Rosetta Stone Version 3 Falls Short of Manufacturer's Claims

Katharine Nielson, MA

Suzanne Freynik, MA

Executive Summary

PURPOSE

CASL review and research in 2007 revealed shortcomings with Rosetta Stone® Version 2 (V2) software. Rosetta Stone recently released Version 3 (V3), claiming significant improvements over V2 of its product. To determine whether our V2 review is still accurate and whether V3 may have affected the outcomes of our previous empirical study, we compared the two versions.

CONCLUSIONS

1. As in V2, the images in V3 are not culturally relevant. In fact, 90 percent of the photos in the Spanish and Arabic programs are identical.

2. As in V2, information about how a language's writing system works is not provided in V3. Learners are shown examples of the language's script and are expected to figure out how the writing system works on their own.

3. Unlike V2, V3 includes exercises that might help learners with basic greetings and other real-life conversations. The new Milestone feature attempts to mimic conversational practice through multiple-choice drills. However, these drills fall short of the "real-life simulations" that the website claims have been included in V3.

4. Although the V3 lessons are organized in more or less the same way as the V2 lessons, V3 includes grammatical concepts that V2 did not include. The new concepts include features of Arabic or Spanish that would be especially difficult for second-language learners.

5. Although some features of the V3 software provide more dynamic elements, the program is not truly interactive. For example, V3 tells learners when they should try an exercise again but does not give them a chance to participate in conversations.

6. While some features of Rosetta Stone V3 have changed, these changes would not have affected the outcomes of our empirical study.

PURPOSE—To determine whether CASL's review of Rosetta Stone® Version 2 is still accurate and whether Version 3 may have affected the outcomes of our previous empirical study.

CONCLUSIONS—Some features of Rosetta Stone V3 have changed; however, these changes would not have affected the outcomes of our empirical study.

RELEVANCE—As a stand-alone package, Rosetta Stone is unlikely to be the solution to the U.S. Government's language learning needs.

TTO 2101 E.4.1 | CDRL A017 | DID DI-MISC 80508B | Contract No. H98230-07-D-0175

Learners in our study might have had more familiarity with conversational vocabulary, but they would not have had more conversational practice. Further, those who disliked the lack of explicit writing instruction or the tedium of the drills would not have had a substantively different experience with V3.

RELEVANCE

Rosetta Stone's claims about the innovativeness of the product, as well as the language learning outcomes possible after its use, are generally overstated. The software does not provide the dynamic environment required to practice using the language in context. Rosetta Stone might be a useful tool to supplement vocabulary acquisition in a more well-rounded language course, but as a stand-alone package Rosetta Stone is unlikely to be the solution to the U.S. Government's language learning needs. ■

Executive Report

PURPOSE

To help meet its language learning goals, the U.S. Government has invested millions of dollars in the online foreign-language training program Rosetta Stone®. In 2007 the University of Maryland Center for Advanced Study of Language (CASL) empirically examined the effectiveness of Rosetta Stone Version 2 (V2) in Arabic, Chinese, and Spanish. Results of that study and a previous CASL evaluation describing the program's potential effectiveness[1] revealed shortcomings with the software, both with respect to the delivery of claims made by the company and in the usefulness of Rosetta Stone for self-study by beginning language learners. For example, the software did not provide resources for learners to practice using the language in a communicative context, and the courses lacked any consideration of culture.

After CASL published its review of V2 and while the empirical study was underway, Rosetta Stone released Version 3 (V3) in several different languages,[2] including Spanish and Arabic. The company has claimed that "version 3 significantly improves our product," and should compensate for many of the problems identified by the CASL review and empirical study.

What this study investigated

Because V3 was released after our review and during our empirical study of V2, our current evaluation[3] aimed to determine (1) whether our expert review of the software is still accurate and (2) whether V3 has any potential implications for the results of our empirical study.[4]

In this report we first revisit the issues with V2 raised by CASL researchers, identify the new V3 claims made by the software manufacturers, and review V3 in both Spanish and Arabic. We then consider whether using V3 during our empirical study might have changed the results.

