

NISTIR 7712

**APPROACHES AND BEST PRACTICES:
Data Collection of Audio Dialogues to
Support the Training of Speech-to-Speech
Translation Systems**

Brian A. Weiss

Craig Schlenoff

Ann Virts

U.S. DEPARTMENT OF COMMERCE

National Institute of Standards and Technology

Intelligent Systems Division

Gaithersburg, MD 20899-8230



National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

NISTIR 7712

**APPROACHES AND BEST PRACTICES:
Data Collection of Audio Dialogues to
Support the Training of Speech-to-Speech
Translation Systems**

Brian A. Weiss

Craig Schlenoff

Ann Virts

U.S. DEPARTMENT OF COMMERCE

National Institute of Standards and Technology

Intelligent Systems Division

Gaithersburg, MD 20899-8230

June 2010



U.S. DEPARTMENT OF COMMERCE

Gary Locke, Secretary

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

Patrick Gallaher, Director

APPROACHES AND BEST PRACTICES:

Data Collection of Audio Dialogues to Support the Training of Speech-to-Speech Translation Systems

Last Updated: July 19, 2010

National Institute of Standards and Technology¹

Brian A. Weiss

Craig Schlenoff

Ann Virts

¹ Certain commercial companies, products and software are identified in this paper in order to explain our research. Such identification does not imply recommendation or endorsement by NIST, nor does it imply that the companies, products and software identified are necessarily the best available for the purpose.

TABLE of CONTENTS

1.	Introduction	3
2.	Scenario Generation	3
2.1.	Scenario Creation/Identification	4
2.2.	Inventory of Existing Scenarios	4
2.3.	Topic Identification	5
2.4.	Scenario Writing and Review	5
3.	Data Collection.....	6
3.1.	Data Collection Site Selection.....	6
3.2.	Noise Masking and Wizard of Oz Setup	7
3.3.	Participants	9
3.4.	1 st Morning Kickoff.....	11
3.5.	Workflow	11
3.6.	Subsequent Morning Kickoffs.....	13
3.7.	File Naming Convention and Post-Collection Delivery.....	13
3.8.	Transcription/Translation	14
3.9.	Data Usage.....	15
4.	Cost.....	15
5.	Conclusion.....	16
	APPENDIX A: Example Iraqi Data Collection Scenarios.....	17
	APPENDIX B: Example Afghan Data Collection Scenarios	20
	APPENDIX C: Previous NIST Collection Output.....	24
	APPENDIX D: Wizard of Oz – Speech Flow and Implementation	25
	APPENDIX E: 1 st Morning Briefing Points	27
	APPENDIX F: Example Transcription/Translation Guidelines	31

1. Introduction

The purpose of this document is to describe the best practices that personnel from the National Institute of Standards and Technology (NIST) have developed and implemented to efficiently and effectively capture two-way, free-form speech-to-speech audio dialogues within recording studios. These dialogues, produced to support the development and evaluation of machine translation technologies, are conducted by English and foreign language speakers conversing with one another in their native languages through the mediation of an interpreter. NIST personnel have collected over 500 h of bilingual audio data sets encompassing more than 1100 dialogues across three unique language pairs (English/Iraqi-Arabic, English/Dari, and English/Pashto) since it became involved in this work in 2007.

This document will present the methods the NIST team has designed and employed allowing the successful capture of audio data. In addition to the data collection protocols including personnel training and workflow, data collection scenario generation and speaker recruitment protocols will be discussed. To further understand this work, the following acronyms (presented in Table 1) will be used throughout the document for brevity and clarity.

Table 1: Document-Specific Acronyms and their Meanings

ACRONYM	MEANING/DESCRIPTION
FLE	Foreign Language Expert - a bilingual speaker fluent in English and the target collection language
IET	Independent Evaluation Team - the NIST-led team in charge of the data collection process
MOS	Military Occupational Specialty - used to identify specific job function(s) of a US military personnel
NIST	National Institute of Standards and Technology
OC	Overall Coordinator - the individual in charge of the weekend data collections
OEF	Operation Enduring Freedom - refers to the conflict in Afghanistan
OIF	Operation Iraqi Freedom - refers to the conflict in Iraq
SME	Subject Matter Expert - US military personnel playing the role of the English speaker

2. Scenario Generation

Before data collections can begin, it is crucial to generate relevant and realistic scenarios to govern the individual dialogues. Developing scenarios for the data collections included a series of steps listed below. Note that these steps are presented in no particular order.

- **Scenario Format Creation/Identification**– The format in which the scenario is presented is critical to the success of the dialogues. This directly impacts the amount of information that the speakers will blend in with the written content along with influencing their engagement in the material.
- **Review Existing Scenarios** - If available and the current work is a continuation of previous efforts, review any existing scenarios and determine what (if any) are relevant to current project's collection goals. This review can include both data collection scenarios and any scenarios used to support previous test events.
- **Rework Existing Scenarios** - Depending upon the age and nature of the previous scenarios along with their general domain, previous scenarios could require changes anywhere from minor adjustments to significant rework. It's also quite possible that some scenarios may no longer be applicable in which case they should no longer be considered.

- Create New Scenarios – Each new scenario is based upon information gathered from focus groups, individual interviews, media articles, and/or previous scenarios (if available). The focus groups and individual interviews should be conducted with subject matter experts who belong to the targeted group of technology end-users.
- Scenario Review - Have a diverse scenario review committee examine each scenario for accuracy and relevance. This should include subject matter experts (as described above), foreign language cultural experts, data collection administrators, and evaluation designers (if identified and available).

Each of these steps will be elaborated upon in the following subsections.

Note – some of the scenarios used to support the English/Iraqi-Arabic bilingual data collections can be found in APPENDIX A: Example Iraqi Data Collection Scenarios. Example scenarios used to support the English/Dari and English/Pashto bilingual data collections can be found in APPENDIX B: Example Afghan Data Collection Scenarios.

2.1. Scenario Creation/Identification

The *Spontaneous* scenario format was developed to allow the speakers (both English and foreign language) to have reasonable freedom and latitude in their dialogues. The overall format begins with the specification of the overall domains. For each domain, multiple English-speaker motivations were specified that provided some background and situational information. Additionally, each English speaker's motivation was paired with numerous talking points with the intent of giving the English speaker specific topics that they can include in their dialogues, but not limit them to specific questions.

In turn, each foreign language speaker was provided with their own specific background (its inclusion is dependent upon the domain) and motivations. For some of the domains, “hides” are included with each foreign motivation. The purpose of the “hides” was to provide the FLE with a realistic object, piece of equipment or material that the SME would discover throughout the course of the scenario. Specifically, these “hides” take the form of the highlighted words written in large print on a single sheet of paper, such as **SHOVELS AND PICKS**. Prior to the start of a scenario, the FLE was given the paper with the hide and has a chance to practice their dialogue away from the SME. The rehearsal coach decides upon a reasonable place that the “hide” should be found. For example, if the “hide” was a bag of money, then it could be decided that the FLE would reveal this when the SME specifically states they are going to search the glove box (because a bag of money could reasonably fit in there). Likewise, the SME would be told by the dialogue coach that they are specifically to state (in their dialogue) that they are searching the glove box (knowing that something will be revealed then). However, the SME does not know the nature of the “hide” until it's revealed in the scenario. This enables the SME to be more spontaneous in their response to the “hide” and realistically alter their line of questioning according to what was just found. This format requires the military speakers to include their own experiences, but still enables the evaluation team to get data within specific domains.

2.2. Inventory of Existing Scenarios

It was discovered that there were both similarities and differences between US military interactions with Iraqi-Arabic (IA) speakers in Iraq as compared to interactions with Dari and Pashto speakers in

Afghanistan. This makes it important to dissect any previously-generated scenarios to see what content (if any) is still relevant. Based upon the *Spontaneous* scenario format and the nature of a previous effort's scenarios, it's possible that the same domains could be reused where only the SME and FLE motivations would need to be revised and/or updated. It's also possible that the previous effort's domains no longer apply where new scenarios need to be generated.

2.3. Topic Identification

The primary source of information for identifying Afghan (or Iraqi) scenario ideas stemmed from discussions/informal interviews conducted with US service personnel at numerous military bases throughout the country. Evaluation team members visited multiple military installations where they conversed with US servicemen and women about their experiences overseas in Afghanistan (or Iraq). Focus groups consisted of enlisted personnel and officers who have had interactions with the local population. It was determined that the most successful focus groups were those that were split between officers/enlisted and enlisted in various Military Occupation Specialties (MOSs). For example, at one military base, three separate focus groups were conducted. The first group contained Non-commissioned Officers (NCOs) who performed infantry tactical functions, the second group contained Hospital Corpsmen Third Class (HM3s) who discussed their medical engagements, and the third group contained other NCOs whose primary functions included leader engagements. Another example featured two focus groups where one included NCOs and the other included 1st Lieutenants.

Specifically, the groups were asked to describe their interactions with Afghan (or Iraqi) speakers including those with civilians, government officials, and foreign police/military personnel (relevant to their specific military function). Additionally, it was noted which interactions were frequently supported by interpreters along with those situations that regularly had little to no access to bilingual speakers. Since many of military personnel had previous deployments to Iraq, they highlighted the similarities and differences between operating within these two countries. Additionally, it was also noted where (and in what situations) the military encountered Dari-speaking Afghans as compared to Pashto-speaking Afghans. The compilation of our notes allowed us to determine which domains and which motivations should be added to the relevant pre-existing scenarios and how the pre-existing scenarios should be specifically updated.

In addition to capturing these personnel experiences, additional topics were identified (and reinforced) by examining numerous military websites including (www.mnf-iraqi.com which is now www.usf-iraq.com) monthly Army newsletters (such as *Tien Bien Times*) and Marine Corps news bulletins that reported news and experiences of Soldiers and Marines in theatre. Likewise, major online newspapers including www.washingtonpost.com and www.nytimes.com, were explored for articles. Additional information was gathered from several military documents detailing lessons learned from recent experiences in Afghanistan.

2.4. Scenario Writing and Review

The scenario writing process began with organizing the notes from the Soldier/Marine interactions/training observations and the articles referenced from the Internet. This information was then used to evolve the IA data collection scenarios into Afghan-specific data collection scenarios. Before the scenarios were finalized, they were subject to a review process. This was

enabled through a selective email distribution list. The personnel included several members of the following groups who were tasked with addressing specific questions:

- Military personnel – Is the scenario relevant and accurate from a military perspective?
- Foreign language/culture experts – Is the scenario relevant and accurate from an Afghan (or Iraqi) perspective?
- Data collection/transcription/translation experts – Will this scenario work from a data gathering perspective? Are there issues with this scenario that would make the speaker(s) uncomfortable and/or put them in danger?
- IET personnel – Will this scenario work from an evaluation perspective?

After receiving feedback, the scenario author incorporated the suggested changes, redistributed the scenarios to the group and the scenarios were finalized within several days. The scenarios were ready for use after finalization.

3. Data Collection

Bilingual scenario recordings were collected by NIST-led IET personnel in California in order to gather training and evaluation data. The data collected consisted of conversational audio between 2-way SME/FLE pairs (conversations between a SME and FLE using an interpreter). Since late 2007, NIST has collected nearly 500 h of bilingual data across over 1000 scenarios total across English/Pashto, English/Dari, and English/Iraqi Arabic pairs. The details of these collections can be found in APPENDIX C: Previous NIST Collection Output. The data collection process is presented in detail in the following subsections.

3.1. Data Collection Site Selection

NIST selected a recording studio to host the data collections based upon the following requirements:

- Be in close proximity to duty station(s) of US Marines and/or Soldiers
- Be in close proximity to the desired foreign language-speaking community
- Be available to do recordings on Saturdays and Sundays, since many of the participants have regular jobs that would prevent them from being present during the work week
- Provide three distinct functional areas:
 - Noise-proof recording areas large enough to accommodate a SME, FLE, and interpreter. Each recording area was complemented by a corresponding audio engineering room where the Quality Assuror (QA) and recording engineer could see (either through two-way video or direct line-of-sight) and talk to the speakers through their headphones.
 - An area where a SME, FLE, and interpreter could practice with a rehearsal coach prior to recording.
 - A break area where the SMEs and FLEs could relax when they were not rehearsing or recording.
- Have at least three recording studios and be capable of supporting simultaneous recording sessions where each studio's operations would not interfere with each other.
- Have both the equipment and expertise to fulfill the noise masking and *Wizard of Oz* requirements (discussed later in this document).

- Provide qualified recording engineers to capture and export the necessary audio mixes.
- Provide an office to serve as the overall coordinator's base of operations.

Also, NIST supplied its own microphones (AKG Model C420) to maintain consistent audio quality across multiple speakers and recording sessions. These microphones were selected for the following reasons:

- They had a cardioid polar pattern which “prefers” sounds arriving from in front of it (from the user's mouth). This minimizes the chance of the microphone picking up speech from other speakers or any ambient noise.
- They featured a head-worn design with a behind-the-neck headband. This allows the speakers to freely move their head where the microphone always stays in the same position relative to the speaker's mouth. The other benefit to this feature is that the speakers have their hands free to use the Wizard of Oz Setup (discussed in the following section).

