Objectives, methods, and techniques for ‘surge’ Somali acquisition and instruction

Christopher Green, PhD, Ewa Golonka, PhD, Nikki Adams, PhD, Carrie Bonilla, PhD, Martyn Clark, PhD, Medha Tare, PhD, and Evan Jones, MA

CONTENTS

Introduction ........................................................................................................................................ 2
Somali language situation .......................................................................................................... 2
Perspectives on ‘surge’ language acquisition ............................................................................. 2
SLA scholarship relevant to ‘surge’ language instruction ............................................................ 4
Recommendations for student selection ..................................................................................... 4
Recommendations for instruction ............................................................................................... 5
Current methods and materials for Somali language instruction ..................................................... 13
Classroom learning .................................................................................................................. 13
Online learning ........................................................................................................................ 13
Supplemental learning materials .............................................................................................. 15
Proposed linguistic correlates of proficiency .................................................................................... 17
Characteristics of Somali ................................................................................................................. 17
Orthography .................................................................................................................................. 17
Typological considerations ....................................................................................................... 18
Methods and techniques for Somali instruction ............................................................................. 19
Sample activities incorporating CASL HLT tools ........................................................................ 21
Concluding remarks ......................................................................................................................... 25
References ...................................................................................................................................... 26
Appendix A: Proposed Linguistic Correlates of Proficiency for Somali ..................................... 29
Appendix B: Challenging linguistic characteristics of Somali..................................................... 32
INTRODUCTION

This report summarizes methods and techniques recommended for the ‘surge’ or ‘rapid-rise’ scenario for Somali learning from *ab initio* (Interagency Language Roundtable (ILR) Level 0) to general working proficiency (ILR Level 3). This ‘surge’ scenario requires students to begin to apply their language knowledge to real-world tasks for their jobs or assigned mission needs at an accelerated pace. Therefore, to be effective, instruction should be focused on developing these abilities through a task-based approach that approximates and/or incorporates job- or mission-related scenarios. Instruction should also be based on current Second Language Acquisition (SLA) theory and pedagogical findings and aim to be automated, to the extent possible. An additional challenge is that Somali is highly complex and typologically distinct from most commonly-taught Indo-European languages. Platform developers and instructors should be aware of these unique characteristics and draw upon Human Language Technology (HLT) tools, including those developed by CASL, to supplement instruction.

Somali language situation

Somali is a Cushitic language spoken primarily in Somalia, as well as in parts of Djibouti, Ethiopia, and Kenya. Recent estimates suggest that there are sixteen million first language (L1) speakers of the language, only approximately eight million of whom reside in Somalia itself (Lewis, Simons & Fennig 2013). Large diasporic communities of Somali speakers are found in several cities in the United States (e.g., Atlanta, GA, Buffalo, NY, Columbus, OH, Minneapolis, MN, San Diego, CA, and Washington, DC); the largest Somali population in North America, however, is in Toronto. Sizable communities are also found in the United Kingdom, Finland, and the United Arab Emirates.

‘Standard’ Somali is recognized as the dialect spoken predominantly in northern regions of Somalia and is the official language of the Somali government and military. There are a number of other dialects that vary in their intelligibility with the standard. Important among these are the coastal, or Benaadir dialect, and Al-Ashraaf or Marka, which is most widely spoken in and around the national capital, Mogadishu. Less mutually intelligible, but still considered closely related, are languages like Dabarre, Garre, Jiiddu, and Tunni, which are spoken primarily by ethnic clans of Somalis. Maay, a language spoken widely in southern Somalia, is not generally considered to be mutually intelligible with Somali, although they are genetically related.

Somali contains a large number of borrowed words from Arabic, as well as a sizable number of borrowings from European languages like English, Italian, and French and South Asian languages such as Persian Farsi and Hindi. However, because the phonologies (sound systems) of these languages often differ significantly from Somali, borrowed words may be difficult to recognize, as they may look and/or sound quite different from their source language.

Perspectives on ‘surge’ language acquisition

The need for ‘rapid-rise’ or ‘surge’ language acquisition and subsequent mission readiness arises in response to changing world events and provides a significant challenge to language professionals, strategic planners, and mission managers. While such events are difficult to maneuver even for well-resourced languages, the challenge is compounded when the target of acquisition is a low-density, understudied, and/or under documented language. A prime example of such a language is Somali. Little is known about the science of Somali language acquisition or pedagogy and the appropriate trajectory for Somali language learning. This state of affairs is further complicated in that although much linguistic documentation has been done on Somali, few high quality, cohesive, and comprehensive learning or reference materials are available for the language.

The Somali example mentioned above illustrates the shortcomings of language policy and planning in the US. According to the Strategic Market Forces Framework for Language Policy and Planning (Brecht & Rivers,
2000; Figure 1), events in a specific world region trigger the need for expertise in a particular language; the launch of Sputnik, 9/11, and crises in Kosovo or Somalia triggered the need for expertise in Russian, Arabic, Serbian/Croatian, and Somali, respectively. Ideally, the need for expertise in a particular language should influence demand (actual current requirements for expertise in a language), which, in turn, should reflect supply (available expertise). Supply, consequently, should meet demand and should determine capacity (capability to supply language expertise in the future). However, as Brecht and Rivers (2000) argue, the situation in the US does not reflect this ideal model. In particular, the authors maintain that need and demand are poorly related, with demand usually not responding to strategic need, and that capacity is often affected by demand instead of strategic need and supply. To compensate for these faults, they propose to strengthen US government programs such as Title VI/Fulbright-Hays so that they better meet the strategic need. A surge scenario that requires rapid-rise language instruction is another way to provide adequate supply at the tactical level, which, in the case of Somali, is almost non-existent. While all of these undertakings provide a quick solution for producing adequate supply for Somali, long-term strategic planning of Somali instruction that reflects strategic need and is determined by the actual supply is necessary. Had the situation in the US reflected the assumptions of the Strategic Market Forces Framework, there would be no need for rapid-rise instruction in any language.

Figure 1. Idealized Strategic Market Forces Framework adapted from Brecht & Rivers (2000), p. 21.

Cognizant of these facts, and drawing on CASL expertise in language description, analysis, and second language acquisition, this report details the Somali language situation and provides an overview of learning objectives, methods, and techniques recommended for Somali learners to reach ILR Level 3 (i.e., general working proficiency) in reading and listening from an ab initio (i.e., Level 0) starting point.

The report covers the following topics as they relate to LanguageNation ‘surge’ scenario for Somali:

- SLA scholarship relevant to ‘surge’ language instruction
- Current methods and materials for Somali instruction
- Characteristics of Somali
- Proposed linguistic correlates of proficiency (LCP) in Somali
- Methods and techniques for Somali learning and assessment
- Concluding thoughts and remarks
SLA SCHOLARSHIP RELEVANT TO ‘SURGE’ LANGUAGE INSTRUCTION

In this section, we discuss specific steps that can be undertaken to compensate for the lack of supply of expertise in Somali in the US. In particular, we propose rigorous student selection for surge courses and optimal instruction that is based on research findings from SLA and cognitive psychology. CASL offers expertise in the following student characteristics and SLA-based pedagogy in order to take them into account for surge language instruction.

Recommendations for student selection

Previous language experience

An important consideration when selecting students for a rapid rise course is their prior language experience. In the case of a surge scenario, where specific language skills are needed immediately, and the target language might not be well-researched or documented, selecting students who have already demonstrated proficiency in a closely related language could compensate for the lack of instructional materials and other resources because learners would bring to these courses background knowledge that can be transferred to the target language situation. There are some documented cases (and many more anecdotal, self-reported instances) of adult third language (L3) courses that were shorter than regular second language (L2) courses but produced similar results. For example, Corin (1994) reported on a Serbian/Croatian conversion course at the Defense Language Institute that retrained 40 Czech linguists in Serbian/Croatian in a 3-month period, rather than the typical 9-month period that Category 2 languages usually require. Moreover, even in L3 instruction where languages are unrelated, the experience gained in L2 learning can still make learning a third language easier, because experienced language learners have a much better idea of how to learn a new language. For instance, L3 learners show a high tendency towards learner autonomy (Rivers, 2001), build awareness of their learning styles and preferences, and develop learning strategies (Ramsay, 1980; Wenden, 1999). Rivers (1996) reported that cross-training courses in Georgian and Kazakh (with Russian as L2) were one third the length of basic courses in these Category 3 languages. Anecdotal information from students and instructors of Somali, as well as our own expertise in Somali structure, suggest that cross-training and conversion into Somali from a lexical (vocabulary) perspective, may be easier for L2 learners of Arabic, while from a structural (grammatical) perspective, the same could be said of individuals with expertise in the more closely-related and highly agglutinating language Amharic.

Aptitude

Another important consideration when selecting students for a rapid rise training program is their language learning aptitude. Researchers have determined that some learners have an aptitude for language learning, or an “exceptional language learning ability” (Doughty, 2013 p. 154). Specifically, an aptitude for language learning means that the learner has high scores on measures of the types of cognitive and perceptual abilities that are required for language learning above and beyond individual characteristics such as motivation, personality, and learning experience. Recently, CASL researchers have provided validation evidence for cognitive aptitude measures that predict adult learners’ high-level scores (3+) on the Inter-Agency Language Roundtable scale using the High-level Language Aptitude Battery (Linck et al., 2013). Some of these constructs (e.g., working memory, explicit learning) have also been shown to be predictive of intermediate-level language attainment which occurs over a relatively short period of time (Bunting et al., 2011).

1 Difficulty categories, first developed by Foreign Service Institute (FSI), categorize languages based on the difficulty native speakers of English may have in learning a given language. For instance, Category 1 (easiest) includes languages that are most typologically similar to English, for example, Spanish and Italian; Category 2 includes German and Romanian; Category 3 includes Russian, Persian, and Hebrew; and Category 4 (hardest) includes Arabic, Chinese, Japanese, and Korean.
Abilities predictive of success include short-term memory, associative learning, perceptual acuity and implicit learning. Short-term memory represents the ability to remember new information for short periods of time, associative learning represents the ability to associate meanings with new words, and implicit learning represents the ability to unconsciously determine patterns from the input received. Measuring these abilities and others, such as working memory, in potential learners would provide information on which learners may be able to better handle a rapid rise situation, in terms of cognitively processing a new language quickly and accurately.

