UTEP Technical Report UTEP-CS-06-02

The UTEP Corpus of Iraqi Arabic

Nigel G. Ward, David G. Novick, Salamah I. Salamah

Department of Computer Science
University of Texas at El Paso
500 West University Avenue
El Paso, TX 79968-0518

email: nigel@cs.utep.edu
voice: 915 747-6827
fax: 915 747-5030
URL: http://www.cs.utep.edu/nigel/

January 23, 2006

Abstract

The rules governing turn-taking phenomena are not well understood in general and almost completely undocumented for Arabic. As the first step to modeling these phenomena, we have collected a small corpus of Iraqi Arabic spoken dialogs. The corpus is in three parts. Part A is 110 minutes of unstructured conversations. Parts B1 and B2 are 176 minutes of direction-giving dialogs, most including a greeting phase, a smalltalk phase, a request phase, and a direction-giving phase. Parts A and B1 were recorded with 13 native speakers of Iraqi Arabic, interacting in pairs. In Part B2 the direction-getter is an American with only very basic Arabic knowledge. This document describes the intended uses of the corpus, the dialog types, the speakers, and the details of the data collection.

1 Aims

This corpus was recorded as part of the project entitled "Beyond Words: Identification of Back-Channel Communication Rules in Arabic and Development of Training Methods", funded by DARPA's Defense Sciences Office, via the Department of the Interior and the Information Sciences Institute of the University of Southern California. The corpus was designed to have general value, but also to meet the project aims. To quote from the project charter:

> The rules governing real-time interpersonal interaction are today not well understood. With only a few exceptions, there are no quantitative, predictive rules explaining how to respond in real-time, in the sub-second range, in order to be an effective communicator in a given culture. This can be a problem in intercultural interactions;

> if an American knows only the words of a foreign language, not the rules of interaction, he can easily appear uninterested, ill-informed, thoughtless, discourteous, passive, indecisive, untrusting, dull, pushy, or worse. Short of long-term cultural exposure, there are today no reliable ways to train speakers to understand and follow such rules and attain mastery of interaction at the sub-second level. The purpose of this research is to increase our knowledge and know-how in this area.

In this project the specific focus is on these phenomena in Iraqi Arabic, in the context of intercultural communication problems faced by Americans in Iraq. Immediate application to the ISI's Tactical Language Trainer[1] (Johnson et al., 2005) is planned. The underlying scenarios of interest involve obtaining information from an Iraqi civilian. The goal is to enable users to learn how to accomplish their tasks by effectively communicating, and in particular using the behavior patterns involved in being a good listener in Arab dialog.

In the first year the main tasks (again from the project charter) are to:

C1 Discover the basic rules governing back-channel behavior (production of *uh-huh*, etc. as a display of serious listening) in Eastern Arabic

C2 Compare these with the rules governing such behavior in American English and quantify the importance of following the right rules

C3 Develop methods for training American speakers to understand and emulate the rules in Arabic and measure learnability

C4 Produce a toolset for the automatic discovery of new rules in new languages and cultures, perhaps based on didi [2] (Ward, 2006)

C5 Apply these training methods in ISI's Tactical Language Trainer

C6 In addition, help add back-channel behaviors to the animated characters of the Trainer to improve verisimilitude

The corpus was also designed to support some more technical enhancements to the trainer:

A1 Dialog Pacing in Direction-Giving In the Trainer's cafe scene, the native gives the user directions to the leader's home. This is currently a fixed chunk of speech, about 8 seconds of fluent Arabic, which no learner can understand.

Currently this chunk functions as the reward: if the user does everything else right, he gets to this stage. It also helps reinforce the user's understanding of directions ("left", "right" ...) in Arabic.

