Guidelines for Annotation of Arabic Dialectness

Nizar Habash, Owen Rambow, Mona Diab, Reem Kanjawi-Faraj

Habash, Nizar, Owen Rambow, Mona Diab and Reem Faraj. Guidelines for Annotation of Arabic Dialectness.
In Proceedings of workshop on Arabic and its local languages (LREC), Marrakech, Morocco, 2008.

Guidelines for Annotation of Arabic Dialectness

Nizar Habash, Owen Rambow, Mona Diab and Reem Kanjawi-Faraj

Center for Computational Learning Systems
Columbia University
New York, NY, USA
{habash,rambow,diab}@ccls.columbia.edu

Abstract

The Arabic language is a collection of variants with phonological, morphological, lexical, and syntactic differences comparable to those among the Romance languages. However, throughout the Arab world, the standard written language is the same, Modern Standard Arabic (MSA). MSA is based on Classical Arabic and is itself not a native spoken language. MSA is used in some official spoken communication, such as newscasts, parliamentary debates, etc. Other forms of Arabic (generally referred to as "dialects" of MSA) are what people use for daily spoken communication. In this paper, we focus on the issue of creating standard annotation guidelines identifying dialect switching between MSA and at least one dialect in written text. These guidelines can form the basis for the annotation of large collections of data that will be used for training and testing automatic approaches to dialect identification and automatic processing of Arabic which exhibits dialect switching. We report on some initial annotation experiments: we discuss statistical distributions of labels in a small corpus (~19K words) that we annotated according to the guidelines and present inter-annotator agreement results.

1. Introduction

The Arabic language is a collection of variants with phonological, morphological, lexical, and syntactic differences comparable to those among the Romance languages (see for example نخله 1959, Butros, 1973, Omar 1976, السامرائي 1981, Brustad 2000, Bateson 2003, Holes 2004; for computationally oriented summaries of the linguistic situation, see Habash 2006, Diab and Habash 2007). However, throughout the Arab world, the standard written language is the same, Modern Standard Arabic (MSA, فصحى *fuSHaý*),[1] also used in some official spoken communication (newscasts, parliamentary debates, etc.). MSA is based on Classical Arabic and is itself *not* a native spoken language. Other forms of Arabic (generally referred to as "dialects" of Arabic) are what people use for daily spoken communication. In unofficial written communication, in particular in the now growing electronic media, often ad-hoc transcriptions of dialects are used. Arabic dialects are usually divided into four geographic groups: Gulf (including all dialects of the Arabian Peninsula and Iraqi), Levantine (including Syrian, Lebanese, Palestinian), Egyptian (including Libyan and Sudanese) and Maghreb (covering all dialects found West of Libya). Other factors such as the bedouin/sedentary distinction and the distinction between rural (village) and urban communities also define some sub-dialectal variations.

MSA and the dialects thus form a prototypical case of "diglossia" (Ferguson 1959). In a diglossic situation, a "high" language is used in public or prestigious communicative situations (media, government) and is written, while a "low" language is used in private communicative situations (family, daily life) and is usually not written. The two forms co-exist; all speakers master the low form, and most speakers also master the high language to a greater or lesser extent. Arabic conforms to the original definition of diglossia given by Ferguson (1959) in that the two forms are closely related.

In diglossic linguistic situations, one frequently finds cases of code switching, i.e., the use of one or more languages, language variants, or dialects in one discourse, and often within one sentence. We will use the term "dialect switching" in this paper to mean the use of two or more variants of Arabic in one discourse. Dialect switching is well attested in Arabic; see for example (Badawi and Hinds 1986) for a proposal to divide Arabic into five levels, each characterized by the extent of the contributions from MSA (or classical Arabic), from dialectal Arabic, and from foreign languages. Different levels correspond to different sociolinguistic parameters, including education level of the discourse participants, as well as discourse purpose and setting. We adopt the assumption that different types of dialect switching happen in different discourses.