Heritage

Knowing learners’ language history can also provide valuable information regarding their readiness to learn or re-learn a language quickly. Childhood exposure to a language which is not used in adulthood can still confer advantages in re-learning that language. Recent work by Au and colleagues (Au et al. 2008; Oh et al., 2003) shows that adults who were childhood speakers of a language, but reported a massive drop in language use after age seven, outperformed childhood hearers and late L2 learners of the language on tests of grammar and phonology. For example, childhood speakers were better at making grammaticality judgments about morphosyntactic errors and at perceiving sentences in noisy contexts in Spanish. Thus, knowing if an adult had exposure to a mission-related target language or related language might inform whether he/she can quickly grasp grammatical or phonological elements when relearning the language. This advantage might be particularly valuable in a rapid-rise situation.

Domain knowledge

Research on L2 processing has shown that domain knowledge, or prior first language (L1) familiarity with the topics encountered, can improve comprehension. In a study of listening comprehension, students showed greater recall of information from a passage on a familiar topic than an unfamiliar topic (Schmidt-Rinehart, 1994). This relates to students’ ability to use schematic, or background, knowledge to anticipate or fill in information which, in turn, aids comprehension. As a result, researchers propose that “taking time to assess the conceptual base the listeners bring to the text will enable teachers to go beyond dealing with the linguistic information in order to help students understand and make their learning more meaningful” (Schmidt-Rinehart, 1994, p. 185). These results suggest that selecting students for rapid-rise training who already have the relevant domain knowledge that they will need to accomplish their mission in the new target language will help to make training more successful. This is also in line with the use of the Task-Based Language Teaching (TBLT: Long, 1985; in press) approach discussed below where L2 learning takes place with the specific mission needs in mind so that language functionality is built up using domain-relevant tasks and content.

Recommendations for instruction

Needs analysis

Effective language programs are those that are tailored to meet the specific needs of program participants and stakeholders. As indicated above, these needs can be at many levels, from identifying the societal needs relevant to developing national language capacity (e.g., Brecht & Rivers, 2005), to selecting appropriate individuals to undergo the training based on their particular aptitude or experience, to collecting diagnostic information on an individual learner to align instruction with learner characteristics (e.g., Cohen, 2003). Identifying needs can be established through a process called needs analysis (see Long, 2005; Tare et al., 2013), the goal of which is to provide information that can help tailor instruction to the learners and the mission based on a careful evaluation of the learners’ real world target language needs.
Needs analysis is defined as a process for the “systematic collection and analysis of all information necessary for defining a defensible curriculum” (Brown, 2009, p. 269). Thus, the first step in the design of any foreign language learning course or materials is to consider what the purposes of the training are and what kind of syllabus will be used to structure the course. Conducting a needs analysis involves questioning the learner(s) for their language learning needs and goals and researching the kinds of tasks and language the learner will need to be able to accomplish those goals. This research can include observing and/or interviewing people using the language in the context for which the learners are being trained. The information collected then drives the content of the course (see Brown, 2009).

Once the range of language learning needs and tasks has been identified, the nature of the syllabus should be established. While many traditional language learning materials focus on a grammatical syllabus (e.g., based on starting with simple grammatical concepts, such as subject/verb agreement and moving toward increasingly complex grammatical concepts), alternative, more effective, options for designing syllabi, particularly in a rapid rise scenario, are based on a learner’s ultimate target language needs. This could take the form of a content-based syllabus (e.g., focused on learning relevant subject matter areas in the target language) or, preferably, a task-based syllabus (e.g., focused on the accomplishment of specific target language tasks that increase in difficulty throughout the course). In the learning scenario in this report—rapid-rise—appropriate input will need to run the gamut from language needed for routine daily interactions in the language to the kinds of input that involve higher level language processing for advanced learners, such as biographies, plays or TV shows (for contemporary cultural references), blogs, documentaries, film/book reviews, scientific articles, or newspaper editorials (cf. Minuchehr & Mills, 2012).

Task-based instruction

It is important to emphasize that it is not only tailoring the content of the course that is crucial, but also incorporating relevant tasks that are directly related to the needs of the learners. Task-based language teaching uses target language tasks, and not inventories of grammatical features, as the basis for syllabus design. For example, if a learner eventually needs to be able to conduct business in the target language, having learners study lists of business-related vocabulary, or reading business-themed newspaper articles with multiple-choice comprehension questions is not sufficient for maximal language acquisition. A learner must engage in business-related tasks (e.g., making a sales call or conducting a meeting) in order to merge their needs with the real-life tasks they may engage in (see Doughty & Long, 2003, for a discussion of task-based approaches in distance learning). At the same time, while the learners must be able to cope with real world target tasks at the end of the instructional sequence and should be given ample opportunity to use language in the context of real world tasks, this does mean that learners are expected to endlessly repeat real world tasks until they ‘get it.’ In cases where the real world target task might be beyond the current capabilities of the learners, the target task is broken down into a set of pedagogical tasks which help to scaffold the learner. The important point here is that these pedagogical tasks are derived from real world needs as identified by a careful needs analysis, and are not merely vehicles for the delivery of a grammatically-based syllabus. In the task-based syllabus, drawing learners’ attention to the grammatical features of target language can and does occur, but always in the service of meaningful communication with the goal of developing form/meaning mappings (Doughty & Williams, 1998).

A prototype task-based lesson on giving directions in Korean was developed by Long, Doughty, Kim, Lee, and Lee (2003). A needs analysis with second language learners of Korean revealed a large number of students who intended to visit Korea and would need to understand and possibly give street directions. To prepare learners for this real world target task, a series of nine pedagogical tasks were developed. The first task involved listening to actual samples of target discourse (i.e., street directions given by native Korean speakers in the local context) while the instructor traced the route described on a projected map. For the second task, learners were asked to trace routes on a map themselves, based on increasingly complex authentic direction fragments read by the instructor. This was followed up with another listening task using a more complex map, and several subsequent tasks in which learners worked in pairs to give and follow directions, using transcripts and marked routes at first,
but gradually decreasing the availability of such aids. The final tasks involved learners giving directions unaided to the instructor and peers. Learner progress is assessed in terms of how well they can perform the target tasks. Although many language courses include practice with maps and giving directions, of particular importance in this example is that the task was motivated through an analysis of learner needs and the pedagogical tasks were derived from real world target language discourse.

SLA research has identified several task features that may influence learner performance on various tasks (e.g., Byrnes, 2002; Long & Norris, 2000; Norris, 2002; Norris et al., 2002; Robinson, 2007; Skehan & Foster, 1999). Some task factors place cognitive or interactional demands on learners, and those demands are hypothesized to affect language learning either positively or negatively. A very useful conceptualization of task design as it relates to language curriculum can be found in Robinson’s Triadic Framework for task classification (Robinson, 2007), in which tasks are classified according to three types of factors: task complexity, task conditions, and task difficulty. “Task difficulty” in this framework refers to the characteristics of the learner and will vary across individuals given their cognitive ability profiles and other affective factors. Because the syllabus designer is usually only able to control task complexity, Robinson (2010) presents a general framework for task sequencing which focuses specifically on those task complexity variables listed in Table 1. The presence (+) or absence (-) of these variables determines the relative complexity of the task, with the less complex state indicated first for each variable. Thus, for some variables, the presence of the variable makes the task less complex (+/-), while for others, the absence of the variable (-/+), makes the task less complex.

| Table 1. Task complexity variables adapted from Robinson (2010), p. 257 |
|---------------------------------|---------------------------------------------------------------------|
| **Task Complexity**             | **Thinking**                                                        |
|                                 | Resource-directing, developmental dimensions                        |
|                                 | +/- here and now                                                    |
|                                 | +/- few elements                                                    |
|                                 | -/+ spatial reasoning                                               |
|                                 | -/+ causal reasoning                                                |
|                                 | -/+ intentional reasoning                                           |
|                                 | +/- perspective taking                                              |
| **Reacting**                    | Resource-dispersing, performative dimensions                        |
|                                 | +/- planning time                                                   |
|                                 | +/- prior knowledge                                                 |
|                                 | +/- single task                                                     |
|                                 | +/- task structure                                                  |
|                                 | +/- few steps                                                       |
|                                 | +/- independency of steps                                           |

As seen in Table 1, these complexity variables are divided along two dimensions: resource-directing and resource-dispersing. Resource-directing dimensions are those which direct the learner’s attention on the target language linguistic system to meet task demands. For example, talking about events in a shared context (+ here-and-now) will be less demanding than talking about events that happened elsewhere at another time (- here-and-now); similarly, tasks that don’t require discussing complex reasoning or inferences about the intentional states of other people (- causal reasoning, -intentional reasoning) will be less demanding linguistically than those that do. Resource-dispersing factors, on the other hand, are task characteristics that affect task difficulty but do not necessarily direct learner attention to particular aspects of the language. For example, a task with few steps and sufficient planning time (+ single task, + planning time) will be less demanding than a task with many steps and no planning time.

All of these task complexity factors can potentially be manipulated by instructors or software developers to sequence pedagogic tasks in a manner conducive to language learning. In general, one starts with the least
complex version of a task that is reasonable for a given learner. In LanguageNation, this decision will be informed by the learner model. Then, resource-dispersing dimensions of the task are increased. This allows the learners to develop automaticity with their current linguistic resources. Next, the resource-directing dimensions are increased, promoting the restructuring of interlanguage through new form/meaning mappings. Principles of input processing (described in a later section) could be particularly relevant here. Note that not all complexity variables will be relevant for all tasks, and, as in the Korean map example above, the ultimate aim is to map pedagogical task demands to real world target task needs in a coherent way (Robinson, 2010).