There is an opportunity here to teach one more skillset: pace control. That is, this scene can be extended to teach the user how to indicate that he needs a rate speed up or rate slow down, and also whether he wants the direction-giver to go on, repeat, or pause to allow time

[1] http://www.tacticallanguage.com, http://www.isi.edu/isd/carte/proj_tactlang/

[2] http:/www.cs.utep.edu/nigel/didi/

to process the information received. At another level, this involves learning how to signal understanding, temporary confusion, or complete misunderstanding. The user can also be taught how to detect whether the speaker intends to go on or is done.

This requires corpus data, both to discover the rules to teach to users, and to discover how to script the agent to behave properly.

A2: Endpointing The Trainer currently uses fairly crude endpointing, which means it can misjudge whether the user has ended his turn. This implies two complementary problems: First, if the user momentarily pauses to decide what to say next, the system is liable to cut him off. Second, even if the user is really done, the system waits for a half-second of silence to be sure, meaning that there is dead time and that the interactions are less tight than they should be. If the Trainer were to just use a more sophisticated English endpointer, these problems would be alleviated to some degree. Even better would be to create an Iraqi endpointer, that is, figure out how Iraqis indicate that they are holding the turn vs. ending the turn, train users to do both, and make the system follow the Iraqi rules. Specifically the skills involved here are: indicating that your turn is over, indicating that you want to take a turn, and indicating that you want to keep the floor.

The need for this was seen in a video of soldiers, who had learned the basics of Arabic with Trainer, attempting to talk to an Iraqi. Some tended to produce long jumbled conglomerations of the phrases they had learned, hoping one would be the right one. Among other things, this seems to indicate the need for explicit training in how to yield turns, hold turns, and hold the floor politely.

A3: Back-Channel Detection An issue arises of detecting when the user is producing a back-channel vs. a full utterance. The technical issue is that invoking HTK, the speech recognizer, involves significant delay, so that a swift preliminary detection of back-channels would support faster responsiveness, economize on CPU load, and reduce speech recognition errors.

A4: Gesture Choreography The Trainer team plans to add new gestures and finer temporal control of gestures to a future version.

Finally we wanted to allow for the possibility that some corpus utterances could be directly used:

A5: Voice Data Gathering The corpus could allow use of the audio collected for the Trainer or ancillary systems, and may allow identification voice talent for future activities.

The primary goal of the corpus is to support the above tasks, especially discovery of the rules for C1/C6, A1, A2, and A3. It should also help explore ideas on how to do the evaluation in C2, and help A4, and A5.

2 Corpus Parts

The corpus has three parts: unstructured dialogs, task-oriented dialogs, and dialogs with a learner of Iraqi Arabic.

2.1 Part A: Unstructured Conversations

This part of the corpus was collected to support tasks C1, A2, A3, and C6.

The size of this part, 110 minutes, was based on past experience doing similar work with Japanese and English (Ward and Tsukahara, 2000). We wanted about 360 back-channel tokens to allow rule discovery and evaluation, and expected to see about 4.2 back-channels per minute, based on the frequency in Egyptian Arabic Callhome(Ward and Bayyari, 2006). We also wanted at least a dozen speakers: although this may not be enough to support evaluation in the sense of demonstrating that the regularities found are truly general and statistically solid, this is acceptable given that we plan to evaluate the results mostly experimentally (C2).

Each dialog was between two Iraqi speakers.

Participants were recorded sitting facing each other across a table.

Parenthetically, it was difficult to decide whether the corpora should be gathered in face-to-face conditions or without visual contact. The Trainer has no vision system, so a corpus collected without visual contact would better support development of rules that the Trainer could work with. On the other hand, in the Trainer scenarios the user is face-to-face with the on-screen agents, so this suggested that the corpus dialogs also be face-to-face. The deciding factor was the desire to collect gesture data (A4) at the same time as the audio: this required that the dialogs be face-to-face. This may complicate analysis, since the availability of the visual channel affects the dynamics of the audio interaction. However, to the extent that visual cues appear redundantly to auditory ones, rather than substituting for them, the results of analyzing this corpus may be of general value.