3.2. Noise Masking and Wizard of Oz Setup

Noise masking was used to prevent the FLE from understanding the English spoken by the SME. The masking signal is triggered by the speech of any of the two persons (SME or FLE) and is routed to the headphones of the other person. The system diagram (see Figure 1) gives a general idea how the system is set up. In effect, when the SME speaks then the interpreter hears clearly, but the FLE gets the triggered signal, which exactly masks the SME's speech, utterance by utterance. When he/she does not speak, then masking is not triggered. Reverse Chechen (Chechen played in reverse with all the pauses deleted) provides a continuous voice for the masking signal (note that this is in lieu of the white noise specified in Figure 1). No problems with the Lombard effect were experienced as people used the natural conversational voice level when they spoke as the three participants (SME, FLE, interpreter) were standing at a comfortable distance. The natural flow of the interpreted dialogue thus seemed flawless as if no electronic equipment were present. It is important to note that the noise-masking can either be generated using separate noise-gate hardware or can be produced solely through the selected studio's audio software. The exact implementation will be dependent upon the studio's specific capabilities.

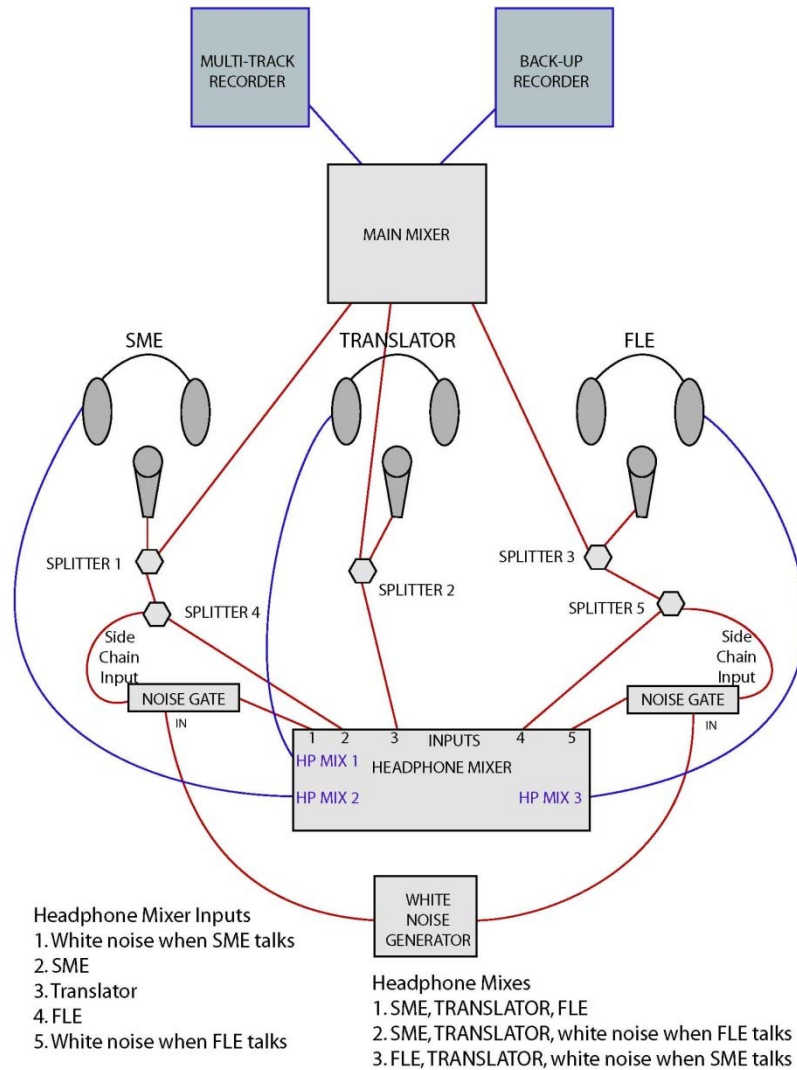
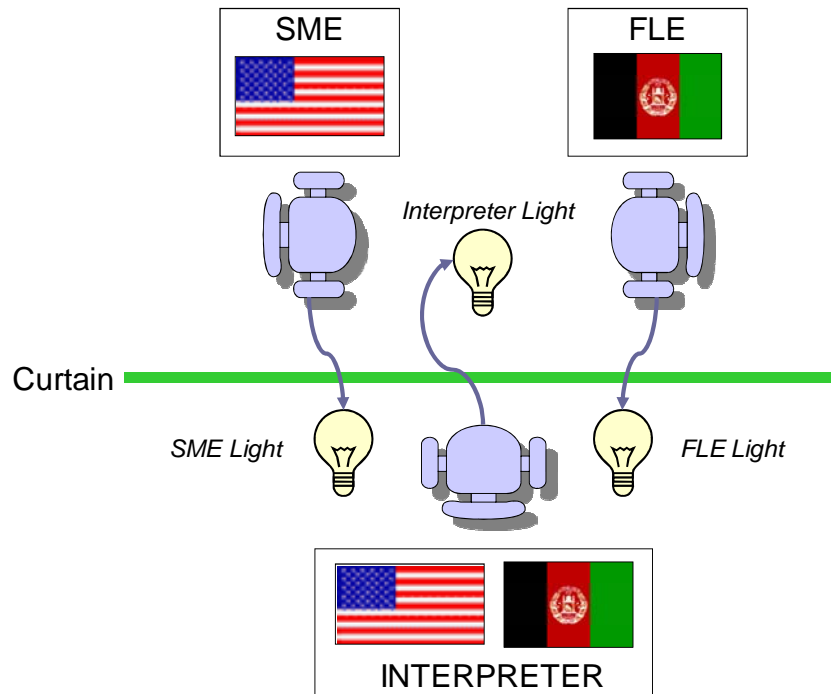


Figure 1: Data Collection Noise Masking Diagram

Once the noise masking setup was in place, the conversation flow of the speakers was addressed. The conversation is supposed to flow SME -> Interpreter -> FLE -> Interpreter -> SME -> Interpreter ->... where the SME and the FLE never speak immediately after one another and the speakers do not interrupt each other. It was also desired that the SME and the FLE speak more directly with one another (be more engaging) as opposed to directing their speech towards the interpreter. To ensure that this speaker flow remained consistent, that multiple speakers were not talking at the same time and the SME and the FLE better engage one another, the NIST team developed a *Wizard of Oz-Inspired* implementation for each recording studio. Each speaker was able to switch on/off a light to indicate to the other speakers when they are talking. Additionally, a curtain was placed between the interpreter and the SME/FLE to enhance their direct conversation with one another. Figure 2 presents the conceptual layout of the setup.



NOTES:

- Each light is to be controlled by its namesake (i.e. the SME light is controlled by the SME using a push-to-talk button, etc.)
- Only the SME and FLE can see the Interpreter's light and only the Interpreter can see the SME & FLE lights
- Chairs are shown above to represent the position of the speakers. All participants stood during the collections.

Figure 2: Wizard of Oz-Inspired Implementation to Support the Data Collections

The further details regarding the Wizard of Oz setup can be found in APPENDIX D: Wizard of Oz – Speech Flow and Implementation.

3.3. Participants

For each weekend collection event, NIST secured nine military personnel to act as SMEs and 16 bilingual foreign language speakers where eight acted as interpreters and eight acted as FLEs. These per weekend personnel numbers were selected in order to maximize the amount of data hours collected and maximize the variety of speaker voices while minimize individual speaker down-time and costs. New SMEs and FLEs were targeted for each weekend collection with the only repeats being two to three outstanding interpreters. The motivation behind repeating some interpreters was to generate some continuity between data collection events and to give the inexperienced interpreters the opportunity to learn from those that have previous collection experience.

The recruited SMEs comprised a range of backgrounds, deployments (primarily Afghanistan with most also having served in Iraq), and MOS. It was important to note their specific training and experiences prior to the collections so they could be matched up the appropriate scenario domains for each SME. The information the SME recruiters collected included the following:

- | | | |
|---------------------|--------------------------|--------------------------|
| • Rank | • Job Title | • No. of OEF deployments |
| • MOS (description) | • Current Unit | • Date/s in Iraq |
| • Service | • No. of OIF deployments | • Billets (jobs) in Iraq |

- Experience with Iraqis
- Billets (jobs) in Afghanistan
- Experiences with Afghans
- Date/s in Afghanistan

SME screening included collecting all of the pertinent information listed above. Bilingual foreign language speakers were screened via a telephone interview by a foreign language expert to ensure that their background experience and language capabilities were acceptable. Foreign language speakers who had relevant experiences but limited English ability were assigned as FLEs. Also, those foreign language speakers that did not appear to perform well during dialogue improvisation (due to a lack of domain knowledge of the dialogue topic or were not comfortable role-playing) were assigned to be interpreters.

Besides SMEs and FLEs, there were other essential data collection personnel present at these weekend events. These included three Quality Assurance (QA) personnel, two to three Rehearsal Coaches, four Recording Engineers, and one Overall Coordinator. These roles will be discussed in greater depth below.

- QA (one per studio) - This person monitors all verbal exchanges while the recording was taking place, and interrupts if problems occurred, such as:
 - Drifting off the topic
 - Using words/concepts that were too technical or too difficult which weren't consistent with the needs of the program
 - Interrupting a speaker before he/she has finished talking
 - Rehashing the same discussion topics over and over
 - Inappropriate interpreter response (interpreter was supposed to speak in first person as if he himself is the English- and Dari- speaker, not to use styles like "He said he is pleased to meet you")

If a QA detects any issues during the scenario, their first action should be to hold up the appropriate sign (see APPENDIX E: 1st Morning Briefing Points) for the speaker(s) to see so the QA does not have to verbally interrupt the conversation. If the problem persists and holding up signs does not solve the issue, the QA should then verbally break into the conversation (they should alert the recording engineer of this intent so that the engineer pauses the recording) to inform the speakers on the issue.

- Rehearsal Coach (three required on Saturdays where only two are needed on Sundays) - During the rehearsals, these individuals assist the speakers in understanding their motivations. They then conduct high-level rehearsals (making an effort to minimize their involvement in the conversation) to ensure that the participants fully understood the roles they were playing and had enough information to produce a successful dialogue.
- Recording Engineer (one per studio + one spare engineer) - Each studio is staffed by a recording engineer (provided by the recording studio), whose responsibilities included assuring that participants were wearing microphones and headsets correctly, speaker volume and noise levels were calibrated, and files were saved and named according to specific naming conventions. At the end of each recording, the Recording Engineer outputs two speaker wave file mixes (one mix contained the SME on the left channel and the FLE on the right channel while the other mix contained the SME and FLE on the left channel and the interpreter on the right channel). Each of these output mixes has a purpose. The two-speaker mix mentioned above has benefit to the research teams since they receive data from the SME and the FLE on separate audio channels. The three-speaker mix has both benefit to the research teams, the transcription/translation

service provider and the evaluation team, since the entire conversation can be easily followed and understood (since the interpreter is included).

After these files are created, they are uploaded to the studio's server where the files were organized by the spare recording engineer and stored until downloaded to IET drives. Besides verifying and organizing each studio's data, the spare recording engineer provided short breaks to the studio engineers so they could get something to eat, use the restroom, etc. To ensure that all recording engineers at all studio facilities performed their tasks consistently, an audio recording specialist from the IET trained them, oversees their activities, and checks their deliverables during the first few data collections. After that, the Overall Coordinator assumed these IET responsibilities.

- Overall Coordinator - This role included acts as liaison to the manager of the recording studio for all business-related matters and providing instructions to participants at the beginning of the day, which includes a walk-through of the recording facility. The Overall Coordinator also handles the scheduling duties. This includes implementing an efficient flow of participants from the waiting area to rehearsal areas to recording areas that required an established schedule at the beginning of each recording day. They monitor the availability of rehearsal rooms (and rehearsal coaches); when a given FLE-SME-Interpreter team finished rehearsing, they vacate the rehearsal room to wait for a recording studio to become available, and the coordinator sent a new FLE-SME-Interpreter team to begin rehearsing. Maximum efficiency is achieved by diligently monitoring both the availability of rooms and the whereabouts of participants.

3.4. 1st Morning Kickoff

The 1st morning of a data collection event is critical in terms of ensuring that everyone including SMEs, FLEs, interpreters, evaluation personnel and studio staff are aware of their responsibilities throughout the weekend. This initial morning brief can last anywhere from 45 min to 75 min depending upon the amount of questions asked. The talking points for the morning brief can be found in APPENDIX E: 1st Morning Briefing Points.

3.5. Workflow

Since there were three studios used during each data collection event, the work-flow was designed to minimize studio downtime. There were always three speaker groups (each group consisted of a SME, FLE, and interpreter) in rehearsal at any given instant to ensure that when a studio became available, it was immediately utilized. This also covered the extreme case when all three studios became available at the same time. Figure 3 presents the workflow at a high level.

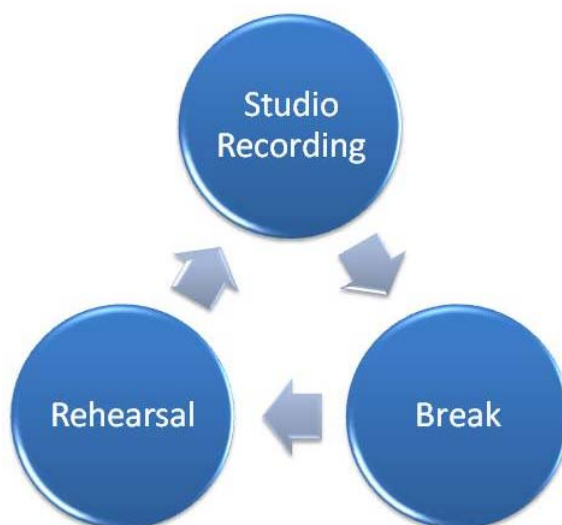


Figure 3: Workflow for Recording Sessions

Once the kickoff meeting concluded, the OC assigned speaker groupings with scenarios and sent them back to the rehearsal room. Specifically, the OC used *Sticky Notes* where a single note contained all of the necessary information for a group to conduct a scenario: SME #, FLE #, INTERPRETER #, and Scenario Code (alpha-numeric that designated Domain/SME Motivation/FLE Motivation). The SME took the *Sticky Note* from the OC and took his assigned FLE and INTERPRETER to rehearsal.