CONCLUSIONS

Our review of Rosetta Stone V3 reveals that while some features have been updated and improved, many have remained the same. In this section, we revisit a number of problems encountered with Rosetta Stone V2, present the claim made by Rosetta Stone addressing each problem (if available), and review the content of the new course with respect to the problem identified.

1 **As in V2, the images in V3 are not culturally relevant.**

Problem with V2[5]: The images used throughout the course are the same regardless of the language being taught, so they are not culturally relevant.

Claim by Rosetta Stone about V3[6]:

> ❝Our new global tapestry of people images reflects the diversity of people and cultures from which our language products derive. We hope you'll find these fascinating.❞

Review: While the images have been updated for V3, 90 percent of the photos are the same for both Spanish and Arabic.[7] The photos are not language-specific and are not culturally relevant. In other words, they do not immerse language learners in the target culture because they depict images of people in all cultures.

2 **As in V2, information about how a language's writing system works is not provided in V3.**

Problem with V2: Rosetta Stone V2 does not explain how the writing system works. Learners are expected to figure out the writing system of each language on their own. The learner-log comments from the CASL 2007 empirical study indicated that the participants were particularly frustrated with the non-Roman scripts, which caused users to either seek outside resources in order to learn the writing system or abandon the program altogether.

Claim by Rosetta Stone about V3:

> ❝Version 3 uses Contextual Formation™ to ask the learner to write new language in response to conversational prompts."[8] In addition, "Version 3 immerses you immediately in the new language, providing just the right context to prompt you to speak, pronounce, read and write in the very first lesson! Intuitive, sequential learning builds progressively and makes every lesson count.❞

Review: The prompts for the writing drills in the newest version of Rosetta Stone are more conversational than the prompts used in V2. That is, learners are sometimes asked to respond in writing to conversational exchanges. However, V3 continues to take an inductive approach to teaching the written language. It does not explicitly explain how the writing system for a language works, and it assumes that learners will be able to map the written language onto the sounds they are hearing (while they are simultaneously learning to map those sounds to the meaning pictured).

For example, the very first writing lesson in Lesson 1 shows four different images with the noun describing each image written in the target language. Students are asked to match the writing to the image. The next screen shows a letter or syllable from one of the previously displayed words written out, and students are expected to match it to the appropriate sound. The software provides no explanation with reading or writing activities, and learners are expected to learn these skills through trial and error.

3 **Unlike V2, V3 includes exercises that might help learners with basic greetings and other real-life conversation.**

Problem with V2: At the end of the entire course, learners have not acquired fundamental material, such as basic greetings or how to introduce themselves. In addition, there is no dialogue or interaction, but instead

short, unrelated sentences that have no real-life counterpart. Finally, the single participant in our empirical study who completed the 200-hour V2 course had the following comment: "While Rosetta Stone does teach a lot of words, they are not always the words you need to have an actual conversation."

Claim by Rosetta Stone about V3:

" New Contextual Formation™ feature uses real-world simulations to give you the benefits you need to succeed. New Milestone feature lets you try out your new language knowledge in real-life simulations. "

Review: Rosetta Stone V3 appears to be more relevant in terms of communicative language use. Nearly all of the grammar drills in V2 encouraged learners to use the third person. V3, on the other hand, includes a variety of grammar drills that encourage learners to use the first person, which is more likely to be needed in conversation.

In addition, while V2 overlooked very basic items like greetings, V3 includes these from the start. However, the greetings are not necessarily presented in a communicative context. A number of drills begin with a greeting, go on to address unrelated grammar and vocabulary, and then end with an equally unrelated farewell. For example, the Core lesson in Unit 1 begins with two images: a man and a woman. The word *hello* is written above each picture and is heard as each image is highlighted. Learners do nothing on this screen, but they do see the words and hear them. Then the actual lesson begins and learners match words (which have nothing to do with greetings) to images and complete simple grammar drills. The last screen of the Core lesson is of two different people leaving. The word *goodbye* is written above each picture and learners hear it as the images are highlighted. The *hello* and *goodbye* screens are unrelated to the material presented in the rest of the lesson. While presenting greetings and salutations in this fash-

ion does expose learners to the words, it does not show them how they are used in an actual conversation. Rather, the words simply mark the beginning and the end of the lesson.