In this paper we focus on the issue of creating standard annotation guidelines identifying dialect switching between MSA and at least one dialect in written text (including transcripts of spoken Arabic). These guidelines can then be used to annotate large collections of data that will be used for training and testing natural language

[1] All Arabic transliterations are provided in the Habash-Soudi-Buckwalter transliteration scheme (Habash et al., 2007). This scheme extends Buckwalter's transliteration scheme (Buckwalter, 2004) to increase its readability while maintaining the 1-to-1 correspondence with Arabic orthography as represented in standard encodings of Arabic, i.e., Unicode, CP-1256, etc. The following are the only differences from Buckwalter's scheme (indicated in parentheses): Ā آ (|), Â أ (>), ŵ ؤ (&), Ă إ (<), ŷ ئ (}), ħ ة (p), θ ث (v), ð ذ (*), š ش ($), Ď ظ (Z), ς ع (E), γ غ (g), ý ى (Y), ã ً (F), ũ ٌ (N), ĩ ٍ (K).

processing (NLP) tools which can identify when dialect switching occurs in a document, and which dialects are involved. This is important, as dialect switching is quite common – as we discuss in Section 3.1, even in edited newswire (which should be all MSA), we find 3.8% of segments to have dialectal influence, and this percentage goes up to 72.3% for broadcast conversations on television (in which participants are usually expected to communicate in MSA as well). The problem is that most existing NLP tools have been developed for pure MSA, for example morphological taggers (Habash and Rambow 2005, Diab et al 2007) or parsers (Bikel 2002, Maamouri et al. 2004a, Maamouri et al. 2006). To have these tools handle Arabic as it is actually produced (namely, with dialect switching), we need to be able not only to handle pure MSA and pure dialectal data,[2] but we also need to develop computational models that can detect and incorporate dialect switching. The proposed annotation guidelines will provide a first step towards creating such computational models, and, generally, towards a broad empirical study of dialect switching in Arabic.

The task of defining annotation guidelines for Arabic dialect switching (or in fact any dialect switching) is complex, because the boundaries between MSA and a dialect are not well defined. If we are studying code switching between, for example, Arabic and English, we can almost always determine from which language a word, a morpheme, even the pronunciation is taken. The only major methodological difficulty is distinguishing borrowings (words that have entered the general vocabulary of language A from language B, such as *algebra* or *oud* in English) from nonce borrowings (words from language B used spontaneously by a speaker while speaking predominantly in language A). In the case of dialect switching, we cannot readily identify borrowings (whether nonce borrowings or regular borrowings) at all. Especially complicating is the fact that Arabic orthography often omits all short vowel diacritics, which may otherwise distinguish between different dialects.

In the rest of this paper we present and exemplify our guidelines (Section 2) and discuss a preliminary study annotating a small corpus with these guidelines and computing inter-annotator agreement (Section 3).

2. Dialect Switching Annotation

When investigating dialect switching (and code switching in general), we can distinguish several (potentially) orthogonal levels of annotation: phonology, morphology, lexicon, and syntax. Since we are interested in written language (including transcribed speech) for the sake of this paper, we omit the phonological level and replace it by orthography. One approach to annotation would be to annotate each of the levels separately. However, in this paper, we group the annotation decisions into two decisions for the word and the segment level (plus one additional annotation decision for the document level).

[2] For example, see (Habash and Rambow 2006) and (Chiang et al. 2006) for NLP tools for the dialects.

The decisions at the word and segment level are judgments on a precisely defined scale, which we refer to as the "dialectness" scale.

Given the difficulty of determining whether a word (or a morpheme) is a borrowing at all in a dialect code switching situation, we have decided to *always* code a word (or morpheme) which can be construed in context as being an MSA word (or morpheme) as MSA. Put differently, our default assumption is that the text is MSA, and we are annotating only clear evidence of dialect influence.

Our approach – two different judgments at the word and segment levels – represents a calculated shortcut, in that it does not provide detailed information at the orthographic, morphological, lexical, and syntactic levels. Instead, based on a preliminary study, we make the assumption that certain combinations are so rare that a simplified annotation scheme is warranted. We acknowledge that a more complex annotation scheme may be required for certain types of studies or computational needs.