Once in the rehearsal room, the speakers were given their own motivation (the INTERPRETER was given a copy of the FLE motivation) while their ability to see their counterpart's motivation was strictly scenario dependent. For those scenarios where the speakers are assigned roles where they are collaborating with one another, conducting a training session and/or gather information, the speakers could see each other's motivations since they'll rehearse their scenarios together. Those scenarios where it's desirable to maintain maximum spontaneity and the speakers' roles have little to no information regarding the person they are about to converse should require the speakers to (1) rehearse separately and (2) not be informed of their counterpart's motivation. After the speakers have practiced parts of their dialogue, everyone is aware of the key vocabulary that could be used, and a studio is available, the three enter the sound booth for their dialogue. It's up to the rehearsal coaches to monitor the overall situation in the rehearsal room to see which group(s) is ready for the next studio opening and which group(s) requires some assistance. Also, it's important to strongly encourage the speakers to take notes on their individual motivation sheets. This minimizes the chances of the speakers forgetting conversation topics and dialogues ending quickly.

After a speaker group returned to the break area following their recording, they were re-grouped such that the SME would interact with a different FLE/interpreter pair during their next scenario collection. The FLE's and interpreters maintained their same pairing throughout the weekend (since not all FLE's could be interpreters and vice-versa). The reason FLE and interpreter pairings were maintained was to further ensure the speakers were comfortable in their dialogues and allow them to build their own rapport to facilitate smooth translations. An example schedule and rotation can be found in Table 2. Note that the *Scenario #*, *SME*, and *TERP*, *FLE* information are provided to each

speaking trio as they are summoned to rehearsal. Once they finish their recording and go on break, the OC notes the studio they used.

Table 2: Example Schedule and Rotation

SCENARIO #	SME	TERP, FLE	STUDIO
A23	1	2, 1	A
F13	2	4, 5	B
E61	4	10, 13	C
B11	3	8, 3	C
D31	5	6, 7	A
D13	7	12, 11	B
C31	8	14, 9	B
D11	9	16, 15	C
B41	10	2, 1	A
C43	6	4, 5	B
F32	1	6, 7	A
F21	2	10, 13	B
F22	5	12, 11	C
A13	4	14, 9	C
C22	7	8, 3	B
E41	8	2, 1	A
E52	3	4, 5	C
D22	10	16, 15	B
C12	6	12, 11	A
C43	9	10, 13	A
B12	5	6, 7	B

3.6. Subsequent Morning Kickoffs

After the first day of collection, the following morning(s) briefs should be much shorter than the previous day. The only discussion points for the following morning(s) include keeping everyone excited and energized about this next day of collection and discussing any areas and/or situations that should be improved. Beyond that, the SMEs and FLEs should be familiar enough with the data collection protocols that the only thing participants are waiting on is the OC to inform them of their pairing and scenario assignment.

3.7. File Naming Convention and Post-Collection Delivery

Because the recording engineers create the two specified waveform audio format (.wav) file mixes for each scenario, they are labeled according to their identifying information. Figure 4 presents the specific naming convention that the NIST IET has applied to the audio files they have collected. The audio data was output as wave format at 44.1 kHz, 16-bit resolution.

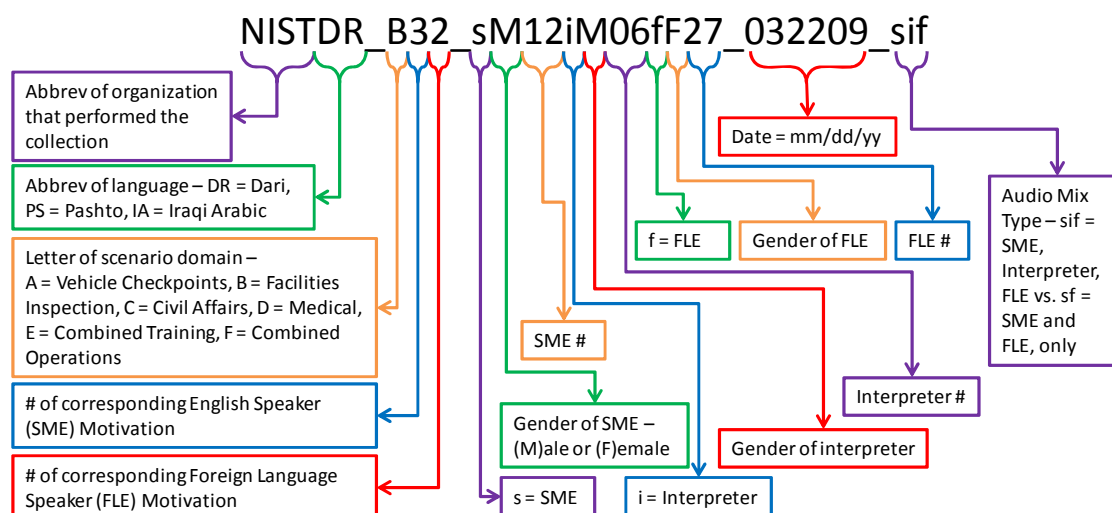


Figure 4: Data Collection Audio File - Naming Convention

Once the recording engineer output the necessary sound mixes and uploaded them to a central ftp server, they were packaged and the external transcription/translation service provider was alerted that the files were ready for ftp transfer.

3.8. Transcription/Translation

Once the external transcription/translation service provider downloaded all of the audio files via ftp, they transcribed and translated the conversations in addition to creating a lexicon of the foreign language words. An example of guidelines used in past efforts can be APPENDIX F: Example Transcription/Translation Guidelines. On average, the transcription and translation of a weekend collection took between two to four months to complete. This process included:

1. Bilingual personnel transcribe the conversations according to the guidelines.
2. Quality assurance personnel review the transcriptions for their accuracy and conformation to the guidelines.
3. Bilingual personnel translate the transcriptions into the other languages (for example, English -> IA and IA -> English).
4. Quality assurance personnel review the translations against the transcriptions and/or audio files for accuracy and completeness.

Additionally, the words captured in each weekend collection are funneled into the creation of the lexicon. There are numerous factors that affect the estimated delivery schedule of the transcriptions and translations including:

- When in the effort the data is captured and processed - earlier collections tend to take longer since the transcription/translation team is still getting up to speed while later collections are processed at a quicker pace).
- The specific foreign language that is being processed – each transcription/translation service provider has different capabilities in different languages and therefore has personnel with varying expertise.

- Amount and quality of data collected during that particular weekend – the more data collected, the longer it takes to process. Likewise, if some of the speakers talk too quickly, stutters, mumble, etc, then it will take the transcription/translation personnel longer to process this data.

The transcriptions and translations are output in .tdf file format and contain the following information:

- File (Column A) - Presents the file name (using the convention described above).
- Channel (Column B) – Presents the channel # of the speaker.
 - 0 = SME speaking English
 - 1 = FLE speaking foreign language
 - 2 = Interpreter speaking English
 - 3 = Interpreter speaking foreign language
- Start (Column C) – Time (in seconds) that the speaker began their utterance.
- End (Column D) – Time (in seconds) that the speaker concluded their utterance.
- Speaker (Column E) – Speaker’s information according to the naming convention above.
- Speaker Type (Column F) – Gender of the speaker.
- Speaker Dialect (Column G) – Dialect of the particular speaker. Provided for each FLE by the data collection team.
- Transcription or Translation (Column H) – Utterance of the corresponding speaker.
- Section (Column I) – Language specific – defined in guidelines
- Turn (Column J) – Language specific – defined in guidelines
- Segment (Column K) – Language specific – defined in guidelines
- Section Type (Column L) – Language specific – defined in guidelines
- Su Type (Column M) – Language specific – defined in guidelines
- Speaker Role (Column N) – interviewer (role of the SME), interpreter, or respondent (role of the FLE).

3.9. Data Usage

Typically, a majority (if not all) of the output data from the collections, including audio, transcription and translation files, is provided to the technology developers so that they may train their systems on the collected languages and dialogues. In most cases, some of this data is held back from the technology developers by the evaluation team to test the technologies’ speech-to-speech translation capabilities. This forms what is known as the representative set where a small percentage that is indicative of the overall data is solely kept by the evaluation team for testing purposes.

4. Cost

The numbers in this section are provided for estimation purposes only. The costs that another group may incur could be significantly lower or higher than the costs listed below.

Table 3 shows the cost a weekend bilingual collection, as described in the previous sections. In the estimates, we are assuming a yield of 35 h of audio data over the weekend. These estimates do not include the cost of scenario creation or equipment purchases (e.g., for the Wizard of Oz setup), which are often one-time fees and are not associated with individual data collections.

Table 3: Cost Per Data Collection

Item	Cost	Notes
Studio Rental per Weekend	\$ 8,000.00	3 recording booths, all studio personnel
Administration Cost	\$ 15,000.00	data collection administration team travel and time, assumes two people
SME personnel	\$ 16,000.00	travel, lodging, and hourly fee for nine SMEs
FLE personnel	\$ 25,000.00	travel, lodging, and hourly fee for eight FLEs, eight interpreters, and three QAs
Recruiting fee	\$ 4,000.00	for SMEs and FLEs
Transcription	\$ 50,000.00	assumes 35 hours of audio @ \$1,390 per audio hour
Translation	\$ 37,000.00	assumes 170,000 words @ 0.22 cents per word
Lexicon Development	\$ 15,000.00	assumes approx. 17,000 unique words @ 0.88 center per word
Total	\$170,000.00	

5. Conclusion

The data collection process, including scenario development and data transcription/translation, has been led and implemented by NIST since 2007. This document has presented the NIST IET's best practices after many individual processes have been developed and refined through numerous collection efforts. These methods and techniques have been used successfully to collect and process large amounts of speech-to-speech data.

APPENDIX A: Example Iraqi Data Collection Scenarios

A – TRAFFIC CONTROL POINT/VEHICLE CHECKPOINT

- ☐ **SME Motivation 1 – NO SUSPICIOUS ACTIVITY** You are manning the checkpoint and have been advised that there has been no suspicious activity of note in the area.

TALKING POINTS

- a) Intended Destination
- b) Personal Information
- c) Vehicle Occupants (if any)
- d) Vehicle State
- e) Vehicle Contents

FLE Background (applies to all) – You are driving in your vehicle approaching an American-controlled checkpoint in a city/area South of Baghdad.

- o FLE Motivation 1 – You have just left your house and are on your way to work. You are supportive of the American presence and appreciate the security they provide to your town. **NOTHING FOUND** or **SHOVELS AND PICKS** or **VIDEO CAMERA**
- o FLE Motivation 2 – You have just received a call that your father is very ill and you are rushing to his house immediately. You have a neutral attitude towards the American presence. You do not care for the American checkpoints and have always seen them as a big waste of time. **150 POUNDS of FERTILIZER** or **MEDICAL SUPPLIES – BOTTLES FULL OF PILLS** or **MOTOROLA TALK-ABOUT RADIOS**
- o FLE Motivation 3 – You dislike the Americans being here and have some friends that are insurgents. You are transporting a large sum of money on their behalf. You were not aware that this checkpoint was active and are hoping to make it through without them finding the money. **\$10,000 of US DOLLARS** or **\$750 of US DOLLARS** or **EMPTY HIDDEN STORAGE COMPARTMENT UNDER BACKSEATS**
- o FLE Motivation 4 – You are an active member of a local insurgent cell and are preparing for your next attack on the Americans. You are smuggling wires and timers to another insurgents house. You realize that the Americans may find these items but are hoping to convince them that you work for a local electronics store. **WIRES AND TIMERS** or **LAMP CORD** or **5 WASHING MACHINE TIMERS**

- ☐ **SME Motivation 2 – IED HOTSPOT** You are manning a checkpoint in the interior of Iraq that is an IED hotbed. You have been advised to thoroughly search all vehicles and screen all personnel passing through.

TALKING POINTS

- a) Intended Destination
- b) Knowledge of Recent IED Activity
- c) Knowledge of Suspected Insurgents
- d) Vehicle Occupants
- e) Vehicle Contents

FLE Background (applies to all) – You are driving in your vehicle approaching an American controlled checkpoint in a city/area West of Baghdad (not near any foreign borders).

- FLE Motivation 1 – You are on your way to your grandfather's house for dinner. You do not like the Americans occupying your homeland, but only believe in non-violent means of protest (you do not support the insurgency). **SUITCASES OF CLOTHES** or **BOXES OF BOOTS** or **6 ROLLS OF DUCT TAPE**
- FLE Motivation 2 – You are a mechanic working at a local auto repair shop. You have just left work for the day and are anxious to get home after your long day. You are glad that the Americans are in your country because they have rebuilt much of your neighborhood that Saddam's men have destroyed in the past. **CUTTING TORCH** or **POWER DRILL AND SAW** or **BOX OF RADIO PARTS**
- FLE Motivation 3 – You are on your way home from work as a guard at a nearby prison. You routinely take your pistol home with you to protect your family. You know you are not supposed to do this, but your commander turns a blind eye to it since he sometimes does it, too. You have a neutral attitude towards the Americans. **PISTOL** or **EMPTY PISTOL MAGAZINE THAT SMELLS LIKE POWDER** or **TRASH BAG FULL OF BLUE UNIFORM SHIRTS**

B - FACILITIES INSPECTION

- **SME Motivation 1 – POLICE STATION** You are meeting with the chief of police to inspect the station under his command. You are very concerned about the morale of the officers and have heard rumors of possible equipment theft from the station.