Part B1: Direction-Giving Dialogs

These dialogs were semi-structured dialogs, typically including a greeting phase, a smalltalk phase, a request phase, a direction-giving phase, and a thank-you/goodbye phase. The direction-giving phase directly supports task A1. The other phases make the dialogs feel complete and are potentially useful for helping plan the evaluation in C2, for making the dialogs better match the scenario in the cafe scene, and for gathering video data on greetings (A4).

Based on past experience (Iwase and Ward, 1998), 10 dialogs totaling 30 minutes would probably have been adequate, but we collected more.

Before entering the recording studio, each participant was asked to think of some place he could give directions to confidently; for example from the metro station to his workplace, or from the metro station to the recording studio. We then told the recipient this place, telling them to, e.g., ask the other person the way to get to his office. However it was apparently not always clear to the participants what we wanted, and so it was sometimes necessary to re-direct participant in the recording studio or even after a dialog had started.

Each pair of participants did the direction-giving exercise twice, with the direction-giver and the direction-getter swapping roles for the second run.

In normal direction-giving dialogs, participants often refer to a map or take notes. However

in the current Trainer scenarios the person getting the directions does not have a map nor a way to take notes, so we similarly had the direction-getter just listen. To some extent this may have made the dialogs unrealistic, especially when the directions given were too long for the receiver to plausibly remember.

We initially considered having the direction-requester initially be standing, and approach the direction-giver who was initially sitting down, to match the staging in the Trainer's cafe scene, but we recorded them all sitting down; so the conversation openings and closings were perhaps abbreviated. This was done to simplify logistics (since we used head-mounted microphones, we were afraid participants might trip if they were moving around), and to allow static camera placement.

2.2 Part B2: Dialogs with an American

These dialogs were the same as those in Part B1 except that one speaker, the direction-receiver, was an American male learner knowing only the Arabic acquired in about 20 hours of working through Mission Skill Builder exercises in the Trainer, especially those on greetings, simple questions, and directions. The learner also received about one hour of tutoring from a native speaker of Arabic

These dialogs were collected because we were concerned that the Iraqi-Iraqi dialogs (B2) were not entirely appropriate to use as models for the dialog patterns that an American speaking with Iraqis could be expected to produce nor those he could expect to hear. When talking to a learner, the native is likely to change his behavior, and to have different expectations for his interlocutor's behavior. For example, an Iraqi talking to an American may alter his behavior by (a) talking slower and (b) giving exaggerated turn-taking cues, and (c) allowing more time for the American to reply. Comparison of the dialogs in Parts B1 and B2 may make it possible to quantify parameters like a, b, and c.

3 Participants

Participants were recruited for us by a contact in the Iraqi-American community.

The participants were all Iraqis. All had been living in the United States for at least two years at the time of the recording. Many spoke a number of languages. Five had lived longer outside Iraq than within it, but all had lived in Iraq at least through their mid-teens.

We had intended that all would have Arabic as their first language, but in fact two of the five above had Kurdish as their first language, although it appears that from school age they also spoke Arabic.

The participants were all male, as the project focus was on male speech patterns. This was for two reasons. First, we thought that American personnel in Iraq would be mostly interacting with male Iraqis. Second, we thought that American personnel in Iraq, being mostly male, would generally benefit most from being trained in male Arabic speech patterns.

The participants ranged in age from 25 to 70+, with most middle-aged. There was substantial

diversity in dialect background and in occupation, on which we collected information. We did not ask for information on ethnic background, education, or social status, but there appeared to be substantial variation here also. Our recruiter indicated that he had arranged for the participants to represent diverse social and ethnic groups within the Iraqi population.

4 Recording Conditions

Recordings were done in stereo using head-mounted microphones in CD quality (44100 samples per second, 16 bits per sample). Because there was no sound isolation, each track faintly contains the voice of the other speaker.

Dialogs were also videotaped using two cameras. We were concerned that the presence of camera operators would make the participants uncomfortable, self-conscious or cause them to behave like actors. To avoid this, the videocameras were statically placed and adjusted only at the beginning of recording each new pair of speakers. There was no tracking. The cameras were positioned to frame the upper body and thus catch facial and hand gestures. The audio and video recordings were not synchronized to each other.