2.1 Word-Level Dialect Annotation

Decisions at the word level reflect on orthography, morphology and the lexicon. We annotate the dialectness of each word as a choice among four levels. To determine the correct level, we must first determine the *lexeme* and the *inflectional morphemes* of the word. The two notions are closely related. The lexeme is an abstraction over all possible inflectional morphemes. For example, كتاب *kitAb* 'book', الكتابان *AlkitAbAn* 'the two books', وبكتبكم *wabikutubikum* 'and with your books' all are variants of the same lexeme, which, like all nouns, is usually referred to with the citation form of the singular masculine, كتاب *kitAb* 'book'. Although the word-level annotation task is concerned with the 'words', the context is still relevant to some degree. The annotators need to determine what the meaning of the complete segment is first before proceeding to annotate the different words. This is important since certain words may potentially mean different things in different contexts and deserve different treatments. For example, the word بايع *bAyç* in بايع سيارتك انت؟ *bAyiç say~Artak inta?* 'are you selling your car?' clearly means 'selling' (*bAyiç*) not 'offered allegiance' (*bAyaç*). The first reading is clearly dialectal (from بائع *bAŷiç*) whereas the second reading is MSA. Thus, in this context, this word would be coded as being a dialectal lexeme (since there is no plausible MSA reading).

The following are the four levels for word dialectness annotation:

Word Level 0 is used for *pure MSA* words. These words are standard MSA lexemes with the contextually appropriate MSA inflectional morphology; they have standard MSA orthography with no typographical errors, e.g., يكتبون *yaktubuwn* 'they write', المساجد *AlmasAjid* 'mosques', سيقوم *sayaquwm* 'he will rise', اعيادكم

AçyAdukum ‘your holidays’, بغداد *baγdAd* ‘Baghdad’, سيليكون *siyliykawn* ‘silicon’, and واشنطن *wAšinTuwn* ‘Washington’.

Word Level 1 refers to *MSA with non-standard orthography*. These words are standard MSA lexemes with the contextually appropriate MSA inflectional morphology, but something is strange in the spelling: either a non-standard spelling possibly inspired by non-standard pronunciation, regional variation in spelling, or typographical errors. We do not ask the annotators to choose one of these options as they are often hard to distinguish. The annotators also do not indicate the correct spelling. Examples include مساجذ *masAjið* (instead of مساجد *masAjid* ‘mosques’, presumably a spelling error), فسطان *fusTAn* (instead of فستان *fustAn* ‘dress’, presumably dialectal spelling) and هدا *hadA* (instead of هذا *haðA* ‘this’ dialectal spelling or spelling error).

Word Level 2 indicates an *MSA word with dialect morphology*. These words are standard MSA lexemes but they have at least one morpheme which is clearly a dialectal morpheme (from any of the dialects). A very common example of this is the use of the +ب *b+* prefix in conjunction with an otherwise entirely acceptable MSA verb, for example بيذهب *biyaðhab* ‘he goes’. Note that the +ب *b+* prefix can appear with nouns in MSA but not with verbs. A common spelling of the Levantine/Egyptian verb ‘I write’ is بكتب *baktub*; although this word looks like the MSA *bikutub* ‘in books’, the syntactic and semantic contexts are used to disambiguate. Examples of other dialectal morphemes often found in conjunction with MSA words include +ح *Ha+* (Egyptian/Levantine ‘future tense’), +ع *ça+* (Levantine preposition ‘on/to’), +د *da+* (Iraqi ‘present tense’), +ك *ka+* (Moroccan ‘present tense’), +غ *γa+* (Moroccan ‘future tense’), م++ش *ma++š* (Egyptian/ Levantine negation circumfix) and +ش *š+* (Iraqi question marker).

Word Level 3 indicates a *dialect lexeme.* These words are words which clearly would never be used in written or spoken MSA by an educated speaker/writer. This level does not include orthographic variants of MSA words, these are coded as Word Level 1. Prime examples are the negation marker مش *miš* ‘no/not’ and its variants and other dialectal morphemes that can be spelled separately from the word: ح *Ha* (Egyptian/Levantine ‘future tense’), ع *ça* (Levantine preposition ‘on/to’), and عم *çam* (Levantine verbal particle marking progressiveness), etc. Other examples of purely dialectal lexemes include: بزونة *baz~unaħ* (Iraqi for ‘cat’). Dialectal lexemes that have MSA homophones are also marked if the context shows that the word is clearly dialectal, for example عافية *çAfyaħ* (Moroccan for ‘fire’ but MSA/Levantine for ‘health’). Note that at this level, orthography is irrelevant, as there is no standard orthography for the dialects anyway. We also do not consider morphology: a dialect word with MSA-only morphology is also coded as Word Level 3.