TALKING POINTS

- a) Greetings/Pleasantries
- b) Morale/State of Police Force
- c) Condition of Building
- d) Equipment/Weapons Status
- e) Emergency Vehicle Status
- f) Issues with Theft

- FLE Motivation 1 – You are the local police chief who is meeting with American forces to brief them on the status of your station. Lately, not all of the officers have been showing up for work and have noticed that some weapons and ammunition has gone missing over the past few weeks. You know this doesn't look good for you but you just want to make your neighborhood/town safe and can't do anything about it but keep trying to recruit and train more police. If possible, you would like to get replacement supplies (specifically, weapons, ammunition, helmets, and body armor) to outfit your men. **HELMETS AND ARMOR MISSING**
- FLE Motivation 2 – You are the local police chief who is meeting with the American forces to give them a tour of your facility. Although not all of the officers showed up for work today, you believe they are happy with their job because they are well-paid. You are concerned that the American forces may ask about the recent disappearance of equipment because you have taken some of the weapons and ammunition home from the armory. You are in need of new parts to support your police vehicles or would like to purchase new vehicles for your force. **PISTOLS MISSING**

- **SME Motivation 2 – POWER PLANT** You are meeting with the head of operations of the local power/electrical plant to get a status report on the facility. You are very concerned about its security and want to verify that measures are in place to protect it from the insurgents.

TALKING POINTS

- a) Greetings/Pleasantries
 - b) Facility Status
 - c) Equipment Needs
 - d) Security Concerns
 - e) Personnel Status
- FLE Motivation 1 – You are the head of the local power/electrical plant. Although your facility is relatively new (the Americans rebuilt it last year after it was damaged in the war), you believe it is still somewhat vulnerable to insurgent attacks. You have contracted out for extra security guards, but you deem them unreliable (they have been caught sleeping on the job) and would like to obtain the latest technology (flood lights, cameras, motion sensors, etc.) to enhance your security.
 - FLE Motivation 2 – You are the head of the local power/electrical plant. Your facility is at least 20 years old and would benefit from some current technological upgrades. Security is a major concern for you. You have recently fired your contracted security guards because they have abandoned their posts numerous times. You want to know what the Americans can offer/provide you (both in personnel and equipment) with respect to security.
- ❑ **SME Motivation 3 – WATER TREATMENT** You are meeting with the head of the brand new local water treatment facility to verify its status (it should be fully operational). You want to know if it is functioning according to specification (properly treating the sewer water) and if its current workforce is sufficient.

TALKING POINTS

- a) Greetings/Pleasantries
 - b) Personnel Status/Needs
 - c) Equipment Status/Needs
 - d) Current Security Measures
 - e) Outstanding Issues
- FLE Motivation 1 – You are the head of the new local water treatment facility and are receiving some American forces who want a status report. Your facility has been running smoothly since it was first brought online several months ago. However, your workforce is unreliable and does not always show up for work. Since the Americans still offer financial support for the workers you want to see if you can offer them more money as incentive for them to show up.
 - FLE Motivation 2 – You are the head of the new local water treatment facility and receiving some American forces who want a status report. Over the past few weeks your facility has suffered numerous breakdowns and has not been working as it should. You are unsure whether it is due to faulty equipment (some generators and pumps have broken) or sabotage (you have heard rumors that the insurgents may target your facility and that a former worker at the plant has joined forces with them, i.e. they now know less obvious ways to sabotage the plant such as putting objects in the in-flow). Security is a big concern that you have.

APPENDIX B: Example Afghan Data Collection Scenarios

C – CIVIL AFFAIRS

- ☐ **SME Motivation 1 – CIVIL AFFAIRS DESK** - You are manning a civil affairs desk that deals with the complaints/requests of the Afghans off of the street. You are responsible for getting as much information about the situation/complaint as possible from the Afghan. At the end of your discussions, you provide the Afghan with a case reference number and inform them that they should check back with you in a week or so after you have had time to investigate their claim.

TALKING POINTS

- a) Request Situation Explanation
 - b) Requested Compensation (jewelry, weapons, money, return of people, etc.)
 - c) Verify Story by Requesting Additional Information/Re-asking Questions (Ensure that the story remains consistent and is not fabricated)
- FLE Motivation 1 – American forces searched your house the other day and broke your door frame when they entered. You are deeply upset by this intrusion and see this as an opportunity to have the rest of your house rebuilt after it was damaged years ago by the Russians. You have just entered the local civil affairs office and are going to request that the Americans not only pay for a new door frame, but also for additional items including a new door, several walls, and a portion of your roof.
 - FLE Motivation 2 – American forces searched your farm the other day and confiscated a stockpile of weapons you had been storing in your barn. Specifically, they took several AK-47's, some magazines, and one RPK. You have just come down to the local civil affairs office with a receipt the Sergeant provided you when he took the weapons. Although these all belong to you and you are aware that there are limits on the number of weapons you can possess, you will demand to have them back. You have accumulated these weapons over the years for protection from the Taliban and need to get them back immediately since the local police do not provide adequate security.
 - FLE Motivation 3 – American forces searched your house the other day and took your brother from the house. The Americans claimed he is a suspect in some local criminal activity and that they took him for questioning. You have just come down to the local civil affairs office to find out where he is and when he will be released. You are confident that your brother is not involved with any criminals and that he is innocent. You want him released as soon as possible.
- ☐ **SME Motivation 2 – SWET SURVEY** - You are going door to door in a local neighborhood to determine the population's quality of life. You also want to inform the residents that you and your team will be frequently patrolling the area to enhance their security, improve their community, etc.

TALKING POINTS

- a) Greetings
- b) Family Demographics (Head of House/Hold Data, Wife/Wives, Children, Other Family Living in House)
- c) Quality of Life Assessment - Utilities (Water, Trash, Electricity)
- d) Town Authority Figures and Their Impact (Village Elder, Mullah, Malik, etc.)
- e) Criminal Activity (if any)
- f) Outstanding Issues

- FLE Motivation 1 – A group of Soldiers/Marines is approaching your house. You have seen them go door to door in your neighborhood during the past hour. You are unsure what they want, but are hoping they can help you with electricity problems. The local service has been unstable for the past month and your backup generator is very old. Other than that, you have no complaints. Your wife and children are very happy and healthy. However, the children's school has yet to be repaired after being damaged by the Taliban several years ago.
- FLE Motivation 2 – A group of Soldiers/Marines is approaching your house. You have seen them go door to door in your neighborhood during the past hour. You are eager to welcome them into your home. You would like to see what they can do to help you in getting your trash removed on a regular basis. You know that the sub-governor is in charge of hiring people to remove the trash, but believe he is corrupt and not doing his job. The trash has been piling up all over the neighborhood and some of your kids have been getting sick.
- FLE Motivation 3 – A group of Soldiers/Marines is approaching your house. You have seen them go door to door in your neighborhood during the past hour. You do not know what they want, but you are glad they are patrolling because you suspect your neighbor across the street is an insurgent. You have seen many men come and go from his house carrying wooden crates (some with wires hanging out of them) several nights ago. You have kept your house in good condition (all of the utilities work most of the time) and are only concerned about this neighbor.

- ❑ **SME Motivation 3 – CENSUS** - You are going door to door in a local neighborhood conducting census operations. The goal of this mission is to understand the makeup of the local families to better serve the needs of the population.

TALKING POINTS

- a) Greetings
 - b) Family Demographics (Head of House/Hold Data, Wife/Wives, Children)
 - c) Home Demographics (No. of people living in house, No. of middle-aged males, Crops grown, etc)
 - d) Town Information (Village name, elder/tribal leader, water source, mosque, type of buildings, etc)
 - e) Education (Highest Education Completed, Location of School(s) Attended, Dates of Schooling, Wife's Education, Children's Education, etc)
 - f) Employment (Current Job, Length of Time at Job, Location, Responsibilities, Previous Work, etc)
 - g) Knowledge of Suspicious or Criminal Activity
- FLE Motivation 1 – A group of Soldiers/Marines is approaching your house and it appears they have been asking questions to other people in the neighborhood. You have a shop in the market where you sell motorcycle and car parts. In your spare time, you also repair broken vehicles since you are well-known in the village as a mechanic. You never received any formal training, rather you learned from your father. In turn, you teach your teenage sons about vehicles and hope they will become great mechanics like yourself.
 - FLE Motivation 2 – A group of Soldiers/Marines is approaching your house and it appears they have been asking questions to other people in the neighborhood. You are supportive of the US Military since they have been fighting the Taliban. This has been very beneficial since one of the crops you grow is poppy and the Taliban have been taxing your harvests for many years. You expect your family to work on the farm because it's your livelihood. Each week you take your crops (including poppy along with some fruits and vegetables) to the bazaar to sell so you can buy food and supplies for your family.

F – COMBINED OPERATIONS

- ❑ **SME Motivation 1 – PLANNING A RAID** - You are meeting with a member of the ANA/ANP to plan a raid (*foreign language speaker will provide details on location and intent of raid*). Your goal in this meeting is to ensure that the officer is competent to lead/execute this operation (*level of US assistant during operation will be agreed upon with foreign language speaker*).

TALKING POINTS

- a) Objective of the mission (make arrests, seize materials, etc.)
 - b) Concept of the operation (route taken to the house, who will go into the house, who will provide security, route taken back, what will be done with detainees and/or contraband)
 - c) Perimeter / containment area (establishing 360 degree security)
 - d) Key terrain features (buildings, hills, dead space, bottlenecks, / danger areas, etc.)
 - e) Arrest and detention procedures
 - f) Evidence handling procedures
 - g) Expected response / level of resistance
 - h) Force and materials required for the operation (unit size, special skills, weapons, restraints, etc.)
 - i) Mission timeline (high level - expected duration)
 - j) Potential collateral damage (civilians, property, etc.)
- FLE Motivation 1 – You are a member of the Afghan National Army who is meeting with American Forces to plan a raid on a heavily-walled compound that the Taliban has been using as a bomb factory. Your men have gathered intelligence that this is the meeting spot for the local insurgent cell and where IEDs are assembled. In addition to containing IED materials including diesel fuel and fertilizer, your intelligence also says that the occupants may be armed with AK-47s and RPGs. You would like the Americans' help in planning this raid so that you capture as many insurgents as you can without getting any of your men injured. You will be asking questions related to the planning and execution of a raid on this location.
 - FLE Motivation 2 – You are a member of the Afghan National Army who is meeting with American Forces to plan a raid on a gas station/mechanic shop. You suspect that the shop is being used to manufacture IEDs (Improvised Explosive Devices) and supporting materials for the insurgency. One of your men has seen copper shavings, cutting torches, welding equipment, excessive amounts of wire, etc. while having his car inspected there. You will be asking questions related to the planning and execution of a raid on this location.
 - FLE Motivation 3 – You are a member of the Afghan National Army who is meeting with American Forces to plan a raid on an abandoned house that is a suspected storage and/or transfer point for illegal goods (*speaker may select from weapons, drugs, IED materials, etc. at their choosing*). Your men have observed a group of young males going in and out of this house for the past few days, none of whom are related to the previous occupants. The neighbors have reported frequent night-time activity (*speaker can describe reported activity*). You will be asking questions related to the planning and execution of a raid on this location.
 - FLE Motivation 4 – You are a member of the Afghan National Army who is meeting with American Forces to plan a raid on an industrial complex situated within a heavily populated urban area. You have intelligence that a warehouse within the complex is being used to store stolen (black market) chemicals for use in the construction of explosives (*foreign language speaker will provide details of intelligence reports and surveillance photographs*). You will be asking questions related to the planning and execution of a raid on this location.

- ❑ **SME Motivation 2 – CORDON AND KNOCK/SEARCH** - You are meeting with a member of the ANA/ANP to plan a cordon and knock/search (*foreign language speaker will provide details, including location, of the operation*). Your goal is to ensure that the Afghan officer and his men understand the mission and are capable of carrying it out (*level of US assistant during operation will be agreed upon with foreign language speaker*).