Some of the picture-concept mapping exercises present images that are related to one another in the form of a conversational exchange (e.g., "My house is big." "My house is bigger than your house."). Again, this exchange is more communicative in nature than the examples used in the exercises of V2, but it hardly provides the "immersion" environment claimed by the manufacturers.

The most significant gesture V3 makes toward communicative language use is the Milestone feature that ends each of the four units, where the learner participates in a dialogue with pictured interlocutors by choosing the written utterance that fits each scenario. The interlocutors are presented in the style of a cartoon strip, with photos scrolling across the scene and speech bubbles indicating who is speaking to whom. The dialogue presented in this feature is identical in every language, and while conversation-like, the exchanges are stilted and inauthentic, as in the following example:

> Hello.
> Hello.
> Are all the flowers the same color?
> No, the flowers are not all the same color. What color flowers do you want?
> I would like red flowers.
> How many would you like?
> I would like 35. Thank you.
> You're welcome.
> Good morning.
> Good morning.
> Do you sell eggs?
> Yes, I sell eggs. How many eggs do you need?
> I need 48 eggs.
> Thank you.
> You're welcome.

The above dialogue is unrealistic and, more importantly, does not give learners enough time for this type of communicative drill. Each of the four

milestones is estimated to take 10 minutes. Therefore, of the 200 or so hours necessary for this course, a total of *40 minutes* is designed for practice in communication. These drills are not the "real-life simulations" that the website says have been included in V3.

4 Although the V3 lessons are organized in more or less the same way as the V2 lessons, V3 includes grammatical concepts that V2 did not include.

Problem with V2: The grammar lessons appear to be sequenced by the notion of moving from simple to complex grammar. This simple to complex sequencing is known as a covert grammatical syllabus. It is commonly found in mass-market language materials, despite publishers' claims to be adopting a more innovative approach. In addition, V2 provides no information on cultural-grammatical issues such as the difference between formal and familiar verb forms in Spanish or the difference between dual and plural noun marking in Arabic.

Claim by Rosetta Stone about V3:

" Improved intuitive, sequential learning makes every lesson count and build progressively. "

Review: Lessons in V3 do not appear to be organized differently than in V2. The organization of the materials is still based on the shift from simple to more complex grammar, though the focus in V3 is on different grammatical concepts. V3 does incorporate some of the important grammatical concepts that the second version overlooked.

Specifically, the drills in V3 provide more practice on some points of Arabic or Spanish grammar that might be difficult for speakers of other languages. For example, V3 Arabic includes additional drills that illustrate the difference between dual and plural marking on nouns, case-marking on nouns, and the conjugations for verbs according to gender as well as

number, all of which were introduced only sporadically in V2. V3 Spanish illustrates the difference between formal and informal second person and no longer focuses extensively on the infrequently used present progressive tense. In both Spanish and Arabic, V3 introduces first, second, and third person agreement all within the first lesson, which is an improvement over V2, where first person agreement is not introduced until Lesson 5 and second person agreement is ignored until the last lesson.

In order to accommodate this more diverse array of grammatical forms, V3 dispenses with some of the unusual vocabulary that permeated V2; V3 does not have as many drills about ducks and horses and people crouching under tables. Rather, V3 introduces a sparser (and arguably more useful) core vocabulary and then drills the same items repeatedly with grammatical manipulations. For instance, a set of drills in V2 might depict people jumping off of many different things, while a comparable set of drills in V3 depicts different combinations of people writing. The V2 manipulation illustrates the names of all the things people can jump off of, whereas the V3 manipulation illustrates how the verb *write* agrees with different subjects.

Unfortunately, in addition to paring down vocabulary items, V3 also does away with some necessary grammatical concepts. While V2 had lamentably few exercises involving the past and future tenses, V3 has even fewer, adhering almost exclusively to the present tense. In addition, the attempts to include language-specific grammatical information are inconsistent. While it is standard conversational practice in both Spanish and Arabic to drop pronouns before conjugated verbs, only the Spanish version shows examples of this throughout the drills. The Arabic version of the software remains a translation of the English version and retains pronouns in every example, which does not reflect how the language is actually used.