2.2 Segment-Level Dialect Annotation

Beyond word annotation, the annotators are asked to judge the whole segment (sentence/utterance) all at once in terms of the quality of the MSA. This annotation is necessary to address cases that at the word level seem all MSA, but it is clear that the sentence is dialectal or mixed because of lexical issues involving multi-word lexemes, or because of syntax. The judgment is on a scale from 0 to 4.

Segment Level 0 is defined to be *perfect MSA***.**

Segment Level 1 is *imperfect MSA*. Here, the source is trying to produce MSA, but some dialectal phenomena are sneaking in (dialectally inspired orthography revealing pronunciation, some dialectal morphology, incorrect case or mood, incorrect subject-verb agreement, or perhaps an isolated dialectal lexeme). A segment cannot be Segment Level 1 if there are more than minimal number words of Word Level 3 – in practice, we have yet to define this threshold.

Segment Level 2 is *Arabic with full dialect switching***.** It is not clear whether the writer is aiming for writing in MSA, or in dialect. A segment cannot be Segment Level 2 if all words are Word Level 2 or Word Level 3.

Segment Level 3 refers to *dialect with MSA incursions*. The source is producing dialectal Arabic, but uses some clichés or words clearly borrowed from MSA. A segment cannot be Segment Level 3 if all words are Word Level 3.

Segment Level 4 is used to mark *pure Dialect.* Here, the writer is producing pure dialectal Arabic. A segment of Segment Level 4 has, in general, at least one word of Word Level 2 or Word Level 3.

2.3 Source-Level Annotation: Home Dialect of Speaker/Writer

After the previous two annotations are done, the annotator may have a reasonably good idea of the home dialect of the speaker/writer. This guess will be marked. The annotators are encouraged to use any knowledge they can muster, e.g., the language used, the words used, the topic, the home country of the newsgroup or TV station, etc. We specify a hierarchy of specific names for dialect regions (such as Maghreb>Tunisian and Levantine>Palestinian) and sub-dialectal features (such Urban or Bedouin). This allows a degree of reasonable approximation in case of doubt.

3. Annotation Experiments

In this section, we present some preliminary results on an annotation task using our guidelines. We first describe the annotation and detail statistics on the distribution of labels across different genres. Then we present inter-annotator agreement results on a portion of the annotated data.

3.1 Corpus Annotation

We annotated a small corpus of 59 documents (19,160 words) in four genres: newswire (NW), web text (WT), broadcast news (BN) and broadcast conversation (BC). The annotation was done as part of MT error analysis research (Kirchhoff et al. 2007), and our corpus choice was dictated by this task. NW and WT were naturally occurring text as opposed to BN and BC which have been transcribed by the Linguistic Data Consortium (LDC).[3] NW data came from Agence France-Presse, Xinhua News and Assabah. WT data came from different Google and Yahoo groups, such as IslamToday or YaMuslim. BC data came from different shows on AlJazeera and LBC. BN data came from AlJazeera, LBC and Dubai TV. Our annotator was a female Levantine Arabic speaker. The data overall was rather free of some of the dialectal phenomena that influenced many of our decisions (particularly dialectal orthographic inconsistencies): the transcripts of the BN and BC data were carefully produced according to LDC guidelines (Maamouri et al. 2004b); and the religious theme of the WT data made it more MSA-like than some other web texts we have seen which are much more dialectal. Thus, this set of texts is not truly representative of the types of dialect switching we expect to find.

Table 1 presents various statistics over the annotated set. The first four columns of data belong to the four genres annotated. The next two combine the two transcribed spoken genres (BN and BC) into BX, and the two written genres (WT and NW) into TX. The last column combines all the data. The first data row shows the number of documents. The second shows the number of segments. BX data had more segments (lines) that were shorter (average 8 words) than TX data (average 19). The next five rows show the distribution of segment level labels. Overall BX data has a larger number of segments in level 1 than level 0. This stands in stark contrast to the TX data whose segments are almost all in level 0. Within BX, BC exhibits a higher degree of level 1 than BN. This is a consistent trend with expectations about the genres: BC is less rehearsed and is more reactive (and thus more dialectal) as opposed to BN which is primarily read. Within TX data, we do not see the trend we expect; namely that WT data is more dialectal. This is perhaps a result of the data being of a higher MSA quality than is commonly the case in other news groups because the groups' themes are religious.