**High-frequency and domain-relevant vocabulary**

When planning the vocabulary component of a well-balanced language course, it is important to consider word frequency and range as the basis for vocabulary selection. Research shows that a learner needs between 95% and 98% coverage of the running words to read a text in English without assistance (Hu & Nation, 2000; Laufer, 1998; Nation & Chung, 2009). One way to get learners to such high level of coverage is to teach them the right words. High-frequency words with a wide range of use are a relatively small class, yet they are the most important group of words that occur often in a language. Nation and Chung (2009) argue that because the 2,000 most common English word families typically cover between 80% and 90% of the running words in a given text, depending on the type of text and the corpus used to estimate coverage, it is reasonable to teach vocabulary according to frequency of occurrence. Using a corpus of The New York Times, the authors calculated that the first thousand word families² from the frequency list covers about 76% of the newspaper’s texts. Then the coverage figures drop rapidly: the second thousand words adds about 8% of coverage, the third and fourth thousands add 3% each, and the sixth thousand adds less than 1%. High-frequency vocabulary lists can be created using word frequency dictionaries, if available, or using representative corpora of a given language. Another group of words that should be selected for instruction is specialized vocabulary (e.g., academic, technical, business, and medical) that meets future language use needs of the learner, which can be determined through needs analysis. For LanguageNation, CASL has generated frequency-based pedagogical and news corpora for use in creating such lists for Somali instruction.

**SLA principles**

The surge scenario necessitates building on the principles that are crucial for initial second language acquisition, i.e., sufficient, varied, and comprehensible input combined with opportunities for output and interaction, and provision of feedback for learners to process and correct errors. These principles originate in the Interaction Framework, the widely-accepted position in SLA that the interrelated processes of input, interaction and output, and feedback are necessary (although may not be sufficient) for psycholinguistic gain; e.g., changes in the learner’s interlanguage toward a target language norm (Gass & Mackey, 2006). Methodological principles for optimization of these elements are listed below, based on Lafford et al. (2007), Doughty and Long (2003), and Chapelle (1998). See also Nielson et al. (2009) for a review of online language course design.

*Input (receptive exposure to the target language)*

- Input should be authentic (culturally, sociolinguistically, and pragmatically) and rich (Lafford et al., 2007). To achieve high levels of proficiency, learners must be able to deal with a wide variety of written and spoken texts.
- Input must be comprehensible (Chappelle, 1998; Doughty & Long, 2003). The goal of language is to convey meaning. As such, the following input principles should be followed:

---

2 A word family consists of a base word and its inflected and derived forms (Bauer & Nation, 1993).
Input should be elaborated, not simplified, to the appropriate development level of the student (Doughty & Long, 2003; Long and Ross, 2009; Yano, Long, and Ross, 1994). Elaboration, rather than simplification, helps to maintain the richness of the language while making it comprehensible.

Lexical and grammatical (syntactic, semantic, and morphological) structures can be made salient (e.g., through textual highlighting; Lafford et al., 2007). Learners need to be aware of the language forms used to convey various meanings even if they cannot explain every lexical and structural rule.

A variety of resources should be available to help learners understand the input (e.g., dictionaries, glosses, and pictures; Lafford et al., 2007). This helps them become self-directed learners.

Multimodal input (video + subtitles, written text + pictures) will aid learners to process input more deeply (Lafford et al., 2007). Multimodal input can increase saliency and comprehensibility.

Output (production of the target language)

- Learners need to produce the target language in various modes (oral and written) multiple times in various contexts as well as in isolation (Chapelle, 1998; Lafford et al., 2007; Swain, 1985). Fluent language use requires practice in using the language.

Interaction (receptive and productive communication in the target language)

- Learners need to interact in the target language to facilitate negotiation for meaning (Chapelle, 1998; Long, 1991), which promotes noticing of target language features – a necessary component for acquisition (Long, 1991; Schmidt, 1990). This interactional adjustment can make input comprehensible.

Feedback (positive or negative feedback to target language performance)

- Learners need feedback on production to assist with noticing target language features (Doughty & Long, 2003); as well as opportunities to correct errors (Chapelle, 1998; Lafford et al., 2007). Learners need to be able to convey meaning in the language, but achieving high levels of proficiency also demands a requisite level of control over lexical and grammatical structures.

- Teachers can facilitate occasional focus on form to draw learners’ attention to target structures (Doughty & Long, 2003; Doughty & Williams, 1998; Long & Robinson, 1998). Activities should focus primarily on meaning, but learners also need to recognize how language forms encode that meaning, especially for more subtle features of the language.

Incorporating these four elements into tasks or activities that learners will engage in throughout a course can take many forms. In the discussion, we will return to these general principles of SLA to make more specific recommendations for the Somali surge scenario.
Cognitive principles of learning

In addition to SLA principles that are specific to learning foreign languages, it is useful to consider general learning principles, derived from years of research in cognitive psychology and education. Out of the 25 general learning principles identified by a group of 30 experts from different areas of the learning sciences, such as psychologists, science instructors, and representatives from the National Science Foundation, we present four that are particularly relevant to adult L2 learning and are supported by empirical evidence (Graesser, Halpern, & Hakel, 2008; Halpern & Hakel, 2003). A variety of activities provided through the LanguageNation platform can be driven by these principles.

Dual Coding Effect

The Dual Coding effect refers to the fact that both the initial encoding of information and the memory for that information are improved when information is presented in different modes (e.g., verbal vs. pictorial), modalities (e.g., auditory vs. visual) or media (e.g., computer vs. paper; Greasser et al., 2008). In a theory first put forward by Allan Paivio, it was postulated that human learning unfolds primarily through verbal and visual associations, and these two types of information are processed and stored separately (Paivio, 1969, 1971). Thus, these types of information can serve as independent representations of a given piece of information, thereby increasing the likelihood of remembering it (Clark & Paivio, 1991; Mayer & Anderson, 1991). Some evidence exists that this principle may be especially helpful for L2 vocabulary learning. For example, Shen (2010) found that dual coding of vocabulary items increased both initial learning performance, and retention performance after a 1-day interval for students learning the meaning and character shapes of abstract Chinese words. In the context of LanguageNation, this principle can be implemented through the use of text+picture glosses or providing captions for videos.

Desirable Difficulties Effect

The Desirable Difficulties Effect refers to the finding that some activities that make learning initially more difficult can lead to better long-term retention of information (Halpern, 2008; Schmidt & Bjork, 1992). Such difficulties might be instantiated by requiring the learner to search out and organize information, rather than passively receiving information organized by the teacher (Greasser et al., 2008), by alternating tasks, rather than blocking them (Richland, Bjork, Finley, & Linn, 2005; Taylor & Rohrer, 2010), or by presenting new vocabulary in thematic categories such as “beach,” and “winter” rather than grouped into categories under a shared hypernym such as “food” or “clothes” (Schneider, Healy, & Bourne, 1998). Although initial learning may be slower and more error-prone when such difficulties are introduced, long-term retention of information should be greatly improved. Technology can support this principle by modifying input to make it more challenging (e.g., by omitting key information that the learner must supply).

Testing Effect

The Testing Effect reflects the finding that frequent testing promotes learning (Karpicke & Blunt, 2011; Karpicke & Roediger, 2007). When tests are given frequently, students are encouraged to maintain a study schedule, teachers and students receive important feedback on students’ progress, and students are less likely to forget material (Greasser et al., 2008). In fact, research shows that practice at retrieval is a very important variable in promoting long-term retention of information (Halpern & Hakel, 2003). That is, all practice at retrieving information (regardless whether or not graded by the teacher, or self-tested by students) may help long-term memory. Karpicke and Roediger (2008) found that the attempt to retrieve learned information through testing, rather than passive study, was the critical variable in memory for that information after a delay, as measured by recall of foreign language word pairs. Such tests should provide the learner with minimal cues and require the learner to generate the answer (e.g., tests of recall, not of recognition; Graesser et al., 2008). This
principle can be supported by technologies such as electronic flashcards because the learner is required to attempt recall of a word’s meaning and then receives immediate feedback on his or her accuracy.

Spaced Practice Effect

The Spaced Practice Effect, also known as distributed practice, refers to the finding that learners show higher learning performance and greater long-term retention for material when practice is distributed across several smaller sessions (spaced), rather than completed in a large block (massed; Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). In the context of LanguageNation, this principle might be implemented through the use of mobile technologies to encourage practice of material at intervals throughout the day or at regular daily intervals. Thus, the system could prompt students to practice frequently for shorter periods of time (e.g., 20 minutes per session) rather than for extended periods only occasionally.

Processability Theory

When designing a language course, it is essential that any specific structures being targeted for instruction or feedback correspond to a learner’s developmental stage. While individual differences such as language aptitude or motivation may set learners apart, studies on cross-linguistic patterns of language development have often shown that learners follow fundamentally similar routes, especially in the beginning stages. Processability Theory [PT] (Pienemann, 1998, 2005) is one theory that has evolved from several decades of research on the order of acquisition of syntax and morphology to account for these developmental similarities. PT is centered on the premise that the ability to produce speech in a second language is limited by the one-by-one acquisition of five speech processing procedures, all of which are the same procedures by which a mature L1 speaker generates grammatical utterances. In other words, the main claim of PT is that learners’ ability to produce language is limited by the gradual acquisition of language-specific processing procedures. These procedures are based on Levelt’s (1989) speaking model (Pienemann, 2005, p. 2) and complemented by a theory of grammar, Lexical-Functional Grammar [LFG] (Kaplan & Bresnan, 1982), which is a model of grammar that reflects many of the psycholinguistic principles prominent in Levelt’s (1989) theory of production. The hierarchy predicted by PT has been accepted as a valid theory of second language development, e.g., in German (Baten, 2011; Jansen, 2008; Pienemann, 1998, 2005), Italian (Di Biase & Kawaguchi, 2002), Chinese (Zhang, 2005), and Spanish (Bonilla, 2012).

Table 2 illustrates the PT processing hierarchy. The first column shows the stage, and the subsequent columns show which general morphological and syntactic structures are predicted to be acquired at a given point in time (T). Importantly, learners cannot skip stages; rather, each stage must be acquired in the order predicted because learners are limited by the processing capabilities of their current stage.
Table 2. Processing procedures predicted in PT.