5 Although some features of the V3 software provide more dynamic elements, the program is not truly interactive.

Problem with V2: The Rosetta Stone Dynamic Immersion Method[9] environment is quite impoverished in comparison with the rich immersion environments of naturalistic language acquisition. Rosetta Stone V2 is restricted to a limited vocabulary surrounding the set of images. The only dynamic and interactive elements are the distinctive tones and symbols intended to indicate that a selection among four choices is correct or incorrect and the comparison of a visual representation of the learner's pronunciation to a native speaker's graph.

Claim by Rosetta Stone about V3: Rosetta Stone does not claim to have changed their Dynamic Immersion Method environment; however, the company describes a new feature that attempts to provide another form of customization.

> New Adaptive Recall™ Language feature tracks progress to reinforce your strengths and revisit needs. New proprietary speech recognition technology gets you speaking from the start and new speech analysis tools perfect your pronunciation.

Review: In terms of the dynamic and interactive components, the course continues to provide feedback via a negative tone. In other words, learners are not given any hints on how to select the correct answer, even after choosing the incorrect answer repeatedly. Rather than receive the correct answer, learners receive feedback about when they should attempt the lesson again. The V3 Adaptive Recall feature records user performance, determines how many items the learner answered correctly, and then makes suggestions for future study. When accessing the program on the suggested date, the learner is prompted to complete the old lesson before continuing with new work.

Similarly, although the technology behind the speech tools appears to be improved, learners still do not receive any feedback other than the chance to compare the graph of their utterances to that of a native speaker. Unlike in V2, learners have more time to practice speech in V3. The lessons that rely on the speech recognition software allow learners to link to a separate screen where they can listen and record each statement multiple times before continuing to the next question. They are able to spend more time working on pronunciation, but they have no chance to actually interact in the target language.

6 While some features of Rosetta Stone V3 have changed, these changes would not have affected the outcomes of our empirical study.

The updated "global tapestry of images," while more visually appealing, does not change the format of the program and would have had no effect on our empirical study. And while the addition of the Contextual Formation and Milestone features is a gesture toward more communicative language, the features do not provide practice for authentic conversations with native speakers. Therefore, while learners in our study might have had more familiarity with conversational vocabulary, they would not have had any more practice with completing conversations. It doubtful that the single participant of the study who completed the Oral Proficiency Interview would have been better equipped to converse in Arabic.

The inductive method of instruction has remained the same, so those students who disliked learning through guessing would not have had a different experience with the new product. In addition, the technology required to run the program is the same, so using a new version would probably not have changed any of the technical issues, such as learners being unable to use the software from secure facilities or while using dial-up Internet access.

While the overall look and feel of the course is more modern, it remains very much the same program. Learners still do little more than match pictures to words and sounds. The language is somewhat more communicative, but the program does not provide the opportunity for communicative practice. The "simulations" involve selecting words from multiple-choice drop-down menus, so they do not prepare learners to actually speak in the target language. Finally, the one-size-fits-all-languages approach has not changed, and while there have been some attempts to adapt the courses to fit language-specific needs, these have been haphazard. Since the structure of the course remains essentially the same, those who disliked the lack of explicit writing instruction or the tedium of the drills would not have had a substantively different experience with V3.

RELEVANCE

Our review of Rosetta Stone V3 reveals that while some problems with V2 have been addressed, there is still more room for improvement. Further, it is unlikely that using V3 would have had much of an effect (if any) on the outcomes of our previous empirical study, which examined the effects of V2 on language learning.

Therefore, our conclusions and recommendations remain the same. The claims made by the Rosetta Stone manufacturers concerning the innovativeness of their product as well as the language learning outcomes possible after use are generally overstated. While it is possible that learners using this product might learn some conversational phrases, the software does not provide the dynamic environment required to practice using the language in context. Rosetta Stone might be a useful tool to supplement vocabulary acquisition in a more well-rounded language course, but as a stand-alone package it is unlikely to be the solution to the U.S. Government's language learning needs. ■

ENDNOTES

1 In July 2007 CASL prepared a technical report reviewing several technology-mediated language training (TMLT) products, which included evaluations of Rosetta Stone Version 2 (V2) in Spanish and Arabic. In May 2007 CASL began conducting an empirical study testing the effectiveness of Rosetta Stone V2 in Arabic, Chinese, and Spanish, and a report with the findings from this study was prepared in March 2008.