Next in Table 1 is a row showing the number of words per genre. This is followed by the distribution of the four word levels. The distributions here are consistent with what we expect: BC is more dialectal than BN than WT than NW. The jump in NW at level 3 is the result of the Levantine annotator considering the spelling of some month names as dialectal because they are different from her MSA: e.g., the name used for the month 'February' was فبراير *fibrAyir*, which is acceptable in Egypt as MSA but not in the Levant (corresponding MSA month name is شباط *šbAT*) (Omar 1976). We will revisit this issue when we discuss inter-annotator agreement in Section 3.2. Overall, at the word annotation, BX has less of level 0 and more of higher levels than TX.

	BC	BN	WT	NW	BX	TX	ALL
Doc's	6	8	17	28	14	45	59
Seg's	640	437	280	287	1077	567	1644
Level 0	27.7	57.4	98.9	96.2	39.7	97.5	59.7
Level 1	67.5	41.2	0.7	2.4	56.8	1.6	37.8
Level 2	3.1	0.5	0.4	1.4	2.0	0.9	1.6
Level 3	0.5	0.2	0.0	0.0	0.4	0.0	0.2
Level 4	1.3	0.7	0.0	0.0	1.0	0.0	0.7
Words	4619	3919	5042	5580	8538	10622	19160
Level 0	96.0	97.9	98.7	99.3	96.9	99.0	98.1
Level 1	2.2	1.5	1.2	0.2	1.9	0.7	1.2
Level 2	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Level 3	1.7	0.6	0.1	0.5	1.2	0.3	0.7

Table 1 Statistics of annotated corpus (levels areprecentages). BC is transcribed broadcast conversation, BN is transcribed broadcast news, WT is web texts (web discussion forums), and NW is newswire. BX is the combination of the two transcribed speech genres (BC and BN), while TX is a combination of the two text genres (WT and NW).

3.2 Inter-annotator Agreement

We double annotated around 20% of the data (11 documents) roughly equally distributed over all genres. Our second annotator is a female Egyptian Arabic speaker. We consider here word-level and segment-level annotations only. At the word level, the overall agreement (as accuracy) is over 98.5%. The most common label agreed on is *level 0* (97.1%). The corresponding Cohen's (1960) kappa score is 0.72 indicating a good degree of agreement. The largest disagreement is between levels 0 and 1. A large portion of the disagreement is the result of our guidelines' lack of specification of MSA standard spelling of some words, particularly proper names; although MSA is standard across the Arab world, some regional varieties of spellings exist. Our Egyptian and Levantine annotators disagreed on whether كوبنهاجن *kuwbinhAjin* 'Copenhagen' is level 0 or 1: it is level 0 in Egyptian MSA, but the preferred spelling in Levantine MSA is كوبنهاغن *kuwbinhAγin*. Similarly, the spelling of names of continents such as 'Africa' using a Taa Marbuta (افريقية *Afriyqyaħ*) was not accepted by the Egyptian annotator who preferred (افريقيا *AfriyqyA*). There is a small portion of levels 0 and 3 disagreement too. These were primarily the result of interpreting interjections literally (as MSA) or functionally (as dialect). For example, the use of والله *waAll~ah* '(lit. by God)' as a semantically empty interjection led one of our annotators to mark it as lexically dialectal (level 3). The segment-level annotation inter-annotator agreement is lower than word-level annotation: basic accuracy agreement is 78% and the kappa measure is 0.56.

[3] http://www.ldc.upenn.edu/

4. Conclusions and Future Work

We have presented a proposal for annotation guidelines identifying dialect switching between MSA and at least one dialect in written text. We have reported on some initial annotation experiments showing that dialect annotation has distinct distributions in different genres. Our initial results on inter-annotator agreement are encouraging. However, much more work is needed in clarifying and detailing the guidelines. In particular, the results of the inter-annotator agreement analysis suggest a need to address the existing variations of MSA in different regions in the guidelines to specify a reference point and/or make the annotators aware of these variations: we cannot have an annotator flag something as dialectal though it is perfectly acceptable MSA in another part of the Arabic-speaking world. Once the guidelines have been updated, we plan to annotate additional data of different variety in the future.