TALKING POINTS

- a) Objective of the mission (warn citizens of a problem, gather intelligence, search for weapons or enemies, etc.)
 - b) Available Intelligence (*foreign language speaker will provide a range of information*)
 - c) Search procedures for homes and businesses (ROE, security aspect, etc.)
 - d) Arrest / detention procedures
 - e) Perimeter / containment area
 - f) Key terrain features (buildings, hills, dead space, bottlenecks / danger areas, etc.)
 - g) Force structure required for the operation (unit size, special skills required, etc.)
 - h) Expected response of residents (who is friendly, who is fearful, etc.)
 - i) Evacuation procedures for nearby residents (where to take them temporarily, supplying them with food and water, transportation, if necessary etc.).
- FLE Motivation 1 – You are a member of the Afghan National Army who is meeting with American Forces to plan a cordon and search for a known Taliban leader. You have received reports from another ANA unit that the suspect was recognized at a nearby traffic checkpoint, but escaped. You are able to provide information from residents and/or neighbors of the search area, and results of conversations with the checkpoint personnel who made the report. Your intent is to jointly plan the search with the US forces, although your men will be the ones to carry out this mission.
 - FLE Motivation 2 – You are a member of the Afghan National Army who is meeting with American Forces to plan a cordon and search for weapons and ammunition that have been reported missing from the local Afghan outpost. You have received a tip from a local citizen that the weapons have been stored in several caves at the base of a mountain just on the outskirts of town. You want to act quickly believing they will attempt to smuggle them out of the area very soon (*foreign language speaker will provide details of the surrounding area, reliability of the tip, etc.*). You are unsure if insurgents are presently in the cave with the weapons, so you would like to request the assistance of the American forces for the operation.
 - FLE Motivation 3 – You are a member of the Afghan National Army who is meeting with American Forces to plan a cordon and knock on a local metal shop. You have received multiple anonymous tips that the metal shop has been supplying local criminals with IED materials including steel pipes and scrap metal although the shop itself is known not to contain any explosive materials. The goal of the operation is to question the owner of the shop without arousing suspicion to get information about his customers. You will ultimately arrest him once you have identified the criminals. You and your men are very familiar with the neighborhood, so you would like them to lead the operation with the Americans in a support role.

APPENDIX C: Previous NIST Collection Output

ENGLISH <-> PASHTO				
2009 NIST Pashto Collections			AVG. MIN PER SCEN.	SCEN. TOTALS
DATE	DATA (GB)	HOURS TOTAL		
2009 - April 25-26	49.79	39.00	24.38	96
2009 - May 30-31	50.03	39.19	21.98	107
2009 - July 24-25	50.31	39.41	26.87	88
Pashto Totals	150.13	117.60	24.41	291
ENGLISH <-> DARI				
2009 NIST Dari Collections			AVG. MIN PER SCEN.	SCEN. TOTALS
DATE	DATA (GB)	HOURS TOTAL		
2009 - Feb 28-Mar 1	48.39	37.91	20.68	110
2009 - Mar 21-22	48.56	38.04	23.29	98
Dari Totals	96.95	75.94	21.98	208
ENGLISH <-> IRAQI ARABIC				
2008 NIST Iraqi Arabic Collections			AVG. MIN PER SCEN.	SCEN. TOTALS
DATE	DATA (GB)	HOURS TOTAL		
2007 - Dec 8-9	23.04	18.00	16.12	35
2008 - Jan 11-12	41.59	32.58	16.15	121
2008 - Jan 25-26	21.96	17.20	16.13	64
2008 - Feb 9-10	44.52	34.87	17.01	123
2008 - Mar 1-2	23.14	18.13	21.33	51
2008 - Apr 19-20	49.08	38.45	30.35	76
2008 - June 28-29	42.75	33.49	22.58	89
Iraqi Arabic Totals	246.08	192.71	19.95	559

APPENDIX D: Wizard of Oz – Speech Flow and Implementation

1. The SME presses (and holds) their button which triggers the *SME Light* to turn on. The INTERPRETER now knows that the SME is about to speak since the *SME Light* is on and the FLE knows that the SME is about to speak because they can see they have pressed their button (the button has a light-up LED to show when it is engaged). **SME speaks.**
2. After the SME has finished speaking, they release their button which turns off the *SME Light*.
3. Once the INTERPRETER sees that the *SME Light* is off, the INTERPRETER activates the *TERP Light*.
4. With the *TERP Light* is on, the SME and FLE can clearly see that the INTERPRETER is about to speak. **INTERPRETER speaks.**
5. After the INTERPRETER has finished speaking, they release their button which turns off the *TERP Light*.
6. Once the FLE sees that the *TERP Light* is off, the FLE activates the *FLE Light*.
7. With the *FLE Light* on, the INTERPRETER can clearly see that the FLE is about to speak (likewise, the SME can see the FLE has pressed their button and is about to speak). **FLE speaks.**
8. After the FLE has finished speaking, they release their button which turns off the *FLE Light*.
9. Once the INTERPRETER sees that the *FLE Light* is off, the INTERPRETER activates the *TERP Light*.
10. With the *TERP Light* on, the SME and FLE can clearly see that the INTERPRETER is about to speak. **INTERPRETER speaks.**
11. After the INTERPRETER has finished speaking, they release their button which turns off the *TERP Light*.
12. Once the SME sees that the *TERP Light* is off....**go back to Step 1...**



APPENDIX E: 1st Morning Briefing Points

✓ THANK YOU FOR BEING HERE TODAY!

✓ INTRODUCTION of KEY PERSONNEL...

- Overall Coordinator (OC)
- Marine/Soldier Recruiter(s)
- Foreign Language Expert Recruiter(s)
- Rehearsal Coach(es) (RC)
- Additional team personnel

MFLTS PROGRAM – OVERVIEW

✓ MFLTS – stands for Machine Foreign Language Translation System

✓ PROGRAM GOAL – Development of two-way, free-form, spoken language communication technologies to enable US military personnel to communicate with foreign language speakers within tactical domains without the need for an interpreter.

✓ In order for technologies to be accurately and effectively developed, they must be trained on current, relevant dialogues and they must be evaluated against such dialogues, as well.

✓ DATA COLLECTION GOAL - Collect tactically-relevant conversations between English-speaking US military personnel and speakers of specific foreign languages (*note the language of the weekend collection*) to support system training.

✓ Marines and Soldiers playing the role of English speakers are referred to as SMEs during the collection

✓ Foreign language speakers are referred to as FLEs (for Foreign Language Experts) during the collection

✓ DATA COLLECTION FLOW

- For each recording, a SME will be assigned to a FLE/INTERPRETER pair with specific SME and FLE motivations (OC will provide these). Specifically, the OC will provide the SME with a sticky note of the pertinent information.
- Upon receiving their assignments, the speakers will go into the training room and be given their motivations by a trainer (RC and/or SME recruiters). The SME will show their sticky note to the trainer so they can receive the appropriate materials
- The three speakers will discuss the scenario (unless it's a checkpoint scenario), practice some of the dialogue and ensure that the speakers are aware of any key words that could be used
- For checkpoint scenarios, the speakers will only be together for a minute (or so) so that they can all see the *Vehicle Settings*. After the SME makes a note of these, the SME will go into a

separate room to practice their dialogue. The FLE and the TERP will remain in the rehearsal room with their hide. The trainer will determine a suitable location for the hide to be found and will let both the SME and the FLE (separately) know where the hide is (so that the SME knows what to search and the FLE knows when to reveal the hide).

- When a studio becomes available and the speakers are next in the queue, they will go into the studio and record their dialogue (the SME should hand their sticky note to that studio's recording engineer). Dialogues should be no less than 10 min and no longer than 30 min. Each recording will be monitored by Quality Assurance personnel to ensure that the scenario is on topic, etc. The QA may interrupt the dialogue at anytime if certain issues (discussed below) arise.
- After the recording is completed, the SME is to retrieve the sticky note from the recording engineer and return it to the OC. The SME will confirm the engineer noted which studio A/B/C was used.
- The speakers are to return to the break room and await their next assignment.

GENERAL

- ✓ Everyone should conduct themselves in a professional manner.
- ✓ No cell phones are allowed in the studios and rehearsal areas. Please leave your phone with the OC prior to going back for rehearsal (you may retrieve it from the OC after your session is done being recorded) or put it in your car – NO EXCEPTIONS
- ✓ Reminder to everyone – there is no official time for lunch break. Rather you should eat on your rotational break after food has arrived. Everyone should expect approximately 10-25 min of break time in between their previous recording and their next rehearsal. But on some occasions you will be called directly back into the studio. If you grab some food after sitting for 15 min, you should reasonably expect that you will be called into another rehearsal before you finish eating.
- ✓ Additionally, there are no official times for smoke breaks. If you have to smoke, do so outside away from the building entrance and you should do so immediately after your recording session. That way, you are back inside the building and in the break area so that the OC can easily find you for your next rehearsal.

REHEARSAL and RECORDING-SPECIFIC

✓ IT IS THE RESPONSIBILITY OF BOTH SPEAKERS TO MAINTAIN A PRODUCTIVE, SUBSTANTIAL CONVERSATION...THIS STARTS BY HAVING AN EFFECTIVE REHEARSAL

- English speaker should inform foreign language speaker of various tactical vocabulary that could be used in the scenario so that both the FLE and the INTERPRETER are knowledgeable of the word in the target language
 - Both speakers should make notes during rehearsal:
 - English speakers should augment the talking points based upon their own experiences
 - Foreign language speakers should also be taking notes on their back story (who they are and what they are doing) and specific tactical vocabulary
1. SMEs and FLEs should always speak in 1st person (I want vs. tell him that I want, etc.).
 2. INTERPRETERS should translate the SME and FLE speech, verbatim. Do not paraphrase, condense or exaggerate any dialogue.
 3. All speakers should speak naturally without speaking fast or quickly.
 4. SMEs and FLEs should conduct a reasonable amount of rapport building (no more than 3 min) in the beginning of those scenarios where it would reasonably occur. For example, rapport building is natural between military officers prior to planning an operation, but is unlikely between US military personnel and a person being stopped at a checkpoint.
 5. SMEs and FLEs should avoid *Yes* and *No* responses to questions. All responses and utterances should contain a 'substance' but don't necessarily have to be long-winded. For example, when asked *Do you sell your crop in the market?*, you should not respond with a simple, *yes* or *no*. Rather, an acceptable response would be along the lines of *Yes, I sell pomegranates and watermelons in the market every Tuesday*.
 6. SMEs and FLEs are encouraged to be creative in their dialogues, so long as they remain on topic within their specific scenario domains and motivations. Do not stay on one specific topic for too long.
 7. FLEs will not be able to hear any SME speech due to the noise-masking that will be used in their headphones. In the event that the noise-masking is malfunctioning during a dialogue, the FLE should only respond to what they are told by the INTERPRETER.
 8. All speakers will use the Wizard-of-Oz pushbutton/light setup so that they know whose turn it is to speak so that speakers do not interrupt one another.

9. The SME and the FLE should engage one-another in conversation (look at each other) as opposed to looking in the direction of the INTERPRETER. They should be standing across from one another.
 10. SMEs and FLEs should not speak such that their utterances are too long for the INTERPRETER to accurately translate. Additionally, SMEs and FLEs should not have multiple exchanges with the INTERPRETER prior to the INTERPRETER translating the utterance to the other speaker
 - a. INCORRECT FLOW – SME -> INT -> FLE -> **INT -> FLE** -> INT -> SME
 - b. CORRECT FLOW – SME -> INT -> FLE -> INT -> SME -> INT
 11. FLEs should only speak in their native language during the recordings. Do not respond in English or intersperse English words within your native language responses.
- ✓ All QA signs should be shown to all of the speakers so that everyone is knows which rules they correspond. The signs include:
- *START* – SME should initiate the dialogue
 - *WRAP UP* – SME and FLE should start to conclude their dialogue (this does NOT mean stop immediately, but speakers should finish within the next 1 to 2 min)
 - *STOP* – Speakers should stop their dialogue
 - *1st Person* – See #1 above
 - *Complete Translation* – See #2 above
 - *Speak Naturally* – See #3 above
 - *SLOW DOWN* – See #3 above
 - *Avoid “Yes” or “No”* – See #5 above
 - *MOVE ON* – See #6 above
 - *Make Eye Contact* – See #9 above
 - *Don’t Sit Down* – See #9 above
 - *Stand Up* – See #9 above
 - *Shorter Sentences* – See #10 above
 - *NO English* – See #11

APPENDIX F: Example Transcription/Translation Guidelines

TRANSCRIPTION AND TRANSLATION GUIDELINES FOR ENGLISH/IRAQI ARABIC

TRANSCRIPTION FORMAT

A line number and the time stamp for the line is required for each line of transcription. Extra empty lines and line numbers should be prepared so that the translators know where to type their translations. The file sent from transcription to translation should look like:

Line 0001-[interviewerEH](3.211-4.001): good morning sir may I help you %noise
 Line 0001-[interviewerEH2IAH]:
 Line 0002-[interpreterIAH](4.001-4.601): صباح الخير سيدي ممكن أساعدك
 Line 0002-[interpreterIAH2EH]:
 Line 0003-[respondentIAH](4.601-5.233): إي بلي عندي شوية مشاكل بسيارتي
 Line 0003-[respondentIAH2EH]:
 Line 0004-[interpreterEH](5.233-5.902): yes I have some problem with my car
 Line 0004-[interpreterEH2IAH]:
 Line 0005-[interviewerEH](5.902-6.61): what is the problem
 Line 0005-[interviewerEH2IAH]:

The former labels INTV, RESP, and ITPR which may have been used in some data sets to identify the speakerRole have respectively become interviewer, respondent, and interpreter. The speakerRole suffixes EH and IAH indicate the language (English or Iraqi Arabic and that the speaker is human. Correspondingly, ES and IAS would indicate that the speaker is a system.

There should be a beginning and end time-stamp per speaker turn for multi-speaker transcription/translation and per utterance for single-speaker short recordings. All beginning and end timestamps should appear at the start of an utterance within () brackets, separated by a hyphen and followed by a colon and whitespace. There must be at least one digit (tenths of a second) to the right of the decimal point; hundredths of a second is encouraged if meaningful.

Line 0001-[interpreterEH] (3.211-4.001): good morning sir may I help you %noise

Start times of the timestamps must appear in non-decreasing order. The final time-stamp of an utterance should not overlap with the first time-stamp of the next utterance unless there is true overlapping speech. Timestamps can never indicate that a speaker overlaps with himself/herself. The example below would be considered unacceptable unless the noise were marked as overlapping with the “yes”.

