Phonology. The sound system of a language.

Phonetics. The way sounds are pronounced.

SATTS (Standard Arabic Technical Transliteration System). System for transliterating Arabic script on a one-to-one basis.

Appendix C. Dialect ID Worksheet

C.1 Worksheet

You may find the following worksheet helpful for recording the steps you took to identify a dialect. A printable PDF of this worksheet can be found on the AVIA CD as AVIA Worksheet.pdf

Arabic Variant Dialect Identification Aid – Dialect ID Worksheet																			
What I heard in the sample	Where I looked in the AVIA	Dialects which match input										Notes							
		Bahraini	East Libyan	Egyptian	Emirati	Hijazi	Iraqi	Jordanian	Kuwaiti	Lebanese	Moroccan		North Iraqi	Palestinian	Qatari	Sudanese	Syrian	Tunisian	Yemeni

Figure 15. AVIA Level 1 worksheet.

C.2 Worksheet Example

If we go back to the example in Section 5.6 picking up where the word /kiif/ has been identified, you would mark the boxes in the first row where /kiif/ is used. On line two you would note the form /ma-šuftuuš/. (We have written these words in IPA, though you could equally write in SATTs or Arabic script). Look at the morphological discriminators list. As you have narrowed down the dialects to only three in the previous step, you need only note the dialect from those three which has the negative /ma-šuftuuš/. The example in Section 5.6 using the AVIA worksheet is shown below:

Arabic Variant Dialect Identification Aid – Dialect ID Worksheet																				
What I heard in the sample	Where I looked in the AVIA	Dialects which match input										Notes								
		Bahraini	East Libyan	Egyptian	Emirati	Hijazi	Iraqi	Jordanian	Kuwaiti	Lebanese	Moroccan		North Iraqi	Palestinian	Qatari	Sudanese	Syrian	Tunisian	Yemeni	
1. /kiif/	Word list, using SATTs								X		X			X						<ul style="list-style-type: none"> • Could be one of these three dialects • No /š/ at end of word, so not Tunisian/Moroccan
2. ma-šuftuuš	Morphological discriminators, using Arabic script, ما شفتوش										X									<ul style="list-style-type: none"> • Narrows down to Palestinian • ما شفتوش occurs in other dialects, but not dialects with /kiif/

Figure 16. Example of using the AVIA Level 1 worksheet.