<table>
<thead>
<tr>
<th>Procedures and Stages</th>
<th>T1</th>
<th>T2</th>
<th>T3</th>
<th>T4</th>
<th>T5</th>
</tr>
</thead>
<tbody>
<tr>
<td>S'-procedure (EmbeddedS) (Stage 5)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>S-procedure (Stage 4)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Phrasal procedure (Stage 3)</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Category procedure (Stage 2)</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Word / lemma (Stage 1)</td>
<td>‘words’</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

As illustrated in the first column in Table 2, the stages that PT proposes that all learners must acquire in a one-by-one manner are based on the same procedures by which competent speakers produce utterances. At Stage 1 (word/lemma), learners are limited to producing lemmas in an isolated fashion. Next, at Stage 2 (category procedure), learners begin to mark lexical items with lexical morphemes, e.g., a marker such as –s in English that indicates plurality. Their word order at this stage is predicted to rely on mapping the most prominent theme, i.e., the subject, to the initial position in the sentence. At Stage 3 (phrasal procedure), learners can map features beyond one lexical item, to extend to the entire phrase, for example, marking plurality within the noun phrase in a language such as Spanish. Phrasal procedure extends to the syntax in that learners begin to mark adverbials, or other adjunctions prior to the subject of an utterance. At Stage 4 (S-procedure), learners are able to match features beyond the phrasal level to across the utterance e.g., subject-verb agreement in English. In terms of syntax, learners will be capable of producing more target-like word order in independent clauses. At Stage 5 (S’-procedure), learners show the ability to subordinate clauses within a main clause, which can lead to learners producing agreement marking across clauses, as well as particular word order characteristic for a given language.

In a surge scenario, especially if dealing with less commonly taught languages, it is unlikely that research would be available on the order of acquisition of syntax and morphology in the target language(s). The application of PT theory to the LanguageNation platform is most likely relevant to providing feedback on error or syllabus design. The general stages illustrated in Table 2 might be useful as guidelines for which structures targeted feedback on error might be most useful, especially for low-resourced languages.
CURRENT METHODS AND MATERIALS FOR SOMALI LANGUAGE INSTRUCTION

Classroom learning

Somali, like most other African languages, is considered a Less-Commonly Taught Language (LCTL). Despite the fact that several United States cities have become centers of large diasporic communities of Somali speakers, formal university classroom instruction in Somali is regularly offered at only a few universities: Ohio State University (Columbus, OH), Ohio University (Athens, OH), and University of Minnesota (Minneapolis, MN). Such instruction is available only for the beginning (year 1) the intermediate (year 2) levels. In addition to language instruction, Ohio State University offers a minor in Somali Language & Culture. Other, more popular African languages, such as Swahili, are offered into the third or fourth year at some universities. Somali was previously offered at the Summer Cooperative African Language Institute (SCALI); however this program has been discontinued due to US Title VI funding shortfalls. Outside of the United States, classroom instruction in Somali is offered at several European universities, most notably the University of London, School of Oriental and African Studies (SOAS), which offers courses through the advanced (year 3) level. Outside of the university setting, at least one public high school in Minneapolis, MN, began offering elementary courses in Somali language and culture in 2012.

Some independent (non-university) courses exist for Somali, typically organized by Somali cultural and community centers, such as the Somali Community Services Coalition in Washington state or organizations like the Upper Midwest Translators and Interpreters Association, which offers a course on Advanced Somali Court Interpretation. Similarly, the US National Cryptologic School now offers classroom instruction in Elementary Somali. Somali is also regularly offered via vendor instruction, such as Dictyon (Greenbelt, MD) and the DC Internationals Middle East and South Asia Language Institute (Washington, DC). Somali is taught at the United States Department of State Foreign Service Institute (FSI) by special arrangement only.

Available course curricula and syllabi reveal that topics and trajectories for instruction vary greatly; however in nearly all instances, classroom instruction is supplemented by one or more pedagogical textbooks or workbooks, drawn from the list of materials reviewed below. Stated objectives in most instances suggest that emphasis is placed on function-based or communication-based learning, as well as awareness of important cultural and sociolinguistic topics necessary for effective communication. As indicated below, only Zorc & Issa (2002) references overall learning in terms of the general proficiency that is described in the ILR scale.

Online learning

Few online courses have been developed for Somali learning. There are, however, a small number of online resources available for Somali teachers and students to supplement classroom instruction.

Online independent-study courses

The United States Special Operation Command (USSOCOM) has developed a web-based 200 hour online self-study course currently available at the Joint Language University (JLU) website. The course is designed to take a student from ILR level 0 to 0+/1 with 20 language lessons in 200 hours. Each lesson has many activities, such as listen and repeat, role play, and circling which word was spoken in a sound clip. Though the majority of the activities associated with the lessons can be completed by the individual, several assume a language study partner, e.g., “exchange greetings with your teacher and your partner.” The activities for which there are determinably correct answers have those answers listed in an answer key for each lesson. The activities make reference to a workbook which is not available online with the course.
An updated version of this course was available online at the JLU website in July 2013 that featured more interactive exercises. This version of the course was also accompanied by a textbook/workbook for the student and a teacher manual which was generic to all language teachers. These materials have since been removed from the JLU website. Upon contacting JLU, CASL researchers learned that there are no current plans in place to make these course materials available.

The JLU website also has Language Survival Kits for Somali. These were developed by Defense Language Institute Foreign Language Center (DLIFLC) and are essentially brief, topic-delimited phrasebooks. Complete English sentences are given with their corresponding Somali translations. There are no vocabulary lists or grammar instruction.

DLIFLC recently launched the Somali Headstart 2 online introduction to the language, which consists of 10 modules under “Sound and Script” and 10 modules under “Military.” The Sound and Script modules are, in fact, more than familiarization with the sounds and orthography of Somali and cover a small amount of basic vocabulary in several functional areas (e.g., “Asking Questions,” “Telling Time,” and “Making Appointments”). The Military modules have more targeted vocabulary and phrases for use in military operations. They are structured such that each situation concerns scenarios military personnel are likely to encounter, e.g., medical situations and cordon and search. The course as a whole is designed for military personnel preparing to deploy to Somalia. The vocabulary is therefore limited and targeted to that end. The course also includes an introduction to Somalia in general and an introduction to Somali culture. Additionally, it has a pronunciation guide focusing on sounds that would be difficult for native speakers of English and a Sound Recorder, where the students can record themselves saying Somali words and listen to the playback. The Sound Recorder does not provide a way to evaluate how accurate the student’s pronunciation is, outside of self-evaluation based upon that playback. Activities for the student include typing/spelling exercises, fill-in-the-blank, and vocabulary matching games. The learning is vocabulary-oriented, and little explicit grammar is taught, though each module has a grammar note. After completing all the activities for all of the modules, and completing each module’s “standards” (a quiz), the student can receive a certificate of completion.

**Online teacher-directed courses**

The JLU web platform also hosts occasional teacher-directed courses. The most recent of these was at “Intermediate-High” level and was taught in summer of 2013. The materials are not available online without registering for the course, so they could not be assessed.

**Online assessment**

The Online Diagnostic Assessment (ODA) Tool at the DLIFLC website is not an instructional tool, but rather is designed to automatically assess a language learner’s strengths and weaknesses and then to provide that assessment as feedback to the student. There are separate assessments for listening and reading. The feedback is given in the form of a test-taker’s estimated ILR level and a breakdown of strengths and weaknesses according to skill set (e.g., understanding content, structure, vocabulary, etc.). Both the listening and reading tests are described as “semi-adaptive;” that is, the difficulty level of each section of the test depends on how the test-taker performed on the preceding section. The assessment takes 1-2 hours to complete per component (listening or reading). Neither the listening nor reading assessments are currently available for Somali, but the DLIFLC website lists the Somali listening assessment as under validation. CASL plans to examine the Somali ODA once it is available.
Online study aids

There are several online Somali dictionaries of varying quality and reliability. These include Marka Cadey (markacadey.net), Free Lang (freelang.net), and Red Sea Online (redsea-online.com). However, the most comprehensive and user-friendly of the online dictionaries is the Somali-English dictionary found at afmaal.com. The website states the dictionary has 30,000 entries. It includes an auto-complete feature that allows the user to type in a portion of a word and to see all possible completion matches. The afmaal.com dictionary is currently only Somali-English, but the website states the English-Somali portion will “be online in a few weeks.” It is unclear when or if the English-Somali portion will be available.

Both the DLIFLC website and the JLU website have ILR-graded text and audio with which students can practice reading and listening in the language. Again, these are not instructional but rather assume a student is already at the particular level of the text or audio that they are accessing. There is no method in these to improve students’ ability or allow them to move to a higher level outside of a new vocabulary list.

Supplemental learning materials

Textbooks and workbooks


This textbook contains 18 lessons and employs a functional, activity-based method of instruction with chapters devoted to providing students with concepts, cultural awareness, and vocabulary needed for them to become increasingly productive and conversational in the language. Grammatical concepts supporting a particular concept or description are also included in each chapter. Explanations are simplistic and often contain conceptual or factual errors. The exercises included in each chapter rely heavily on availability of a dialogue partner; there are no answer keys for more individualized activities. Exercises are almost entirely question/answer or description based; there are no word retrieval or grammar-based drills. Some exercises focus on situational communication, rather than solely upon structure. The content of the text likely aims for students to achieve ILR 2 in speaking; texts are minimal, suggesting perhaps only ILR 1 in reading may be attained.


Despite its name ‘We speak Somali,’ this resource in French is more of a pedagogical grammar than a textbook or workbook. Nearly half the book is devoted to a description of Somali history and culture; grammatical sections are based largely on part of speech, rather than on increasingly complex topics or constructions. The book does, however, explain grammatical concepts in a coherent way in comparison to other comparable resources. Included examples are successful in illustrating concepts, despite a poor glossing scheme. For students with reading proficiency in French, this is likely a suitable reference for learning through ILR 3.


The author describes this text as a resource for intensive beginning Somali that relies heavily on input from an instructor and availability of a dialogue partner. The focus of the course is on listening and speaking, and the method of instruction is largely inductive; grammatical concepts are intended to be induced by the students following an activity. The book contains 18, approximately 3 hour units with accompanying exercise worksheets. The final lessons introduce the past tense, adpositions, auxiliaries, and the use of the impersonal pronoun, suggesting that students completing the course may only just reach ILR 2 in speaking, and perhaps in other skills, if not lower.