Line 0001-[interpreterEH] (3.211-4.001): good morning sir may I help you %noise
 Line 0002-[respondentEH] (4.000-4.902): yes I have some problem with my car

DISCOURSE CONVENTIONS

Turns

The basic transcription unit is the turn - a single act of communication by one individual bounded either by the limits of the transcript or by other turns within the transcript. Each turn begins on a new line of the transcript, preceded by a label to identify the speaker and a time stamp. If any speaker speaks multiple sentences in a very long turn, try to break them into smaller segments at reasonable logical break points. For example:

good morning sir I think you know Corporal Bailey from the other day he suggested we meet and discuss how we might solve your problem

The above text should be broken up like this:

Line 0001-[interviewerEH](3.211-4.001): *good morning sir.*

Line 0002-[interviewerEH](4.001-5.002): *I think you know Corporal Bailey from the other day.*

Line 0003-[interviewerEH](5.002-6.005): *he suggested we meet and discuss how we might solve your problem.*

It is anticipated that anytime the turn includes a mix of statement SU and question SU types, the transcription will separate (at least) those SUs of differing SU_type into separate lines of the transcript, including corresponding timestamps for those SUs. Correspondingly, a line of the transcript is not necessarily an entire turn.

The audio recordings include at least 2 speakers - an interviewer and a respondent. The following codes are required to mark the beginning of each speaker turn.

Speakers

When the **interviewer** speaks *English* mark the beginning with <interviewerEH>.

When the **interviewer** speaks *Arabic* mark the beginning with <interviewerIAH>.

When the **respondent** speaks *English* mark the beginning with <respondentEH>.

When the **respondent** speaks *Arabic* mark the beginning with <respondentIAH>

If the audio includes a **system** interviewer or respondent.

When the **system interviewer** speaks *English* mark with <interviewerES>.

When the **system interviewer** speaks *Arabic* mark with <interviewerIAS>

When the **system respondent** speaks *English* mark with <respondentES>.

When the **system respondent** speaks *Arabic* mark with <respondentIAS>

If the audio includes an **interpreter**:

When the **interpreter** speaks *English* mark the beginning with <interpreterEH>

When the **interpreter** speaks *Arabic* mark the beginning with <interpreterIAH>

Overlap

English Tags: <overlap> and </overlap>

Arabic Tags: <بداية_تداخل> and <نهاية_تداخل>

Mark the sections where multiple speakers speak over each other with <overlap> at the beginning and </overlap> at the end, and transcribe all overlapping speech as accurately as possible. These tags will surround the strings of words that are involved in the overlap.

Line 0014-[interviewerEH] (18.2-19.9): what church <overlap> does </overlap> Ali attend

Line 0015-[respondentEH] (18.9-19.0): <overlap> mosque </overlap>

Silence

English Tag: %silence

Arabic Tag: %انقطاع

Mark long pauses and gaps (> 1 second) within a sentence unit with %silence. It is not necessary to timestamp the silence tag.

Spoken Letters and Acronyms @

All alphas should be transcribed using capital letters. Alphas or alpha strings (not acronyms) should be separated by a whitespace and preceded by an underscore

e.g., we expect _W to address the troops
that's Dana _D _A _N _A

Strings of alphas which form acronyms should be joined by underscores

e. g. C_I_A , F_B_I

Plural spoken letter constructions should be transcribed using s (**no** apostrophe):

e.g., these I_Ds are invalid
the M_R_Es are not popular
more I_E_Ds can be expected
I can find three _Us in Uluru (has **one** spoken letter)

Possessive spoken letter constructions will be transcribed using an s **with** an apostrophe:

e.g., this I_D's expiration data is past
these M_R_Es' contents vary (note position of apostrophe for plural possessive)

Acronyms that are pronounced as words should be transcribed as full words and marked with the @ symbol

e.g., @NASA
@NATO
@ناتو

Arabic, spoken letters are pronounced and transcribed as separate individual words. The Arabic letters for English letters 'j' and 'n' should not be written as ج and ن but as full words جيم and نون.

Letters in non-Arabic borrowed acronyms should be transcribed as separate words but joined with an underscore to align them with English acronym marking.

e.g., أي_بي_سي for 'ABC'

Unintelligible speech (())

If you are unsure of a portion of the transcription, enclose it in double parentheses. For example, if you think a speaker said "looks like this," but are unsure, transcribe it as ((looks like this)). If something is completely unintelligible, transcribe it as (()).

Mispronunciations +

If speakers obviously mispronounce a word such as saying "anenome" instead of "anemone," transcribe the word correctly and mark it with the plus symbol

e.g., +anemone

Fragments -

Partial words or fragments are indicated with a hyphen. If someone says 'thir-thirty', the corresponding transcription would be *thir- thirty*. Please be sure to leave a blank space after the "-" in the above example.

Restarts --

Speaker restarts are indicated with a double dash --. This annotation should be used when a speaker stops short, cutting him/herself off, before continuing with the utterance. Please study the following examples:

brAmj -- brnAmj tEArfy yEny -- برامج
 yEny mA -- mA byh kvyr يعني ما -- ما بيه كثير
 >nA mA d>qdm -- mA d>qdm twjyhy أنا ما دأقدم -- ما دأقدم توجيحي
 w- -- wbEdyn mrAt و- -- وبعدين مرات
 hm yEny h- -- hAy hy هم يعني ه- -- هاي هي
 lkAn >H- -- >Hsn <*A kAn tqdr لكان أح- -- أحسن إذا كان تقدر
 mn E\$- -- mn >kvr mn E\$ryn snp من عش- -- من أكثر من عشرين سنة
 myp w>rbEyn \$hr- -- \$hry? مية وأربعين شهر- -- شهري ؟
 ySyr byh >\$yA' mA k- -- mA knt tqdr tkwn يصير بيه أشياء ما ك- -- كنت تقدر تكون
 lmA Alwld bt- -- bySyr Emrh wAHd wE\$ryn snp لما الولد دت- -- ديصير عمره واحد وعشرين سنة

Interjections and filled pauses (%AH , %UM, etc)

	English	Arabic
Fillers with a vowel sound and no nasal (<i>uh, ah, eh</i>)	%AH	%آه
Fillers with a nasal sound (<i>um, uhm, hm, mm</i>)	%UM	%أم
Positive interjections (<i>uh huh, mm hmm,</i>)	%UH-HUH	%أها
Negative interjections	%UH-UH	%أهوه

Transcribe *yeah*, *yep*, and *nope* if that is what is said. Spell *OK* as *okay*.

Proper Names ^

Proper names are to be labeled with the caret ^ symbol. This symbol is used in front of each word part of the proper name. If the name is comprised of more than one word, the entire name should be enclosed in single parentheses

e.g., (سامية^هندي) (^Samia ^Hindi)
(الرئيس^مبارك) (^President ^Mubarak)

The caret ^ is only used with name entities and name locations - mostly to signal that they could have a variable transcription/transliteration. The names of days, months, and holidays are considered normal nouns and not marked in Arabic.

Note: The previous constraint “Do not use the caret for common nouns that are part of a title or surname” has been removed as it causes confusion for transcribers. All names and titles in a proper name will be marked with a caret.

For English, all proper nouns should be transcribed with initial capital letters as per normal spelling conventions. The caret and single parentheses should also be used to mark names and locations to align with the above conventions. Following are some further examples of correct markup for proper names

eg (^Missus ^Scurr)
(^Toys ^_R ^Us)
(^President ^George ^_W ^Bush)
(^Johanna ^Edwards ^Junior)
(^Sir ^Georg ^Solti)
(^Henry ^Kissinger) P_H_D

Language Switching/Foreign words

English tag %foreign

Arabic tag %أجنبي

Occasionally, a speaker may switch back and forth between two languages (such as English and Arabic) for a short period of time or may use a term from another language. Portions of speech in another language are annotated to indicate the language and wherever possible the words that are spoken in that language transcribed. The language name should appear in all capital letters, and the transcription should be connected with underscores (in place of spaces), exactly as in the following example.

e.g., %FRENCH_merci_beaucoup

If the annotator does not know the name of the language or what is being said, he/she should use the %foreign tag in its place.

ORTHOGRAPHIC CONVENTIONS

- a. Punctuation: Use full stops and question marks along with accompanying timestamps to mark logical break points when transcribing speaker turns. Also retain any punctuation that is part of the standard spelling of a word (for example, *o'clock*, *I'll*, *don't*, and *it's*).

- b. Don't mark the beginning of sentences with a capital letter. Do capitalize the pronoun "I" and proper names, eg. ^Adam, ^Berkeley, and (^New ^York)
- c. Words will be represented in accordance with standard orthographic conventions (as found in a dictionary), unless the speaker's pronunciation of a word is substantially different (e.g. "going to" pronounced "gonna," or "want to" pronounced "wanna.") In these cases, the transcription should imitate the pronunciation. Contractions should be transcribed in the conventional contraction form.
- d. Transcribe everything that was said as words or letters --- don't use numerals. For example, the year 1997 should be spelled out *nineteen ninety seven*, if it is pronounced in the usual way as "19" and "97." The same number could be transcribed as *one thousand nine hundred and ninety seven* or as *nineteen hundred and ninety seven* if this is what the speaker said. Transcription must represent what was actually said. If the speaker pronounces a zero like the letter "O" use the spelling *oh* to represent this pronunciation in numbers, eg. if the speaker pronounces zero as "oh" in a street address like 1901 Center St., it should be transcribed as *nineteen oh one* (^Center ^Street)

NON-SPEECH ACOUSTIC EVENTS

The following tags should be used to transcribe non-speech sounds

Acoustic event	English Tag	Arabic Tag
breath noise and lipsmacks	%breath	%تنفس
laughter	%laugh	%ضحك
cough, throat clear	%cough	%سعال
sneezing	%sneeze	%عطس
clicks and other intermittent noise	%int	%إزعاج_مستقطع
beg. of continuous noise, speech, music	<noise>	<بداية_ضجيج>
end of continuous noise, speech, music	</noise>	<نهاية_ضجيج>

Note: If laughter or coughing extends over time, just mark the starting point. However if it crosses over into a new utterance it will need to be marked again at the start of each new utterance.

NIST notes that background speech that is significantly loud will interfere with ASR and should be separately annotated — but this problem is not likely in the Transtac data.

IRAQI ARABIC TRANSCRIPTION

General spelling Note: Modern Standard Arabic (=MSA)

Adhere as closely as possible to unvocalized MSA spelling and word segmentation in all cases.

e.g. الله not و الله for 'wa Al-l~ah'

When in doubt about the spelling of a word or name, annotators should contact paper or online web references (dictionaries, atlases, etc.)

The following example illustrates the application of these principles: the Iraqi utterance /ʔagul:ič/ ("I tell you" *fem.sg.*) will be transcribed as MSA أَقُلْ لَكَ unvocalized, spelled as two words, using accepted MSA orthography. There are three notable exceptions:

- If the word is listed as a **high-frequency colloquial word** (see below, **HIGH FREQUENCY IRAQI WORDS**) then it should be spelled as indicated in the list and no attempt should be made to render it in MSA.
- If the word is a **colloquial verb** whose morphology deviates substantially from that of its MSA equivalent, then it should be written as indicated in the conjugation paradigms of colloquial verbs. (see below **Perfect Verb Paradigms** and **Imperfect Verb Paradigms**).
- **Nunation** will be transcribed when it is recorded in speech. (see below **Short vowels, diacritics, and Nunation**)

Contractions

Contractions are extremely rare in Arabic. Annotators should limit their use to cases where they are actually produced by the speaker. Annotators must take care to transcribe exactly what the speaker says using standard orthography

e.g., نص for نصف 'half'

Non-standard contractions should be spelled out in full. There are no hyphenated words in Arabic and annotators should consult a dictionary before using a hyphen. Compounds are written as single words, as in قرطوسي or برماني

A good example of a contraction that exists in the Iraqi Arabic form is /has~a/ for /ha+Al+sAEah/. Suggested written form for that word will be هسا instead of the commonly used form هسه can be connected to a current but less frequent variant هسع with the deletion of the final consonant. There is evidence for this form in Tunisian Arabic where the word فيسع 'quickly' is extremely frequent (from في ساعة).

An important annotation decision needs to be made when it comes to choosing whether the word under scrutiny is a 'made-up' word, a misused word, or a non-standard dialect term.