5 Participant Handling

The dialogs were recorded on a Saturday in November 2005, in a hotel in Washington D.C.

Participants were compensated with $100 each.

We attempted to make the participants feel at ease, by supplying glasses of water, by walking them patiently through the paperwork, by generally treating them with respect, and by giving them adequate guidance and support. There was a six-person team involved in this. The roles were roughly as follows: one usher to help welcome participants and guide them to and from the recording studio, one paperwork handler, one recording engineer, one director in the recording studio to give instructions and occasionally call for retakes, one floater, who doubled as one of the participants, and one project director, who doubled as the learner. The director, usher, and floater were native Arabic speakers.

Pairings were staggered: for example, participant 1 and participant 2 were recorded as a pair, then participant 2 and participant 3, and so on. The last participant closed the loop by being recorded with the first participant.

To match the Trainer scenarios, we wanted the dialogs to be representative of dialogs with strangers. This was not entirely possible due to scheduling constraints, so about a third of the dialogs were between pairs of Iraqis who knew each other well. Moreover, since participants unavoidably had some amount of waiting, most participants had some opportunity to see each other before their recording session. The American learner, however, did not interact with the Iraqis before the recording sessions, so the B2 dialogs were all indeed between strangers.

The order of dialogs for each pair was determined by the fact that it was more important for the participants to be relatively mutually unfamiliar in the B1 dialogs, so these were generally

recorded before the A dialogs.

We did not want participants to be overly self-conscious about their speaking style, for fear that it would affect how they spoke. Thus we downplayed our interest in "speech properties and dynamics of interaction", although this was of course mentioned on the Consent Form, by focusing on other goals in the Instructions to Participants.

The data collection was in accordance with the research protocol titled "A Study in Iraqi Arabic Conversation Dynamics", approved by the UTEP Institutional Review Board and assigned number 2141.

6 Appendices

Agreement and Consent form

Instructions to Participants

Participant Datasheet

Dialog Log

References

Iwase, T. and Ward, N. (1998). Pacing spoken directions to suit the listener. In *International Conference on Spoken Language Processing*, pages 1203–1206.

Johnson, W. L., Beal, C., Fowles-Winler, A., Lauper, U., Marsella, S., Narayanan, S., Papachristou, D., Valente, A., and Vilhjalmsson, H. (2005). Tactical language training system: An interim report. USC ISI, adapted from a conference paper presented at the Intelligent Tutoring Systems Conference, September 2004.

Ward, N. and Bayyari, Y. A. (2006). A prosodic feature that invites back-channels in Egyptian Arabic. Paper to be presented at the *20th Arabic Linguistics Symposium*.

Ward, N. and Tsukahara, W. (2000). Prosodic features which cue back-channel feedback in English and Japanese. *Journal of Pragmatics*, 32:1177–1207.

Ward, N. G. (2006). Methods for discovering prosodic cues to turn-taking. In *Speech Prosody*. submitted.

A Study in Iraqi Arabic Conversation Dynamics
--- Agreement and Consent ---

Description of Study

The purpose of this study is to understand the speech properties and dynamics of interaction in Iraqi Arabic. This study is under the direction of Dr. Nigel Ward. You will be asked to hold a few short conversations. This will take less than 30 minutes. Participation is voluntary. If at any point during the study you feel uncomfortable, they may contact Karen Hoover, Institutional Coordinator for Research Review of the University of Texas at El Paso, at 915-747-5680. You may also contact Dr. Nigel Ward at 915-747-6827.