5. Acknowledgements

This work was partially funded by DARPA GALE contract HR0011-06-C-0023 and by NSF SGER grant BCS-0749062. We would like to thank the participants in the April 2008 workshop on the annotation of code switching for their useful feedback.

6. References

السامرائي, ابراهيم. (1981). العربية التونسية (فصل 13). التطور اللغوي التاريخي. دار الاندلس. بيروت.

نخله, رفائيل. (1959). غرائب اللهجه اللبنانية السورية. المطبعة الكاثوليكية. بيروت.

Badawi, S, and Hinds, M. (1986). *A Dictionary of Egyptian Arabic: Introduction*. Beirut: Librairie du Liban.

Bateson, Mary Catherine. (2003) Arabic Language Handbook. Georgetown University Press.

Bikel, Daniel. (2002). Design of a multi-lingual, parallel processing statistical parsing engine. In Proc. of International Conference on Human Language Technology.

Brustad, Kristen E. (2000). The Syntax of Spoken Arabic: A Comparative Study of Moroccan, Egyptian, Syrian, and Kuwaiti Dialects. Georgetown University Press.

Buckwalter, Tim. (2004). Buckwalter Arabic morphological analyzer version 2.0. 12(3):373–418.

Butros, Albert. (1973). Turkish, Italian, and French Loanwords in the Colloquial Arabic of Palestine and Jordan. Studies in Linguistics. Volume 23.

Chiang, David, Mona Diab, Nizar Habash, Owen Rambow and Safi Sharif. (2006). Parsing Arabic Dialects. In Proc. of the European Chapter of the Association for Computational Linguistics (ACL). Trento, Italy.

Cohen, Jacob. 1960. *A coefficient of agreement for nominal scales*. Educational and Psychological Measurement, 20(1):37-46.

Diab Mona and Nizar Habash. (2007). Arabic dialect processing tutorial. The conference of the North American Chapter of ACL, Rochester, NY.

Diab, Mona, Kadri Hacioglu, and Daniel Jurafsky. (2007) "Automated Methods for Processing Arabic Text: From Tokenization to Base Phrase Chunking". Book Chapter. In Arabic Computational Morphology: Knowledge-based and Empirical Methods. Editors Antal van den Bosch and Abdelhadi Soudi. Kluwer/Springer Publications.

Ferguson, Charles A. (1959). "Diglossia". *Word 15,* pp. 325--340.

Habash, Nizar, Abdelhadi Soudi and Tim Buckwalter. (2007). "On Arabic Transliteration." Book Chapter. In Arabic Computational Morphology: Knowledge-based and Empirical Methods. Editors A. van den Bosch and A. Soudi. Kluwer/Springer Publications.

Habash, Nizar and Owen Rambow. (2005). Arabic Tokenization, Morphological Analysis, and Part-of-Speech Tagging in One Fell Swoop. In Proc. of ACL. Ann Arbor, MI.

Habash Nizar and Owen Rambow: (2006). MAGEAD: a morphological analyzer and generator for the Arabic dialects. In Proc. of ACL. Sydney.

Habash, Nizar. (2006). "On Arabic and its Dialects," Multilingual Magazine. #81 Volume 17 Issue 5.

Holes, Clive. (2004). Modern Arabic: Structures, Functions, and Varieties. Georgetown University Press.

Kirchhoff, Katrin, Owen Rambow, Nizar Habash, and Mona Diab. (2007). Semi-automatic error analysis for large-scale statistical machine translation. In Proc. of MT Summit XI, Copenhagen, Denmark.

Maamouri, Mohamed, Ann Bies, and Tim Buckwalter. (2004b). The Penn Arabic Treebank: Building a largescale annotated Arabic corpus. In NEMLAR Conference on Arabic Language Resources and Tools, Cairo, Egypt.

Maamouri, Mohamed, Tim Buckwalter, and Christopher Cieri. (2004b). Dialectal Arabic Telephone Speech Corpus: Principles, Tool Design, and Transcription Conventions. In NEMLAR.

Maamouri, Mohamed, Ann Bies, Tim Buckwalter, Mona Diab, Nizar Habash, Owen Rambow, and Dalila Tabessi. (2006). Developing and using a pilot dialectal Arabic tree-bank. In Proc. of LREC'06.

Omar, Margaret. (1976). Levantine and Egyptian Arabic: Comparative Study. Foreign Service Institute. Basic Course Series.