2 According to the Rosetta Stone website, the following languages were initially launched in Version 3: Arabic, English (U.S.), English (U.K.), French, German, Italian, Portuguese (Brazil), Russian, Spanish (Latin American), and Spanish (Spain).

3 In order to complete the research, CASL obtained licenses to access Rosetta Stone and researchers systematically investigated all of the lessons in Spanish and Arabic.

4 For a review of the content of V3, see Appendix I.

5 All problems were taken from the CASL technical reports mentioned previously (M.7 Review of Technology Mediated Language Training Programs and E.3.2. Rosetta Stone Findings).

6 Unless otherwise indicated, all claims by Rosetta Stone were taken directly from the company's website on February 9, 2008.

7 See Appendix II for a complete comparison of V3 Spanish and Arabic.

8 This claim was excerpted from written communication with Rosetta Stone Technical Support on May 7, 2008.

9 The Dynamic Immersion Method is Rosetta Stone's language learning philosophy. They claim that it "will teach you a new language the same way you learned your first language: by directly associating words (written and spoken) with objects, actions and ideas that convey meaning." The website goes on to say that Rosetta Stone "uses the Dynamic Immersion Method to simulate a real-life immersion experience" (Retrieved May 7, 2008 from http://tinyurl.com/6hpvyw).

Corresponding Author and Reprints: Katharine Nielson, MA, University of Maryland Center for Advanced Study of Language, knielson@casl.umd.edu, www.casl.umd.edu.

UNIVERSITY OF MARYLAND
CASL
CENTER FOR ADVANCED
STUDY OF LANGUAGE

APPENDIX I

Overview of the Changes to the Structure of Rosetta Stone V3

Our review of V3's new features found some variety in the format of the exercises with the addition of a Milestone feature that provides a written simulation of a dialogue. The program continues to rely on the concept-image mapping approach, and students complete exercises where they either match pictures to text or sound, or respond to picture-based prompts in speech or text.

While V2 was largely focused on vocabulary exercises, V3 includes more grammar drills. For example, in V3 Spanish, in Lesson 1 of the first unit, learners are prompted to choose the article that agrees with a noun. In addition, V3 now includes exercises on subject-verb agreement and spelling.

See Table 1 for a comparison of exercises included in versions 2 and 3.

Table 1. A comparison of the exercise topics in Rosetta Stone Version 2 and Version 3.

Version 2 Exercises	Version 3 Exercises
• Adding, subtracting, multiplying and dividing	• Ages
• Body parts	• Body parts
• Clothing	• Calendar terms
• Comparatives and superlatives	• Cities and countries
• Comparing quantities and amounts	• Clothing
• Countries	• Colors
• Descriptive adjectives, including colors	• Comparing and contrasting
• Family relationships	• Descriptions
• Food, eating, and drinking	• Family relationships
• Furniture and instruments	• Greetings
• Giving directions	• Household items
• Names	• Introductions
• Numbers	• Landmarks
• Pets	• Numbers
• Professions and human conditions	• Personal hygiene
• Seasons of the year	• Questions
• Streets, maps, and landmarks	• Shopping
• The alphabet	• Speaking
• Times of day	• Sports
• Vehicles	• Times of day

The V3 exercises are divided between those that require rote memorization of grammatical or spelling rules and those that require learners to recall specific vocabulary words.