This book functions more like a stand-alone, self-study course on Somali and is available commercially with recordings to illustrate dialogues and accompanying exercises. There are 14 lessons that are tied closely to grammatical concepts. Exercises are almost entirely based on translation, with some fill-in-the-blank and matching. All dialogues contain an English translation; thus, no reading comprehension exercises are included. The conceptual and grammatical progression of lessons is more balanced than in comparable materials. This book accompanies a year-long course (120 hours) offered by the author himself at the University of London.


The stated goal of this book is for students to achieve ILR 3 proficiency. The textbook contains 50 short chapters centered upon grammatical topics and drills; however the order of introducing particular topics is questionable relative to other comparable materials. The text is meant to accompany classroom or one-on-one instruction with a native speaker. The vast majority of ‘exercises’ are simply sentences to read and digest, rather than activities that engage the student. There are, however, some exercises containing fill-in-the-blank and translation/conjugation. The authors state that students should endeavor to work with their instructor to design additional exercises, drills, and homework where necessary. The text loosely references conceptually similar materials in other Somali resources published by Dunwoody Press.

Dictionaries and Grammar


While this dictionary is written for Somali speakers learning English, it is also a helpful resource for English students of Somali. The dictionary contains a number of more contemporary vocabulary items not found in resources preceding it. Of note, 'dh' and 'kh' are not treated as separate letters in headwords, rather they occur alphabetically within pages for 'd' and 'k.' This dictionary also includes a number of illustrative examples and easy-to-follow conventions for indicating plural type and corresponding singular vs. plural agreement patterns.


This dictionary aims to provide what many Somali bilingual dictionaries lack, i.e., better representations of scientific and technical terms and a more up-to-date lexicon. The dictionary omits highly technical jargon, uncommon/rare derived forms, and synonyms. The dictionary provides many technical terms that are difficult to find in other Somali bilingual dictionaries, and when needed, the Somali translations are phrasal entries for English technical terms that are difficult to translate in just one word or simply do not exist as a single word in Somali. However, the dictionary also has its downside. For example, the author has excluded parts of speech, left out example sentences, and in many cases, lists a Somali verbal noun as the translation for an English imperative verb.


This general purpose Somali-English bilingual dictionary is perhaps the most widely used resource by Somali students. The beginning of the dictionary contains a brief outline of Somali grammar, including the abbreviations that are used in dictionary entries to indicate different grammatical categories. Entries are then comprised of the Somali headword, alternate forms (if applicable), the part of speech, the English sense, any cross references and dialectal or language borrowing information, if applicable. Although there is an English index, this dictionary is primarily Somali to English, not vice versa. Head words which are spelled the same way but have distinct meanings are listed under separate entries.

These resources are, for all intents and purposes, the same in that they cover the same topics in the same order. The difference between them is that the latter employs more linguistically-sophisticated descriptions of concepts and structures, and the theoretical and analytical assumptions that the author makes are more clearly articulated for an academic audience. Nonetheless, this grammar still suffers from fairly terse descriptions of a number of topics and does not widely draw upon other contemporary literature on the language. Despite its title, Saeed (1993) is suitable perhaps only for pedagogical use by beginning Somali language learners, as it offers a very cursory overview of key topics, while setting many others aside. In what appears to be an attempt at accommodating more native, non-linguist students, the author tends toward ill-suited terminology, mnemonic devices, incomplete glosses, and space-saving omission of examples which greatly decrease the utility of this resource.

PROPOSED LINGUISTIC CORRELATES OF PROFICIENCY

Linguistic correlates of proficiency (LCP), loosely defined, are linguistic structures and concepts that a learner must control to achieve a given level of language proficiency. These may include phonological and grammatical concepts, as well as a particular range of lexical items, collocations and idioms, and even different discourse registers such as formal vs. informal speech (Long, Gor & Jackson 2012). We are not aware of defined LCP for Somali in the published literature; however this state of affairs is far from unexpected, as such clearly defined correlates are largely lacking even for the vast majority of more commonly taught L2s (Long, Gor & Jackson 2012). A comparison of the sequence of topics in three often-used Somali textbooks, interviews with Somali instructors and students, and information obtained in three introductory Somali course syllabi confirms this finding.

It is beyond the scope of the current objective to investigate Somali LCP empirically; however we hope to do so in the future. In the meantime, drawing from interviews with Somali instructors, students, and our own expertise in Somali, we offer the non-empirical recommendations found in Appendix A as a first step to identifying Somali LCP. We take as a conceptual baseline the grammatical topics found in Zorc and Issa (2002); this is the only resource making an explicit claim that students should achieve ILR 3 upon completion of the course. We began by summarizing the functions defined in each ILR level in Reading and Listening, from 0+ to 3. We then mapped grammatical (i.e., linguistic) concepts and constructs necessary to perform these functions and tasks, ensuring that all topics covered in Zorc and Issa (2002) are represented. While the grammatical topics covered are, for all intents and purposes, identical between our approach and Zorc and Issa (2002), our approach differs in its explicit reference to ILR functions and necessarily its corresponding sequencing.

CHARACTERISTICS OF SOMALI

This section offers a simplified overview of Somali. For a more detailed summary of Somali characteristics that may present challenges to students, see Appendix B.

Orthography

Standard Somali is the language of primary education in Somalia and has a strong presence in news media, radio, and the internet. Most scholars consider Somali to have been 'standardized' in 1972 with the creation of a Roman or Latin orthography (Abdullahi, 2000; Andrzejewski, 1978). CASL’s ongoing corpus-based survey of
both pedagogical and news materials in Somali, however, reveals that spelling and other written conventions vary greatly and are anything but standardized. Somali also has a long tradition of having been written in a Perso-Arabic script.

The Romanized Somali orthography contains a number of characters that should be familiar to English speakers; however combinations of consonant letters that would appear uncommon or impossible in English are common in Somali. While most Somali sounds should not be problematic for native English speakers to hear or produce, Somali contains several sounds that may likely present more difficulties. These sounds and their phonetic equivalents (in brackets) are as follows: dh [d], q [q], ’ [ʔ], kh [x], x [ħ], c [ʕ]. Orthographically, ‘dh’ and ‘r’ are often interchangeable. The Somali orthography includes only five vowels; double (long) vowels and double (geminate) consonants are also indicated. Importantly, double vowels and consonants are lexically contrastive in Somali; they can encode a difference in meaning between words (e.g., dhige ‘writer’ vs. dhigee ‘to put a heavy load on something,’ and gadoon ‘rain pool’ vs. gaddoon ‘to overturn’). Some suffixes containing written ‘a’ and ‘e’ are interchangeable.

**Typological considerations**

**Tone and accent**

Like many world languages, differences in relative pitch, known more commonly as ‘tone,’ can encode meaning in Somali. Unlike many of the aforementioned languages, the use of tone in Somali is limited and carries only a small functional load. That is, contrastive tone most often differentiates between grammatical case (e.g., Subject vs. Non-Subject) in a subset of words, and is used in one particular class of nouns to distinguish between number (singular vs. plural) and gender (masculine vs. feminine). Tones, however, are not written in the Somali orthography.

Some scholars consider Somali tone to be a fairly predictable, static accent (Hyman, 1981); others propose that Somali words can contain, at most, a single High tone (Saeed, 1999). Others argue that this diagnostic is overly simplistic, failing to capture the actual function and distribution of tones in Somali (Green & Morrison, 2014).

**Word order**

Somali differs somewhat significantly from English and many other commonly taught second languages in its general lack of a standard sentential word order. While many languages reliably exhibit a fairly static subject-verb-object (SVO) or subject-object-verb (SOV) structure, for example, Somali sentences are much different. Somali sentences are based on a system wherein information that is ‘new’ to the discourse (i.e., in focus) is marked by one of three focus markers (baa, ayaa, and waxa). Depending on whether the subject or object of the sentence is ‘in focus,’ different inflectional patterns on the verb are also necessary. Because this information is encoded in such a way, a strict word order is not needed. Word order is not entirely free, however, as the order of constituents occurring within the Somali verbal piece (i.e., a phrase-like sequence containing object pronouns, adpositions, deictic (direction) markers, and the verb itself; Saeed, 1999) is largely static.

**Morphology**

Somali verbal and nominal morphology is highly agglutinative, i.e., words are most often comprised of several smaller pieces (morphemes) which combine to compose the sense or meaning of the word. As a fairly simple example, the Somali phrase nimanku ‘the men’ is, in fact, composed of four distinct morphemes: i) nin – the stem ‘man’; ii) the plural suffix, –an; iii) –ka, the masculine determiner; and iv) –u, a Subject marker. Verbs are similarly complex, as in the case of joogsaday ‘I stopped,’ which, itself, is composed of five unique morphemes: i) joog – the stem ‘stop’; ii) –is – the causative marker; iii) –at – the middle voice marker; iv) –t – first person agreement; and v) –ay – the first person past tense marker. Because of the complexities of Somali morphology,
there are many possible combinations and permutations available in the design of learning activities and evaluation tasks for a particular lesson module or objective.

**METHODS AND TECHNIQUES FOR SOMALI INSTRUCTION**

Approaches for Somali instruction should follow the general SLA principles and techniques outlined earlier in this report. That is, building an appropriate curriculum involves conducting an analysis to determine the learners’ specific needs for the language and then using this analysis to create pedagogical tasks that step students closer and closer to mastery of the language needed to fulfill their goals. Tasks should be designed that take into account complexity and sequencing appropriate to the students’ current proficiency level. Importantly, focus on noticing of grammatical structures and concepts should occur in the context of meaningful activities using, for example, frequent or domain-relevant vocabulary, so that learners acquire language features in context. The following table summarizes these recommendations as relevant to language instruction in a surge situation.