Word compounds and Multiwords

Annotators should use the standard conventions of segmented and non-segmented words in MSA when it comes to the use of affixes and pronouns with nouns or verbs (including participles) and separately, prepositions with whichever words follow them.

e.g., صار لي (2 separate words) not صارلي (one word)
 عبد الله ^ not عبد الله ^

Regional Variant Spellings

Most foreign words and place names already have established MSA spellings (Washington واشنطن, Los Angeles لوس أنجلوس), but in cases where the MSA spelling has regional variants, follow the Iraqi spelling.

e.g., "garage" (Iraqi: كراج -- contrast with Egyptian: جراج)
 "congress" (Iraqi: كونغرس -- contrast with Egyptian: كونجرس)

Words that are not attested in MSA should be transcribed as expected in MSA, but according to Iraqi orthography. Note that although many computers are able to display "extended" Arabic characters, such as the Persian letters /p/ پ, /č/ چ, and /g/ گ, few systems provide the user with an easy way to actually type these characters on the keyboard. So, although these letters are potentially available for representing foreign sounds, the convention in MSA orthography is to substitute the corresponding and easily-available Arabic letters instead. Therefore, according to Iraqi practice,

e.g., for /p/ use ب (e.g., "Pam" بام)
 for /č/ use تش (e.g., "Chet" تشيت)
 for /g/ use غ (e.g., "Gilbert" غلبرت)

Unsure Spelling of Proper Name ^^

In transcribing foreign names, vowels are often conventionally represented in the script using the long glide (vocalic/consonantal) letters و, ي, and the 'alif' ا to represent the non-Arabic vocalic range [i , e , a , o , u]

e.g., ولسون ^ for 'Wilson'
 ماري ^ for 'Mary'

When annotators encounter a name whose spelling they are not sure of, they should make their best guess and transcribe it with a double caret ^ ^ .

e.g., شيراك ^^ for 'Chirak'

Interdialectal variation (DIALECT word)

Sometimes annotators borrow from other Arabic dialects. In this case, the source dialect should be tagged: The following Arabic Dialects have been included so far

ALG	EGP	GLF	IRQ	LEB	JOR	MOR	PAL	SAU	SYR
%جزائري	%مصري	%خليجي	%عراقي	%لبناني	%أردني	%مغربي	%فلسطيني	%سعودي	%سوري

eg. %مصري_كويس

MSA and IA phonological differences

The following regular phonological differences between MSA and IA do not justify departing from MSA orthography when transcribing the colloquial form:

- MSA velar /q/ → IA voiced velar stop /g/.

Examples: قال (MSA /qa:l/, IA /ga:l/)

- MSA velar stop /k/ → IA palato-alveolar affricate /č/.

Examples: كلب (MSA /kalb/, IA /čalib/)

Ta marbuta

Ta marbuta should be written with the two dots (ة) whether pronounced /a/ or /t/

Short vowels, diacritics, and Nunation

Do not transcribe the short vowels /a/, /u/, /i/, or the diacritics indicating zero-vowel ("sukun") and gemination ("shadda"). When Nunation ("tanwin") is recorded in speech it should be transcribed. The most common examples all involve use of "fatHatan" (e.g., /ʔahlan wa-sahlan/ أهلاً وسهلاً), although some uses of "kasratan" are recorded in educated or elevated speech (e.g., /ʔila Had:in ma:/إلى حد ما/). Note that the "fatHatan" may occur without an alif chair: (e.g., /xa:S:atan/ خاصة/)

Glottal stop (hamza)

All glottal stops to be written according to transcription/translation provider's Hamza Rules below.

Description - Initial Hamza	Insert Hamza	
1 st person singular present tense eg. >aktub, >arsim	YES	
Plural nouns starting with alif eg. >aEmaA	YES	
Pronouns eg. <intuwA, <inta,etc	YES	
Digits eg. >abaEap,.....etc	YES	
Comparatives eg. >akbar, >ajmal, >aHlaY	YES	

Perfect tense starting with alif + kasra & its corresponding noun eg. AistifEal, AistafEaAl, AistiEmal, AistaEmaAl		NO
Perfect tense starting with alif + fatha & its corresponding noun eg. >aElan, <iElaAn	YES	
Imperatives starting with initial alif eg. <ijlis	YES	
Prepositions starting with alif eg. <ilaY	YES	
Medial Hamza		
Medial hamza pronounced as 'yaA' when spelt with hamza in MSA eg. qaA}map, maA}lap (Not to be confused with words without hamza in standard grammar eg. SaAyir)	YES	
Medial hamza omitted eg. yiTam~an, muwnap		NO
Medial alif with hamza not pronounced eg. raAs, yaAxu* (If hamza is pronounced, spell it eg. yis>al)		NO
Final Hamza		
Final hamza not pronounced eg. \$aA', masaA', balaA',etc	YES	
Final hamza on chair pronounced as 'yaA' eg. \$aATi}, maTaAfi}, haAdi}...etc. (Except for when it is a proper noun eg. haAdiy)	YES	

Verbal prefixes and suffixes

Colloquial verbs follow MSA paradigms, but with modifications to the underlying tri-consonantal radicals and with the addition of purely colloquial clitics.

The colloquial Perfect Verb follows MSA suffixation orthography except for:

- the MSA 2nd pers. fem. sg. marker /-ti/ is written *تي* in colloquial Arabic (note: this parallels the colloquial orthography of the pronoun /?inti/ *إنتي*).

Example: *شنو سويتني البارحة ؟*

- the MSA 2nd pers. masc. pl. subject marker /-tum/ is shortened to /-tu/ in colloquial and written *توا* (this parallels the orthography of the Imperfect *يكتبوا*).

Example: *وين رحتوا وين إجتوا ؟*

The colloquial Perfect Verb follows MSA stem orthography except for the following:

- MSA doubled verbs are treated as finally-weak verbs in colloquial.

Examples: /HaT:e:t/ حظيت, /Zal:e:t/ ظليت

- MSA verbs with hamza as third radical are treated as finally-weak verbs in colloquial.

Example: /gare:t/ قرئت (MSA /qara?tu/ قرأت) Note: this rule is identical to the earlier-cited rule of MSA glottal stop → /y/ or vocalic length in various environments

- Hollow MSA verbs with hamza as third radical are resyllabified in IA in the conjugation of the perfect.

Example: /?iǧa/ إجا (MSA /ǧa:?a/ جاء), /?iǧu:/ إجا (MSA /ǧa:?u:/ جاؤوا)

PERFECT VERB PARADIGMS

هو	بات	شاف	قرا	إجا
هم	باتوا	شافوا	قروا	إجوا
هي	بانت	شافت	قرت	إجت
إنت	بنت	شفت	قریت	إجیت
إنتي	بتي	شفتي	قریتی	إجیتی
إنتوا	باتوا	شفتوا	قریتوا	إجیتوا
أنا	بنت	شفت	قریت	إجیت
إحنا	بتنا	شفنا	قرینا	إجینا

The colloquial Imperfect Verb follows MSA orthography except for the following:

- the colloquial verb particle /d-/ is prefixed to the imperfect verb subject marker prefixes y-, t-, n-, and A- (the concatenation sequence is as follows: colloquial verb particle /d-/ + MSA subject marker /y-,t-,n-,A-/ + imperfect verb stem).

Examples: دأعرف , دنعرف , ديعرف

IMPERFECT VERB PARADIGMS

هو	ديكتب	ديشوف	ديقرا	ديجي
هم	ديكتبون	ديشوفون	ديقرون	ديجون
هي	دنكتب	دنشوف	دنقرا	دنجي
إنت	دنكتب	دنشوف	دنقرا	دنجي
إنتي	دنكتبين	دنشوفين	دنقرين	دنجين
إنتوا	دنكتبون	دنشوفون	دنقرون	دنجون
أنا	داكتب	دأشوف	دأقرا	دأجي
إحنا	دنكتب	دنشوف	دنقرا	دنجي

Pronominal suffixes and verbal objects (prepositional phrases and direct objects)

MSA orthography should be followed in all cases. Note, especially, the following:

- The 3rd pers. masc. sg. possessive pronoun should be written -هـ, and not -و. For example, /kita:bo/ should be written كتابه, and not كتابو)

- Verbal prepositional phrase objects should be written as in MSA (i.e., not cliticized). For example, /gultil:ak/ should be written قلت لك, and not قلتلك or قلتلكك.

Pronominal suffixes and active participles

Unlike MSA, IA active particles may take pronominal suffixes. Note the following:

e.g., الكتاب بي معلومات قيمة Not الكتاب بيه معلومات قيمة

Numerals

All numerals should be written out as complete words. Numerals from 11 to 19 show a form of word contraction in most Arabic dialects.

إدعش <idE\$ ‘eleven’

إثنعش >vnE\$ ‘twelve’

خمسطعش xumuSTE\$ ‘fifteen’

Except for the numerals 11-19 (see below), follow MSA orthography (or parallel MSA forms) in all cases. Note, especially, the following examples:

Iraqi Arabic	Number
إدعش	11
إثنعش	12
ثلثطعش	13
أرباطعش	14
خمسطعش	15
ستطعش	16
سبباطعش	17
ثمنطعش	18
تسائطعش	19
ثلاثة آلاف	3000
أربعة آلاف	4000
خمسة آلاف	5000
ستة آلاف	6000
سبعة آلاف	7000
ثمانية آلاف	8000
تسعة آلاف	9000
عشرة آلاف	10,000

Days of the week		Months of the year	
الاثنين	Monday	كانون الثاني / يناير / الواحد	January
الثلاثاء	Tuesday	شباط / فبراير / الثاني	February
الأربعاء	Wednesday	آذار / مارس / الثالث	March
الخميس	Thursday	نيسان / أبريل / الرابع	April
الجمعة	Friday	آيار / مايو / الخامس	May
السبت	Saturday	حزيران / مايو / السادس	June
الأحد	Sunday	تموز / يوليو / السابع	July
		آب / أغسطس / الثامن	August
		أيلول / سبتمبر / التاسع	September
		تشرين الأول / أكتوبر / العاشر	October
		تشرين الثاني / نوفمبر / الإدعش	November
		كانون الأول / ديسمبر / الإلتعش	December

IRAQI ARABIC/ENGLISH TRANSLATION: This is based on the document “TransTac Data Steering Committee (DSC), English/Iraqi Arabic, Version 5, April 19, 2005”

Contents:

- 1. Guidelines for Preparing Transcripts for Translation**
- 2. Guidelines for Translators**
- 3. Guidelines for Checking and Distributing Translations**
- 4. Guidelines for Preparing Data for Translators**

Transcripts should be prepared before being sent to translators. [Note: The material in this section has made its way into “Babylon Transcription Conventions, Updated: April 7, 2005”].

1.1 Line Numbers

The audio recordings include at least 2 speakers: an interviewer and a respondent.

Line numbers and time-stamps should be prepared so that the translators know where to type their translations. The file sent to translators should look like this:

Line 0001-[interviewerEH](3.211-4.001): good morning sir may I help you <noise> ?

Line 0001-[interviewerEH2IAH]:

Line 0002-[interpreterIAH](4.001-4.601): صباح الخير يا سيدي أي خدمة ؟

Line 0002-[interpreterIAH2EH]:

Line 0003-[respondentIAH](4.601-5.233): يا صباح النور الله يخليكم اكو مشكلة مع سيارتي .

Line 0003-[respondentIAH2EH]:

Line 0004-[interpreterEH](5.233-5.902): good morning I have some problem with my car .

Line 0004-[interpreterEH2IAH]:

Line 0005-[interviewerEH](5.902-6.61): what is the problem ?

Line 0005-[interviewerEH2IAH]: شنو المشكلة ؟

1.2 Breaking Long Segments

Long segments in transcriptions will be translated into similarly long segments. These long segment pairs can be hard for machines to learn from. Segments longer than roughly 25 words should be broken into smaller segments, at logical break points, before being sent for translation. If you see a transcript fragment like this:

good morning sir I think you know (^Corporal ^Bailey) from the other day he suggested we meet and discuss how we might solve your problem

The above text should be broken up like this:

good morning sir .

I think you know (^Corporal ^Bailey) from the other day .

he suggested we meet and discuss how we might solve your problem .

2. Guidelines for Translators

2.1 What the Source Text Looks Like

The transcripts of recorded conversations will have line numbers and speaker information on each line, so you can tell which speaker said what sentence. For example:

Line 0001-[interviewerEH](3.211-4.001): good morning sir may I help you <noise> ?

Line 0001-[interviewerEH2IAH]:

Line 0003-[respondentIAH](4.601-5.233): صباح الخير يا سيدي أي خدمة .

Line 0003-[respondentIAH2EH]:

Line 0005-[interviewerEH](5.902-6.61): what is the problem ?

Line 0005-[interviewerEH2IAH]:

Note that transcripts include at least 2 speakers: an **interviewer** and a **respondent**.

The transcript may also include segments spoken by an interpreter. The interpreter text will also be translated.

2.2 Where to Type Translations

You should write the translation on the line below the source utterance:

Line 0001-[interviewerEH](3.211-4.001): good morning sir may I help you ?

Line 0001-[interviewerEH2IAH]: **type your translation here**

Line 0002-[interpreterIAH](4.001-4.601): صباح الخير يا سيدي أي خدمة ؟

Line 0002-[interpreterIAH2EH]: **type your translation here**

Line 0003-[respondentIAH](4.601-5.233): يا صباح النور الله يخليكم اكو مشكلة مع سيارتي .

Line 0003-[respondentIAH2EH]: **type your translation here**

If the source segment is English, you should translate into Iraqi Arabic. If the source is Iraqi Arabic, you should translate into English. A single document may contain mixed source material

(some English utterances, some Iraqi Arabic). We have put in tags like “EH2IAH” (“English-to-Iraqi-Arabic”) to remind you of the translation direction.

2.3 Style of Translation

The goal of these translations is to take the source text -- which was originally spoken, not written -- and translate it, producing a result that sounds as if it were originally spoken in the target language. The translation must of course convey the meaning correctly, and should also sound conversational if you read it aloud.

Important note for English-to-Iraqi-Arabic translation: Translate into Iraqi dialect, not Modern Standard Arabic (MSA). Try to resist the urge to use MSA words, and think instead how you would expect to hear the thought expressed in spoken Iraqi.

Try to translate as literally as possible, but without sacrificing fluency or naturalness. For example, there are many ways to translate “You were hit by a bullet.” The preferred translation is one that is closest to “You were hit by a bullet” (rather than “a bullet hit you” or “it was a bullet that hit you”).