Participant Statement and Signature

I understand and agree that:

1. There are no known risks or benefits involved in participating in this study, other than the monetary remuneration.
2. Participation is voluntary. I may end my participation at any time and for any reason.
3. Recordings will be made during the session.
4. The recordings may be used by members of the UT El Paso Interactive Systems Group and other persons for reasonable education, scientific, and technical purposes.
5. The recordings will be kept confidential. These recordings will be stored on password-protected computers, and after the research is complete will be archived. In particular, recordings in which I can be identified will not be publicly released, will not be used in any public presentation, and will not appear in any publicly available media, unless I approve.*
6. The fact of my participation in these sessions will also be kept confidential, and my name will also be kept confidential, unless I approve.+

Signature: _____ Date: _____

Name: _____

Address: _____

_____

Telephone: _____

Researcher's Signature: _____ Date: _____

*If you would like to allow other uses of these recordings, please initial here

___ Samples of my voice may be used in CDs, software, or other materials for educational purposes.

___ Video clips including me may be used in videotapes, DVDs, software, or other materials for educational purposes.

+If you would like your name to be made available in a follow-on project, please initial here:

___ I am willing to be contacted as a possible actor, voice talent, or consultant for the development of educational materials.

rev Sep 29, 2005
IRB Protocol 2141

A Study in Iraqi Arabic Conversation Dynamics
--- Instructions to Participants ---

Thank you again for participating. One aim of our study is to improve teaching and learning of Arabic.

Please read these instructions. You can refer to them in the recording studio.

Phase 1

Our scenario involves two roles: the Giver and the Receiver. You will take each role in turn.

In the Giver role, you will give the other person directions to some location. Right now, please think of some location you know well (perhaps your place of work), and can easily give directions to. The place is _____

Once in the recording studio, you will sit down at the table. Pretend you are outside at a café. When the other person approaches you, pretend he or she is a stranger.. They will not have a pencil to write things down, so don't make the directions too long at first … just give the first 3 or 4 things they have to do, then say something like ``at that point you'll be close, so just ask someone".

In the Receiver role, we will give you the name of a place, and your job is to obtain directions on how to get there. You will enter the recording studio and see someone sitting down. Pretend they are a stranger who you see sitting outside at a cafe. Greet the person, politely engage them, and ask them directions to the place. Afterwards we may ask you to write down the directions as best you can.

This phase is necessary to ``warm you up" in speaking Arabic, to help us calibrate our recording equipment, and to prepare you for the next phase.

Phase 2

In the second phase, you will interact with an learner of Arabic. We are interested in developing new methods for evaluating the social/cultural competence of beginning foreign language learners.

This dialog will proceed exactly as above. We want this to be an ``authentic" experience for the learner, so please don't use English.

Phase 3

In this stage we will record the two of you discussing the performance of the learner. It will probably not take long to reach a rough consensus on his abilities, however to pay you we need a full 8 minutes of conversation. To fill up the time you can talk about anything you like (weather, pets, food … absolutely anything).

After this, you will each rate the learner's language and culture skills.

A Study in Iraqi Arabic Conversation Dynamics
--- Participant Datasheet ---

number ____

Your age: 18-19, 20-24, 25-29, 30-34, 35-39, 40-49, 50-59, 60+

Your sex: M F

Your occupation:

Your linguistic background:

from	to	location	main languages/dialects used
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____
_____	_____	_____	_____

…….. *fill in below after the recording session* ……..

Did you know your first dialog partner before today? If so how well?

Did you know your second dialog partner before today? If so how well?

Did the learner have the social/cultural/language skills needed to be effective in simple dialogs?

1 2 3 **4** 5 6 7

Was the learner polite by the standards of Arabic interaction?

1 2 3 **4** 5 6 7

Comments, if any:

A Study in Iraqi Arabic Conversation Dynamics
--- Dialog Log ---

Dialog Type: 1-directions 2-directions-with-learner 3-free-conversation

Participants ID#s: ____ ____

Start Time:

Observations (recording conditions, unusual aspects, retakes …)

Dialog Type: 1-directions 2-directions-with-learner 3-free-conversation

Participants ID#s: ____ ____

Start Time:

Observations (recording conditions, unusual aspects, retakes …)