<table>
<thead>
<tr>
<th>Goal</th>
<th>Recommendation</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Improve student selection for surge courses</td>
<td>Consider student previous knowledge and experience</td>
<td>• Select students with background in a related language, demonstrated advanced proficiency in any L2, or heritage speakers; • Select students with higher language learning aptitude, similar domain knowledge.</td>
</tr>
<tr>
<td>Provide rich target language input</td>
<td>Select authentic texts</td>
<td>• Choose texts written by and for native speakers; • Make use of target language corpora; • Provide texts from a range of sources.</td>
</tr>
<tr>
<td>Provide comprehensible input</td>
<td>Consider student’s current proficiency level when selecting texts</td>
<td>• Select target texts for appropriate complexity and difficulty; • Use text-leveling when possible.</td>
</tr>
<tr>
<td>Elaborate texts</td>
<td></td>
<td>• Build in redundancies to the original text (e.g., add synonyms, explanations, paraphrasing); • Incorporate interactive multiple choice glosses or textual and pictorial glosses.</td>
</tr>
<tr>
<td>Occasionally focus on form</td>
<td></td>
<td>• Make lexical and grammatical structures salient; • Use input enhancement (e.g., highlighting, underlining, graphics, animation, font size/color, audio repetition); • Incorporate inductive or deductive activities for focus on the structures of the language, especially for complex morphology and syntactic patterns.</td>
</tr>
<tr>
<td>Utilize resources to aid comprehension</td>
<td></td>
<td>• Dictionaries, glosses, pictures for reading tasks; • Utilize tools such as DYM, parser, concordancer; • Electronic gloss/annotation to electronic dictionaries; • Grammatical information and paradigms as supplementary tools.</td>
</tr>
<tr>
<td>Engage learners in interactive tasks using the target language</td>
<td>Use tasks, not texts, as the unit of analysis</td>
<td>• Use task-based language instruction; • Incorporate simulations, virtual worlds, tutorials.</td>
</tr>
</tbody>
</table>
| Provide opportunities for practice in the target | | • Engage learners in production, not only comprehension checks of the input, e.g., use open-
The goal of language training in a surge scenario is to rapidly prepare learners to deal with the demands of the mission, and this is most effectively accomplished by focusing instruction on the language and skills needed to perform target language tasks rather than trying to exhaustively cover the linguistic features of the language in a decontextualized fashion. Therefore, the organizing unit for the curriculum should be real-world language tasks and mission-relevant content that the learners will encounter post-training. This may include appropriate use of job aids and language tools necessary to augment their language ability and fulfill the mission at a high level.

In addition, as noted by DeKeyser (2003), there is no reason to believe that some attention to explicit learning of target language features would not be beneficial in a language learning context as well. The following are examples of activities and linguistic aids that would draw learners’ attention to relevant, potentially difficult, aspects of Somali through the use of CASL tools while engaging in meaningful content. For details on these challenging linguistic aspects of Somali, see Appendix B.
Sample activities incorporating CASL HLT tools

In this section we suggest how CASL HLT tools can be used for various pedagogic purposes such as: supplementing and providing feedback on linguistic challenges that arise in the course of performing language tasks, creating games for exploring and building vocabulary, and aiding learners’ text comprehension by providing automated annotations. These are not meant to be exhaustive but illustrative of the myriad uses for tools that can be incorporated into LanguageNation to facilitate learning.

Supplementing task-based instruction

One target language need that learners may have is interpreting and reporting details of an event, such as a political demonstration or a natural disaster. One pedagogic task learners could engage in is telling a story with the use of supporting media, such as pictures or video. For example, learners might first be presented with pictures that illustrate actions and the corresponding Somali text. Learners could view and read the text initially then be asked to sequence events, or match pictures and texts. In a more interactive iteration of the activity, learners could collaborate to piece together a story based on the information. Another variation on a collaborative activity would be that learners each have different information about what happened, and must find out the missing pieces from their partner (whether the computer, another learner, or a coach) in order to figure out and reconstruct the sequence of events.

System-generated feedback from CASL HLT tools

One particular challenge that may arise in this kind of activity is identifying the subject and object of the action. This subject or object status is encoded in Somali by two factors: i) the presence of a focus marker, and ii) verb agreement. That is, the location of a focus marker (after the subject or after the object) in the sentence and the agreement patterns that this placement triggers on the verb are key indicators of the relationship of the arguments to the verb. The complexity of this interaction is demonstrated in the following activity where one could expect a high rate of error given the [-few elements, - independency of steps] nature of this relationship.

THE MEN brought some tea.

1) Using the word ba, mark the word ‘in focus’
2) Select an ending for ba
3) Select an ending for the verb

Nimankii shaah keen.

Nimankii ba
   -a
   -y
   -fi

System-generated, or in the case of the collaborative activities, user-generated, feedback would likely be necessary to target errors in basic understanding of agent-patient relationships, i.e., ‘who did what to whom.’ The correct answer could be encoded into the system at the same time the questions themselves are developed, and the students can thus receive feedback on their answers. Alternatively, the correct answer can be automatically determined through use of the morphological parser alongside a syntactic parser. CASL is currently investigating the functionality of an already developed Somali syntactic parser for applicability to the learning platform. Ideally, for the first error, the system would be able to start with implicit feedback, such as providing the user with the correct answer, without resorting to a grammatical explanation. Upon subsequent
errors of similar types, the system would ideally be able to draw on more explicit grammatical resources via CASL’s morphological or syntactic parser.

Deciphering complex clusters and their referents

Another challenging aspect of Somali grammar that may arise for feedback in this kind of task is adposition placement and use. The platform could offer GIFs or simple animations depicting short scenarios or simple actions/movements requiring adposition use. Such exercises could employ simple locative phrases (Waan gurigayga ku akhriyaa/ ‘I am reading in my house.’) requiring a single adposition at basic levels, as in the activity below. From a receptive standpoint, sentences containing adpositions and deictics could be supplied (written or aural) to the learner, requiring the learner to use contextual and grammatical clues to match a morpheme to its referent.

For higher-level learners the event would require combinations of multiple adpositions, forming an adpositional cluster. Further levels of complexity could include deictic markers. This activity tests knowledge of word/morpheme order in clusters, which must be mastered to convey meaning accurately. Feedback could be provided by a syntactic parser to check the grammaticality of the sentence as a whole and by the CASL Did You Mean…? (DYM) tool to verify accurate spelling and to suggest alternatives to spelling errors.

Rich feedback-on-error from tagged corpora

Students face additional challenges in learning even the most common lexical items and phrase types, as Somali is a highly agglutinating language where a single word may contain upwards of four or five meaningful constituents. By relying on system feedback and drawing on feature-rich tagged corpora, lessons targeting problem areas can be generated. CASL’s Somali Morphologically-aware Did You Mean (MaDYM) can spot student errors, suggest the correct alternative, and subsequently analyze words into their constituent parts; sentences in CASL’s corpora containing similar words/sequences of morphemes are then supplied in succeeding lessons. An example is given below how this technology might be utilized to give feedback on a learner’s error.

In learning to describe items in a picture, the student must select an appropriate description of the house. In (1), in order to answer correctly, the student must recognize the vocabulary item, recognize the appropriate plural pattern, and also ensure that the word has subject marking. The student incorrectly chooses a description without subject marking. In (2), the correct word is identified by the system and its morphemes are delimited by the parser in (3). In (4), the corpus returns samples of other words with the same form that can populate additional lessons or practice. This kind of feedback might be useful in conjunction with a meaning-based activity with plenty of examples of the target structure.
Creating vocabulary-building games

Validated vocabulary learning via Did You Mean…?

Vocabulary assessment can be enhanced via CASL’s fuzzy search Did you Mean...? tool. Similar to Hasbro’s The Game of Scattergories, as schematized below, students click to roll an alphabetical die which lands on a letter. Students must provide a word(s) beginning with the targeted letter for every descriptor given. Upon completion, the DYM validates the responses, and fuzzy search suggestions are provided for any incorrect responses.

Ideally, the words targeted would be recycled from other tasks and activities in the learner module, with the purpose of assisting learners to consolidate words that they have been producing or interpreting incorrectly in other activities. Furthermore, this kind of activity would be best complemented by exposure to these target words in context, for example, in a newspaper article. Ideally, the words would be conceptually related to the newspaper article. Learners could either play this game prior to reading for meaning in order to prime vocabulary knowledge, or they could play after reading to draw their attention to words and concepts that arose in the text. The DYM tool could also be used in other activities throughout the module to give feedback and suggestions to learners on their written production.
While completing reading tasks, students often encounter a great deal of new vocabulary or vocabulary over which they have only partial control. Students can click to select unfamiliar words as they proceed through a passage, with either the option of just selecting the word to add to their vocabulary list or turning on the annotated text option where the L2 definition can be obtained by hovering over the word. Once the student has finished the passage, he or she can consult the compiled personalized self-study vocabulary list. With access to the electronic dictionary and parser, the student can choose to have the generated list contain only dictionary forms (the roots of the words in the passage) or one that includes the parse as well. As new words or word forms are acquired, the students can remove items from the list. These words can also be stored in learners’ personal vocabulary notebooks which can be later revisited and updated by the learners with additional information such as example sentences, synonyms, and information on pronunciation.

Vocabulary practice game

Students can participate in a “20 Questions”-like game as a solitary player against the computer or in simultaneous online play with other students. This game can be used to reach ILR objectives for various levels, but in the example given here, it is applied to ILR Level 1 and 1+ (descriptions of people, places, and things).

For solitary play, a vocabulary list (primarily nouns) can be stored along with their values with respect to a predetermined set of questions. For example, the vocabulary list for the solution to the game might include the word “poem.” The predetermined question set can be largely overlapping for all of the possible solutions and would include questions like “Q1: is it a person,” where the value for “poem” has been stored as “Q1:No.” The predetermined questions can be organized on screen by broad category, and the student will expand out each category to select the question he wants to ask the computer. There can be the option of either unconstrained guessing on the part of the student (so that the student must know the word in advance, making the exercise vocabulary practice) or the student can be supplied with a list of possible answers which includes the correct answer to the round (so that the student need not know the word in advance, but may have to look up some words from the list in order to guess the answer, making the exercise vocabulary learning). Solitary play is most easily implemented for ILR Level 1.