APPENDIX II

Lesson-by-lesson comparison of V3 Arabic with V3 Spanish

We conducted an exhaustive comparison of the software in each language to determine whether Rosetta Stone V3 continued to rely on a one-course-for-all-languages approach. In addition, we wanted to determine whether the same photos were used throughout all the courses. See Figure 1 for a graphical representation of the similarities and differences between the Arabic and Spanish versions of the software.
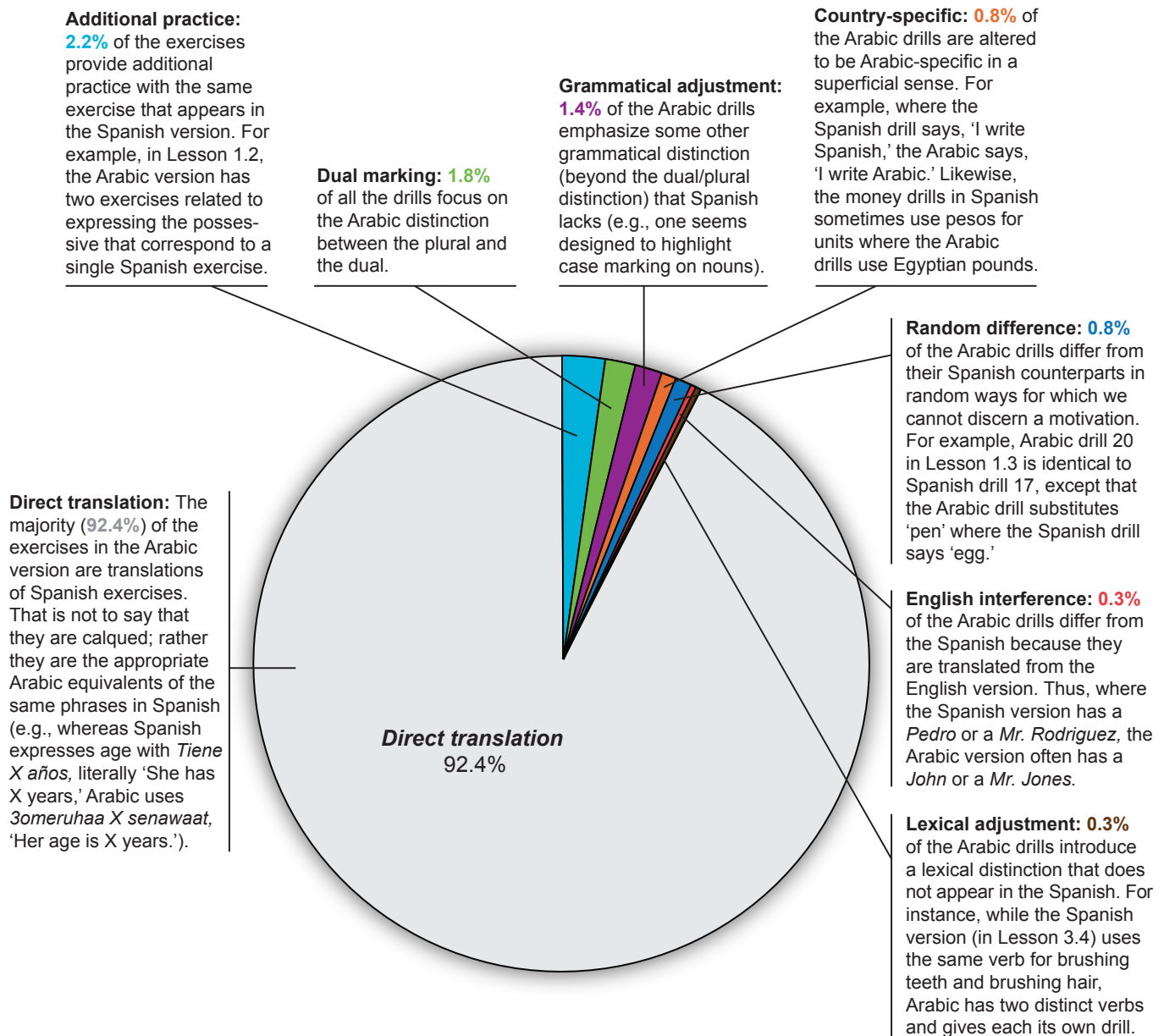
Additional practice: **2.2%** of the exercises provide additional practice with the same exercise that appears in the Spanish version. For example, in Lesson 1.2, the Arabic version has two exercises related to expressing the possessive that correspond to a single Spanish exercise.

Dual marking: 1.8% of all the drills focus on the Arabic distinction between the plural and the dual.