Try to maintain the same speaking style (or *register*) as the source. For example, if the source is polite, the translation should maintain the same level of politeness. If the source is rude or angry, the translation should be rude or angry. However, please do not type curse words in your translations, even if they appear in the source.

Because the source text is an unedited transcription of spoken conversations, it may sometimes be hard to read, and may make more sense if you read it aloud. You will see that the source text sometimes reflects the kinds of “mistakes” people say when they're speaking aloud. For example, “Uh no I'm um I think he's uh um his home is over there.” In this case, the speaker pauses (“uh”, “um”) and restarts the sentence three times, changing what he's planning to say (“I'm, I think he's, his home is over there”). Your translations will also have this “spoken-sounding” flavor, somewhat different from what you produce when you translate prose.

2.4 Spelling Target-Language Words

It is very important for you to follow the writing conventions we supply, at all times. Ordinarily, there is a lot of variation in how words are spelled, when a word is written as a single word or broken up with one or more spaces, and so on. Ordinarily, this variation doesn't matter very much because a human being reading the text will be able to figure out what is meant, and will not be too confused about different spellings and where words are broken up.

However, for our purposes, any variability will create a lot more work and cost a lot more, because the text will be read by computers. The computer does not understand that different spellings may mean the same word, and that a “word” may or may not have a space in the middle. It is therefore extremely important to be as consistent as possible in how things are written, and to proofread very carefully.

- We have supplied transcription guidelines on our Iraqi Arabic writing conventions (extracted from “Guidelines for Transcribing Iraqi Arabic”). *You should use these conventions when translating English to Iraqi Arabic.*

- We have also supplied transcription guidelines on English writing conventions under the section headed ORTHOGRAPHIC CONVENTIONS (extracted from “Babylon Transcription Conventions, Updated: April 7, 2005”). *You should use these conventions when translating Iraqi Arabic to English.*
- *While translations do not require transcription tags, alpha strings which comprise an acronym should be joined by an underscore as per the transcription conventions for both Arabic and English. eg. I_B_M*

Please ask us if you have any questions about writing conventions.

2.5 Punctuation and Case

You will notice that the source texts only contain periods and question marks. Your translations should likewise *not include any other punctuation*.

Even though punctuation such as commas and semi-colons would make a sentence easier for a human to read, it usually makes it harder for a computer to read, and these translations will all be processed by computer.

Use normal upper- and lower-case for typing English (e.g., “Bush” for the person, but “bush” for the plant). The only exception for the *first word of the sentence* – treat it like any other word. That is, don’t capitalize it just because it comes first. For example, these are correct: “John came here” and “I came here” (“John” and “I” are always capitalized in your translations) but “you came here” (“you” is never capitalized in your translations).

2.6 Angle Brackets and Percentage signs < > %

Angle brackets < > and percentage signs % are used in the English transcripts to enclose non-speech things, like background noises %noise, the speaker hesitating and saying %AH or %UM, the speaker breathing especially noticeably %breath, and so on. You will also see tags for marking overlapping speech, such as <overlap> and </overlap>. Don’t translate any of these items.

Line 0001-[interviewerEH](3.211-4.001): %breath he just %AH needs to go <overlap> home </overlap> .

Line 0001-[interviewerEH2IAH]: هو بس لازم يرجع للبيت .

2.7 Parentheses (())

Parentheses (()) are used when the person writing the transcript couldn't hear exactly what the speaker was saying. The words in parentheses are the writer's best guess at what the speaker said. Translate these as if there were no parentheses.

Line 0001-[interviewerEH](3.211-4.001): please take this ((medicine)) today .

Line 0001-[interviewerEH2IAH]: خذ هاي الادوية اليوم .

2.8 Broken Words and Restarts

A dash - is used when the English speaker didn't say the whole word. Do not translate broken words.

Line 0001-[interviewerEH](3.211-4.001): he just wa- -- wants to go home

Line 0001-[interviewerEH2IAH]: هو بس يريد يروح للبيت

It also happens fairly often that someone starts to say a word, then changes his or her mind about what s/he wants to say. The speaker may “restart” the sentence.

Line 0001-[interviewerEH](3.211-4.001): he just wa- -- he just needs to go home

Line 0001-[interviewerEH2IAH]: هو بس هو بس لازم يروح للبيت

(Note again that broken words should never be translated).

Line 0001-[interviewerEH](3.211-4.001): I I just know him by Bill

Line 0001-[interviewerEH2IAH]: آ- آني بس أعرفه باسم بيل

This indicates some hesitation, change of mind, or a particular style on the part of the speaker, and the translations should reflect this hesitation or style. If a word or phrase appears twice (like “he just” or “I”, above), try to translate it twice. If that is impossible, try to imitate the hesitation and re-starting in the target language.

Sometimes an utterance may split across multiple lines in English in a way that doesn't make sense syntactically in the source language. Translate these fragments on separate lines; do not combine lines or change line numbers.

Line 0001-[interviewerEH](3.211-4.001): I'm going to look at %AH

Line 0001-[interviewerEH2IAH]: هسا أشوف

Line 0002-[interviewerEH](4.601-5.233): your arm .

Line 0002-[interviewerEH2IAH]: ذراعك .

Line 0003-[interviewerEH](5.902-6.611): okay go ahead .

Line 0003-[interviewerEH2IAH]: طب تفضل .

2.9 Numbers

You will notice in the source texts that all numbers are “sounded out” using words. Please make sure you also “sound out” number translations. Do not type Arabic or Western numerals.

Line 0001-[interviewerEH](3.211-4.001): I was born in nineteen sixty five .

Line 0001-[interviewerEH2IAH]: انولدت في عام ألف وتسعمية و خمسة و ستين .

2.10 Proper Names

Sometimes proper names are hard to translate. If no ready translation springs to mind, then transliterate the name as best you can, and put the original word in brackets [] just after the transliteration.

Line 0001-[interviewerEH](3.211-4.001): I need to talk to my friend (^Corporal ^Bailey) .

Line 0001-[interviewerEH2IAH]: [Bailey] آني لازم أحكي مع صاحبي اسمه العريف بايلي

However, translate well-known person and place names (e.g., George Bush, John, Baghdad, Kabul) as they are written normally in the target language:

Line 0001-[interviewerEH](3.211-4.001): ^John says that (^George ^Bush) is visiting ^Baghdad .
 Line 0001-[interviewerEH2IAH]: يقول جون إنه جورج بوش ديزور بغداد .

Some proper names may be translated partly by sound, and partly by meaning. Strive for the most natural translation:

Line 0001-[interviewerEH](3.211-4.001): I come from (^Southern ^California) .
 Line 0001-[interviewerEH2IAH]: آني جاي من كاليفورنيا الجنوبي .

If you use acronyms that are sounded out letter-by-letter, please be sure to follow the writing convention guidelines, joining the letters with an underscore _

Line 0003-[respondentIA](4.601-5.233): . آني مع الجيش الأمريكي .
 Line 0003-[respondentIAH2EH]: I am with the U_S Army .

2.11 Context

Always consider the context when translating. For instance, the English word “okay” appears numerous times. It could mean several things in Iraqi Arabic. Make it close to authentic Iraqi Arabic as much as possible.

Line 0001-[interviewerEH](3.211-4.001): okay thank you .
 Line 0001-[interviewerEH2IAH]: زين شكرا .
 Line 0002-[interviewerEH](3.211-4.001): okay .
 Line 0002-[interviewerEH2IAH]: تمام .
 Line 0003-[interviewerEH](5.902-6.61): okay so now what’s your -- what’s your child’s name sir ?
 Line 0003-[interviewerEH2IAH]: زين هسا قل لي يا سيدي شاسم بنتك ؟

2.12 Checking

It is very important to proofread and spell check every file.

3. Guidelines for Checking and Distributing Translations

Research and development sites need parallel texts in a form that they can use directly. A final pass over translators’ work will reduce duplicated effort.

3.1 Spell Checking and Proofreading

Translations should be proofread, postedited, and spell-checked by a native speaker of the target language who is familiar with this document and the goals of the TransTac program.

This is necessarily an interactive process. Early in the process, you will find systematic problems in the translations. Please take these problems back to the translators and explain them, so that future translations will require less postediting.

3.2 Text-Only Format

Translations should be distributed in text-only format, not in Word format.

Distributed files should consist of **source line**, **translation line** -- then repeat. The total number of lines in a distributed file should be divisible by two (2).

New translation data should be appended to previous distributions, i.e., any new distribution should contain all program translation data.

3.3 Character Encoding

Both English and Iraqi Arabic text should be distributed in UTF-8 encoding.

HIGH FREQUENCY IRAQI WORDS

Iraqi Standard Arabic	MSA Equivalent
أتاني	أنتظر
أثاري	لهذا
أخلاق سز	عديم الأخلاق
أدب سز	عديم الأخلاق
أرباطعش	أربعة عشر
أشو	إذا
أفا	عفا
أكو	يوجد
ألو	صيغة استهلال الكلام في التلفون
أبياخ	أواه
إثنعش	إثنا عشر
إدعش	إحدى عشر
ستوني	توي
إفتر	تجول
لهم/لها/له/لكم/لك/لنا	لهم/لها/له/لكم/لك/لن
لاخ	آخر
لخ	أخرى
باوع	نظر
بايق	سارق
باكر	غداً
باق	سرق
بزونة	قطعة
بسط	ضرب
بلاية	بلوى

بدأ	بلش
ربما	بلكي
زورق / قارب	بلم
نعم	بلي

أنبوب	بوري
بأي	بيا
جدتي	بيبي
جدة	بيبية
سنة عشر	سنتطعش
مزح	تشاقى
ثلاثة عشر	تلتطعش
رز	تمن
عصاً	توثية
توي	توني
ثمانية عشر	ثمنطعش
زجاج	جام
جالب	جايب
سيجارة	جكارة
سجائر	جكاير
ظننت	حسبالي
ملعقة	خاشوكة
ابتزاز	خاوة
دخل	خش
أنف	خشم
ضيوف	خطار
مسكين	خطية
دع / خلي	خل
ضع / دع	خلي
مزاج	خلق
تسمخ_رشاء	خمسطعش
دع / خلي	خن

إذا	خو
كريم	خواردة
إذا	خوب
إذا	خوباً
جيد	خوش
بعث / أرسل	دز
معضلة / دوار	دوخة
رمى	ذب
هؤلاء	ذولا
هؤلاءك	ذولاك
رجل	رجال
سوف	رح
وبخ	رزل
ليت	ريت
فطور	ريوق
هطل / هطول مطر غزير	زخ
أفخاذ	زرور
ذكر / رجل	زلمة
طيب / حسناً	زين
عادة	سولة
مداس / نعل	شخاطة
علبة كبريت	شخاطة
قبعة	شفقة
مزاح	شقى
ماذا	شكو
كيف	شلون
رمى	شمر
ما هو	شنو
ما هي	شني
كيف	شون
للقدر / بكم	شقد
هندوء / سكون	صنطة

ذنب	صوج
ضجرة	ضوجة
مقلاة	طاوة
ضرب	طخ
فرقة	طرقاعة
كرة	طوبة
بعدها	عادين
صرخ	عاط
ظن	عباله
ظننت	عباري
لأن	عين
على أساس	عأساس
لأجل	علمود
يا ليت	علواه
على أي شيء	علویش
لأي شيء	علیش
على من	عليمن
بعدها	عودين
صارخ	عابط
إذا	عيل
ضفدع	عكرک
بعض	فد
فداء	فدوة
جفل	فز
إمتلاً	قبط
غطاء	قبع
معطف	قبوط
سرير	فريولة
إبري	قط
حذاء	قندرة
علبة	قوطية
إبريق	كتلي

أجرة	كروة
محتال / نصاب	كلاوجي
كل ساعة	كلسع
جداً	كلش
خزانة ملابس	كنتور
للساعة	لسا
نشل	لغف
لماذا	لويش
لا يوجد	ماكو
خاص	مال
خاصة	مالة
خاصتك	مالتك
لا أحد	محد
مبعثر	مخربط

من هنا	منا
من هناك	مناك
مزكوم	منشول
من	منو
لكن / أليس	مو
بوظة	موظة
طاولة	ميز
إنفلونزا	نشلة
نصف	نص
هذا	هاذ
هذا القدر	هالقد
هذه هي	هاهي
ذلك	هاي
تلك	هايه
ذاك	هذاك
هؤلاء	هنول
هذه الساعة	هسا
توي	هستوني

أهلاً	هلو
كذلك / أيضاً	هم
أيضاً	هماتين
أيضاً	همين
كثيراً	هوايا
جمهرة	هوسة
هكذا	هيكى
رغبة	واهس
يرسل	ودي
ويلك	ولك
ويلكم	ولكم
مع	ويا
معي	وياي
أين	وين
ياقة	ياخة
أتى	إجا
يرعب	يخرع
يخطف	يخمت
يتدحرج	يتدهر
يلانم	يرهم
يبري	يقت
يلغي	يكنسل
عند / بجوار	يم
يهرس	يمرد
يمزق	يمعط
بجوارك	يمك
يعطي	ينطي
يرسل	يودي
يفحص	يتشيك
درجة	باية
مرحومة	بنكة
طريقة / وسيلة	جارة

مطرفة	جاكوج
شاي	شاي
سرير	جرباية
شوكة طعام	جطل
بالإضافة إلى	كمالة
أنبوبة	جوب
كيس	كوشر
كيس من الخيش	كونية