For online simultaneous play with 2 or more students, the computer can provide to one student one word from a list of suitable vocabulary. The course developers can either decide in advance which words are most appropriate for which student proficiency level or this information can be roughly obtained using the word frequency list. The student guessers can then use a platform, like online chat, to ask the student who knows the answer questions about it to guess what it is. Online simultaneous play can be applied to Level 1+ or higher.

Comprehension aids using text annotation

Many activities will naturally require the use of text, which may need to be modified to the student’s current level. One way in which CASL’s morphological parser/generator can be useful in making the text comprehensible is by allowing a learner to see an annotation of component parts of a word in a text. For example, the generator can be used such that upon hovering over highlighted new or targeted words in a string of text, the student is presented with these pertinent pieces of information. This technique would be useful for a more advanced learner who is already familiar with many of the morphological components of Somali. Allowing users this access to more detailed information might be especially helpful, for example, in understanding the complex noun attributes in Somali, where a student must know not only the noun’s meaning, but also the agreement pattern that it follows and the way in which it forms its plural, in order to understand its meaning. An example follows:
CASL’s Somali parser can also be used for text annotations in aid of exposing the student to i) words sharing morphological and phonological characteristics with a target word, or alternatively ii) additional examples within the word family of a target word, as in the following example.

This is applicable to various ILR Levels and objectives, but will be exemplified here for ILR Level 2. One of the Level 2 objectives is to be able to “describe frequently occurring events, simple biographical information, social and technical material.” A corresponding linguistic correlate of proficiency for this objective is having command over all major verb conjugations. As the student reads through practice texts, he can hover over content that he does not understand; the parser expands the word to show its morphological breakdown. The computer can thus lay out for the student explicitly what order verbal morphemes may appear in. Further, if there is one morpheme in particular the student doesn’t know or understand, he can select that part of the parse and ask the computer to give him more words that use that morpheme and provide them in context using Somali corpora. Some of these examples may come from verbs in a bilingual corpus, giving the student help with the English translation of the Somali morphology. In addition, the computer can inform the student about relevant sound changes that occur when morphemes come together, as this information is also encoded in the Somali parser. Similar automated instruction, but focusing on learner errors, is discussed in Antonsen, Huhmarniemi, and Trosterud (2009).

CONCLUDING REMARKS

In this report, we have outlined recommended methods and techniques grounded in contemporary, well-established SLA principles for the objective of ‘surge’ Somali language acquisition. This objective aims to take absolute beginners to general working proficiency at an expedited pace. While there is little published literature on ‘surge’ acquisition, and likewise little information on the particular science behind the acquisition of Somali, SLA research suggests that the principles described above, coupled with a task-based curriculum and supplemented with CASL-developed Human Language Technology tools, will yield positive results for this learning scenario.

Given the complexity of Somali, we hypothesize that these results can be further bolstered by student selection recommendations stated above, e.g., individuals with empirically-suggested aptitude for language learning, previous (heritage) exposure to the language, demonstrable success in L2 language learning, and domain-specific knowledge related to the target needs or mission. Also related to target needs and mission, we
recommend the development of a task-based curriculum drawing on, approximating, or otherwise incorporating such real-world material and objectives. In doing so, students are exposed to subject matter and tasks that will immediately serve their target acquisition needs. Again, due to the considerable linguistic complexity exhibited by Somali, we recommend that CASL HLT tools and corpora be incorporated into the curriculum and the LanguageNation learning platform in a variety of ways, e.g., facilitating error detection, providing feedback on error, and tailoring supplemental activities to computationally-determined lexical or grammatical problem areas.

REFERENCES


Appendix A: Proposed Linguistic Correlates of Proficiency for Somali

ILR 0+: Memorized Proficiency

ILR summary:
Sufficient comprehension to understand a number of memorized utterances in areas of immediate need

ILR objectives:
i) can recognize letters and sounds;
ii) recognize isolated words and phrase;
iii) recognize personal and place names;
iv) familiar with vocabulary related to business, food, time

LCP Somali summary:
Students should have command over sounds, orthography, and numbers 1-20; ability to identify common conversational openings and closings; construct declarative and equative sentences using basic vocabulary; form and answer yes/no questions

LCP Somali concepts:
i) sounds and orthography, including spelling variation and common sound changes;
ii) common greetings and benedictions;
iii) numbers 1-20, dates, months, telling time;
iv) imperative of weak (suffixing) verbs (these are citation forms and are used in greetings and commands);
v) declarative/equative sentences using waa and corresponding yes/no questions using ma;
vi) independent pronouns (function as both subjects and objects outside the verbal piece);
vii) common names, cities, regions, countries

ILR 1: Elementary Proficiency

ILR summary:
Able to comprehend simple connected materials, recombine known vocabulary to achieve new meaning, identify simple subject matter in authentic language

ILR objectives:
i) comprehend formulaic exchanges with high frequency structures;
ii) identify shared vocabulary and cognates;
iii) basic descriptions of people, places, and things;
iv) familiarity with geographic locations and directions;

LCP Somali summary:
Students should have command over most present tense verb forms; ability to recognize common borrowed words from both Arabic and English; should be able to form simple nouns phrases containing attributive adjectives, including familiarity with marking grammatical gender with definite determiners and demonstratives

LCP Somali concepts:
i) affirmative and negative simple and progressive present tense and yahay ‘to be’ and leeyahay ‘to have,’ all present tense forms of non-derived weak (suffixing) verbs;
ii) Arabic and English borrowings;
iii) nouns pluralized by -o, agreement with definite determiners and demonstratives;
iv) attributive adjectives;
v) cardinal directions and locatives, numbers 20-100;
vi) simple conjunctions

**ILR 1+: Elementary Proficiency Plus**

ILR summary:
Able to understand simple discourse

ILR objectives:
i) comprehend simple discourse, announcements, and biographical information;
ii) understand temporal reference;
iii) understand more sophisticated description of people, places, and events

LCP Somali summary:
Students should have command over present tenses for all verb types and begin to describe actions/states using temporal and locative adverbs and basic adpositions; ability to recognize and use all noun classes and recognize their patterns of agreement, including use with possessive determiners; construct more complex noun phrases with adjectival relative clauses

LCP Somali concepts:
i) affirmative and negative present progressive; present tense forms of four main strong verbs
ii) other plural nouns, agreement with definite determiners and demonstratives;
iii) possessive determiners
iv) present tense of causative (-is) and middle (-o,-an,-at) verbs
v) temporal and locative adverbs
vi) adjectival relative clauses, short forms of *yahay* and *leeyahay*
vii) basic adpositions (*ku, la, ka, u*); comparative and superlatives

**ILR 2: Limited Working Proficiency**

ILR summary:
Sufficient comprehension to understand simple, authentic material and straightforward, factual material

ILR objectives:
i) describe frequently occurring events, simple biographical information, social and technical material;
ii) familiarity with high-frequency sentence patterns;
iii) understand factual questions about authentic texts about familiar people, places, and events;
iv) express feelings, wishes, and future plans

LCP Somali summary:
Students should have command over all major verb conjugations, including reduced and auxiliary forms; ability to construct sentences containing focus markers and understand their use, restrictions, and unique characteristics; manipulate particles within the *verbal piece*; begin constructing simple sentences with subordination

LCP Somali concepts:
i) past tense forms of all weak (suffixing) verbs, *yahay* and *leeyahay*;
ii) focus constructions, alongside subject and object pronouns, subject case-marking;
iii) impersonal and reciprocal pronouns;
iv) ‘hybrid’ verbs, i.e., adjectives using *yahay*;
v) adposition clusters and deictics;
vi) the infinitive and auxiliary constructions (conditional, future, potential), optatives;
vii) subordinate clauses using in-, alongside reduced verb forms;
viii) strong verbs in past and present tenses

**ILR 2+: Limited Working Proficiency Plus**

ILR summary:
Understand most factual material in non-technical prose and concrete topics

ILR objectives:
i) separate main ideas from lesser details;
ii) use linguistic and contextual clues to make sensible guesses about unfamiliar material

LCP Somali summary:
Students should begin to manipulate thoughts, ideas, opinions, and descriptions using more complex syntactic constructions and adverbial phrases; recognize constructions utilizing vocative and genitive case

LCP Somali concepts:
i) other constructions containing relative and subordinate clauses;
ii) additional suffixing/tone changing vocative and genitive constructions;

**ILR 3: General Professional Proficiency**

ILR summary:
Read with almost complete comprehension on a variety of subject; read ‘between the lines’

ILR objectives:
i) understand reported text, routine correspondence, and most technical material;
ii) hypothesize, generalize, and support opinions;
iii) read ‘between the lines,’ detect subtlety and nuance

LCP Somali summary:
Students should have full command over sentence formation and should begin to manipulate syntax to form recursive, embedded sentences; should understand the principles and patterns behind common ways to derive new words; linguistic and cultural knowledge to parse and use common idioms to encode pragmatic intent

LCP Somali concepts:
i) derived nouns and verb, compound formation;
ii) conjoining clauses with oo and ee;
iii) common idioms;
iv) complex adverbial constructions;
v) interactional pragmatics, illocution
Appendix B: Challenging linguistic characteristics of Somali

While students likely face challenges when learning any language, certain languages present particular difficulties when target language structures and/or vocabulary differ significantly from their own L1 or well-established L2. Based upon a grammatical comparison to English and other commonly taught second languages, and informed by students and instructors of Somali, we propose that the following characteristics may prove especially challenging to rapid-rise learners of Somali.

**Vowel deletion**

Most Somali words contain two or more syllables. It is common in longer words, however, for a vowel to be deleted from a word when certain conditions are met. This deletion will obscure the citation form of a word, hindering dictionary lookup. An illustrative example is the word *orday* ‘I ran.’ This verb is derived from the stem *orod*; however when the past tense ending –ay is added, a vowel is deleted. In other instances, vowel deletion also triggers a sound change. For example, the verb *arag* ‘to see’ becomes *arkay* ‘he saw;’ there is both a vowel deletion and a subsequent sound change. Other types of sound changes are described below.