Grammatical adjustment: **1.4%** of the Arabic drills emphasize some other grammatical distinction (beyond the dual/plural distinction) that Spanish lacks (e.g., one seems designed to highlight case marking on nouns).

Country-specific: 0.8% of the Arabic drills are altered to be Arabic-specific in a superficial sense. For example, where the Spanish drill says, 'I write Spanish,' the Arabic says, 'I write Arabic.' Likewise, the money drills in Spanish sometimes use pesos for units where the Arabic drills use Egyptian pounds.

Random difference: 0.8% of the Arabic drills differ from their Spanish counterparts in random ways for which we cannot discern a motivation. For example, Arabic drill 20 in Lesson 1.3 is identical to Spanish drill 17, except that the Arabic drill substitutes 'pen' where the Spanish drill says 'egg.'

Direct translation: The majority (**92.4%**) of the exercises in the Arabic version are translations of Spanish exercises. That is not to say that they are calqued; rather they are the appropriate Arabic equivalents of the same phrases in Spanish (e.g., whereas Spanish expresses age with *Tiene X años,* literally 'She has X years,' Arabic uses *3omeruhaa X senawaat,* 'Her age is X years.').

English interference: 0.3% of the Arabic drills differ from the Spanish because they are translated from the English version. Thus, where the Spanish version has a *Pedro* or a *Mr. Rodriguez,* the Arabic version often has a *John* or a *Mr. Jones.*

Lexical adjustment: 0.3% of the Arabic drills introduce a lexical distinction that does not appear in the Spanish. For instance, while the Spanish version (in Lesson 3.4) uses the same verb for brushing teeth and brushing hair, Arabic has two distinct verbs and gives each its own drill.

Direct translation
92.4%

Figure 1. Graphical representation of differences between Rosetta Stone V3 Arabic and Spanish.

APPENDIX II, CONTINUED

The general reliance on translation from English (albeit less rigid than in V2) still gives rise to a number of difficulties. Among these are the tendency to provide subject pronouns even though Arabic is a language that generally omits these. The same tendency has been recognized and corrected in a number of the Spanish drills, but the Arabic version apparently maintains superfluous pronouns because its drills have been translated from English.

Also, Arabic speech acts, such as greeting and thanking, involve different procedures than corresponding speech acts in Spanish and English. For instance, Arabic speakers in responding to greetings tend to amplify them (e.g., a common response to *marhaba* 'hello' would be *marhabtain* 'two hellos'). Many Arabic speech acts also entail formulaic phrases of religious origin (e.g., the response to 'How are you?' is often 'I'm well, thank God.') Because most of the drills in the Arabic version of Rosetta Stone are translated from an English template, they miss these differences.

Rosetta Stone greetings
(from Lesson 1.1)

A.	*Marhaba.*	Hello.
B.	*Marhaba.*	Hello.

(from Lesson 2.4)

A.	*Kaif haalik?*	How are you?
B.	*Ana mariida.*	I'm sick.
	Kaif haalik?	How are you?
A.	*Ana biHair.*	I'm well.

Authentic Iraqi greeting sequence
(from the Linguistic Data Consortium's transcripts of recorded Iraqi Arabic conversational telephone speech)

A.	*Alu.*	Hello.
B.	*Alu.*	Hello.
A.	*Alu Alu.*	Hello hello.
B.	*Alu shlawnik, aHuuii 3aziiz?*	Hello, how are you my dear brother?
A.	*Alu marhaba Allah yesalmik*	Hello, greetings, God rest you.
B.	*Kaif haaluk? Shlawnik ziin inta Inshallah.*	How are you? You're well, I hope (lit. God willing).
A.	*Alhamdullilah biHair, Allah yesalmik . . .*	Thank God I'm well. God rest your soul . . .

Authentic Modern Standard Arabic greeting sequence
(between a native and a nonnative speaker)

A.	*ahlan.*	Welcome.
B.	*ahlan biik.*	Welcome to you, too.
A.	*Kiifik?*	How are you?
B.	*Alhamdulilah.*	Thanks be to God.
A.	*Ma huii aHbarik?*	What's your news?
B.	*aHbarii djayid.*	All of it good.
A.	*hamdullilah . . .*	Thank God . . .