**Sound changes**

Because Somali words are highly agglutinating, and word formation tends to involve the addition of one or more affixes to a stem, a number of predictable sound changes result. In many instances, Somali consonants change according to their place in a word or their proximity to another consonant. Like vowel deletion, sound changes obscure the citation form of a word, thereby presenting a challenge to beginning language learners. Examples of some sound changes encountered in Somali include:

- Simple alternation:  
  - na+ka → naga ‘from us’
  - keen+ay+taa → keenaysaa ‘she is bringing’

- Complex alternation:  
  - meel+ta → meesha ‘the place’
  - maalmo+ka → maalmaha ‘the days’

- Assimilation:  
  - bad+ta → badda ‘the ocean’
  - fur+niin → furriin ‘divorce’
  - dil+n+ay → dillay ‘we killed’

**Vowel harmony**

Written Somali has five short and five long (double) vowels; however the spoken language has twice this number of vowels due to a contrastive feature of the language known as *vowel harmony*. This means that there are two sets, or series, of Somali vowels that are pronounced differently. In one series, vowels are pronounced with the tongue slightly forward in the mouth, while in the other, the tongue is further back in the mouth; think of the difference in *beat* vs. *bit* or *bait* vs. *bet*, in English. The vowels in the citation form of each Somali word are from one series or the other and often change series when certain suffixes are added. While this will not present a challenge for students when reading or writing Somali, it may negatively affect listening comprehension and speaking.

**Noun classes and agreement**

Somali nouns are divided into classes or groups based on notable characteristics that they display. While the precise details defining each noun class have long been debated among scholars (Andrzejewski, 1979; Banti, 1988; Hyman, 1981; Lampitelli, 2011; Le Gac, 2003), most scholars generally agree that nouns can be described with reference to two particular features: i) the way that a noun forms its plural, and ii) its pattern of *agreement*. 
Somali nouns form their plurals in three main ways:

**Suffixation:**
- naag ‘woman’ → naago ‘women’
- baabuur ‘truck’ → baabuurro ‘trucks’
- hilib ‘meat’ → hilbo ‘meats’
- magaaloo ‘town’ → magaaloo ‘towns’

**Reduplication:**
- buug ‘book’ → buugag ‘books’
- san ‘nose’ → sanan ‘noses’

**Tone change:**
- díbi ‘bull’ → dibí ‘bulls’
- Soomaalí ‘Somali’ → Soomaalí ‘Somalis’

The way that a Somali noun forms its plural is intimately linked to its pattern of *agreement*, sometimes referred to as *grammatical gender*. Grammatical gender determines the form of words (e.g., determiners and possessives) that agree with a particular noun. Many commonly taught languages, like Spanish and French, manifest grammatical gender in the type of determiner ‘the’ and the form of adjective that agrees with a particular noun. For example, in Spanish, *el perro blanco* ‘the white dog’ vs. *la vaca blanca* ‘the white cow.’

Agreement in Somali functions similarly, but it is far more complex than in languages like Spanish or French. The complexity lies in the fact that the agreement patterns associated with Somali nouns change (and not always in the same direction) when a noun is pluralized. Thus, a student of Somali must learn the vocabulary item, its method of plural formation, and two patterns of agreement. The following table illustrates this system with an exemplar from each of the five main Somali noun classes.

<table>
<thead>
<tr>
<th>Class 1</th>
<th>kab</th>
<th>‘shoe’</th>
<th>kabta</th>
<th>‘the shoe’</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 2</td>
<td>sannad</td>
<td>‘year’</td>
<td>sannadka</td>
<td>‘the year’</td>
</tr>
<tr>
<td></td>
<td>sannaddo</td>
<td>‘years’</td>
<td>sannadda</td>
<td>‘the years’</td>
</tr>
<tr>
<td>Class 3</td>
<td>hilib</td>
<td>‘meat’</td>
<td>hilbka</td>
<td>‘the meat’</td>
</tr>
<tr>
<td></td>
<td>hilbo</td>
<td>‘meats’</td>
<td>hilbaa</td>
<td>‘the meats’</td>
</tr>
<tr>
<td>Class 4</td>
<td>af</td>
<td>‘language’</td>
<td>afka</td>
<td>‘the language’</td>
</tr>
<tr>
<td></td>
<td>afaf</td>
<td>‘languages’</td>
<td>afafka</td>
<td>‘the languages’</td>
</tr>
<tr>
<td>Class 5</td>
<td>mädax</td>
<td>‘head’</td>
<td>mädaxa</td>
<td>‘the head’</td>
</tr>
<tr>
<td></td>
<td>madáx</td>
<td>‘heads’</td>
<td>madáxda</td>
<td>‘the heads’</td>
</tr>
</tbody>
</table>

The agreement patterns for Somali noun classes are summarized in Table 3.

<table>
<thead>
<tr>
<th>Class 1</th>
<th>T-/D-</th>
<th>K-/G-</th>
<th>Suffixation (-o/-yo)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class 2</td>
<td>K-/G-</td>
<td>T-/D-</td>
<td>Suffixation (-Co/-yo)</td>
</tr>
<tr>
<td>Class 3</td>
<td>K-/G-</td>
<td>T-/D-</td>
<td>Suffixation (-o/-yo)</td>
</tr>
<tr>
<td>Class 4</td>
<td>K-/G-</td>
<td>K-/G-</td>
<td>Reduplication</td>
</tr>
<tr>
<td>Class 5</td>
<td>K-/G-</td>
<td>K-/G-</td>
<td>Tone shift</td>
</tr>
</tbody>
</table>
Focus

As mentioned above, Somali word order is generally loose; the order of constituents like the subject and object are not fixed, rather they are influenced by a system of grammatical focus marking. Nearly every Somali sentence requires a focus marker; those that do not instead require a declarative or interrogative marker. Focus markers follow a word or phrase that is ‘in focus,’ meaning that it is new information to the discourse or is being emphasized. A focused word or phrase can either precede or follow the more static verbal piece. Some research suggests that the exact placement of a focused word or phrase in a sentence yields subtle differences in meaning (Gebert, 1986). The following sentences illustrate general function of focus in Somali; the focus marker in each sentence is underlined.

Ninkii baa gaadhigii cusbaa watay. ‘THE MAN drove the new car.’
man-the FOC car-the new drive-3SM-PST

Ninkii gaadhigii cusbaa buu watay. ‘The man drove THE NEW CAR.’
man-the car-the new FOC-3SM drive-3SM-PST

The concept of marking new information as ‘in focus’ is not itself difficult, rather the challenge lies in other structural and grammatical requirements that focus marking triggers. First, when the subject of a sentence is in focus (as in the first example), a bare focus marker is used; however when the object of a sentence is in focus (as in the second example), the focus marker bears a subject pronoun.

Second, overt subject case marking only occurs when the subject of a sentence is not in focus, as in the first example below. When the subject of the sentence is in focus (as in the second example), it cannot be subject marked.

Ninku wuu imanayaa ‘The man comes.’
man-the.SUBJ dec-3sm come-3sm-pres

Ninka ayaa imanaya. ‘THE MAN comes.’
man-the FOC come-3SM-PRES

Lastly, when the subject of a sentence is focused, the verb of the sentence must occur in what has come to be known as the reduced paradigm. The examples below illustrate such a situation where, in the first sentence, keeneen has 3rd person plural agreement with its subject ‘the men’; however because the subject of the second sentence is now in focus, agreement marking on the verb is entirely different.

Nimankii way keeneen. ‘The men brought it.’
man-PL-the DEC-3PL bring-3PL-PST

Nimankii ayaa keenay. ‘THE MEN brought it.’
man-PL-the FOC bring-3PL-RED

In addition to the types of focus marking described just above, Somali very commonly uses another type of focus that creates cleft sentences. An example in English would be in a comparison between ‘The child wanted the chocolate bar,’ vs. ‘It is THE CHILD that wanted the chocolate bar.’ The latter sentence places emphasis or focus on the subject. In Somali, the constituent (subject or object) focused by a cleft always appears after the verb.

Shaah baan doonayaa. ‘I want SOME TEA.’

Adpositions

Cross-linguistically, the acquisition of adpositions (i.e., prepositions and postpositions) presents particular difficulty for L2 learners. From a linguistic perspective, Somali adpositions represent one of the most challenging aspects of the language’s grammar. One key factor contributing to this challenge is that although Somali adposition particles are located in a fairly static slot within the verbal piece, they are not collocated with their referent. In addition, when multiple adpositions are found in a single verbal piece, they form clusters not only with one another, but also with other morphemes like object pronouns, as discussed further below. The following examples illustrate that adpositions do not necessarily co-occur with their referents. Note that in the second example, the 3rd person singular has no overt form, contributing yet another challenge for learners to tackle.

Pronouns and coalescence

Because Somali is highly agglutinating, it is extremely common to encounter words comprised of many individual pieces. It was introduced above that both nouns and verbs are comprised of several components; however still other Somali words are composed via coalescence of smaller morphemes, several of which are found in examples given above.

A common type of coalescence is found with Somali subject pronouns. These pronouns seldom stand alone, rather they coalesce with other sentence components, such as focus markers, declarative markers, and question markers, among others. In the sentences just above, the focus markers baa and waxa coalesce with the 1st person singular subject pronoun aan to create baan and waxaan, respectively. More complex are instances when the focus marker itself coalesces with a word preceding it to form even larger words, as in the following example where the word maxaydin is comprised of the question marker maxa, the focus marker baa, and the subject pronoun idin.

Subordinate clause structure

Subordinate clauses (including relative clauses) are widespread in Somali, and yet the way that they are formed is challenging for several reasons. First, verbs in subordinate clauses have a special set of endings, similar but
not identical to the reduced paradigm verb forms found with subject focus. Second, subordinate clauses never contain a declarative or focus marker. Third, both relative and subordinate clauses can act like the subject of a sentence; the final word in such a clause may be subject case-marked, even if the final word is a verb. Fourth, subordinate clauses have a special negative marker, aan. Lastly (and pertaining exclusively to relative clauses), relative clauses are used descriptively in Somali where an adjective might be used in another language. For example, one cannot simply say ‘a valuable thing’ in Somali, rather one must say ‘a thing that has value.’ In addition, relative clauses are used in counting and in expressing some types of possession.