



CENTER FOR ADVANCED
STUDY OF LANGUAGE

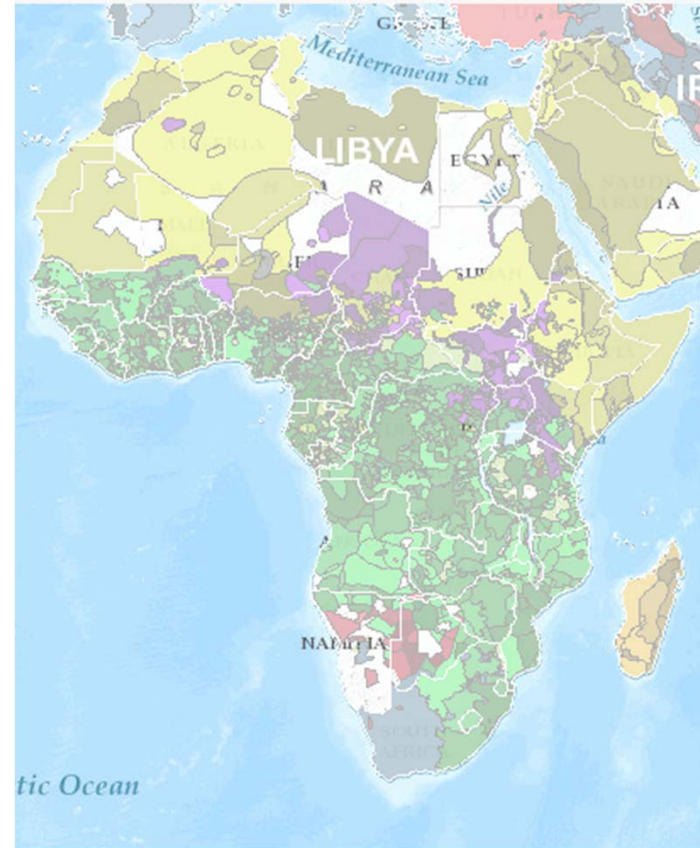
Finding, evaluating, and adapting learning resources for critical African languages

Christopher R. Green

*Center for Advanced Study of Language (CASL)
University of Maryland*

The Africa situation:

- Overall challenges:
 - 54 countries, ~2,100 languages, and countless other dialects and emerging languages...not to mention widespread multilingualism
 - Even the most widely spoken African language are poorly resourced
 - Sparseness of high-quality reference, pedagogical, and learning materials can have a negative impact on diplomacy, defense, and intelligence work



The Africa situation, cont'd.

- Challenges related to learning resources:
 - Resources and materials uncovered on the Internet can provide a skewed, incomplete picture of the quantity/quality of what is 'out there'
 - Developing contemporary materials from scratch takes years
 - Adapting and enhancing available materials can cut time and some expense, yet still requires specialized expertise

Scenario #1

- A need for expertise in *(insert your favorite language here)* arises
- Not even sure what language it is?
 - Geolinguistics/language databases
 - Human Language Technology
 - Ask your friendly neighborhood linguist
- This is a topic for another day



Scenario #2

- Assuming the language needed is known:
 - What is the response timeline?
 - We needed it yesterday...
 - What level of expertise is needed?
 - That depends...we'll get back to you.
 - Who do we have that can assist?
 - That also depends...we'll get back to you.
 - What materials are available?
 - Quick!...Search the internet!



The results are...?

There's a dictionary of Language X on the internet...

...we're all set, right?!



The dangers of search engines

- Search algorithms are based on terms and/or combinations of terms
 - Results provide little/no information on contemporaneity, quality, relevance, authorship, breadth, depth, etc.

dictionary of oromo

Web News Images Shopping Videos More Search tools

About 262,000 results (0.36 seconds)

Oromo Dictionary
oromodictionary.com/ ▾
The Oromo Dictionary Application can be downloaded and installed to the android phones from Google Market. Oromo Dictionary is available in Android Market ...

Translation Oromo-English-Oromo - FREELANG online ...
www.freelang.net/online/oromo.php ▾
Jan 8, 2014 - Online version of Freelang's Oromo-English dictionary and English-Oromo dictionary.

Free English Oromo dictionary and translator - Android or PC
www.freelang.net/dictionary/oromo.php ▾
Oct 15, 2007 - Oromo-English-Oromo dictionary to download for free (PC and phone) or look up online.

Oromo Dictionary - Android Apps on Google Play
https://play.google.com/store/apps/details?id=com.oromo.dictionary... ▾
If you want to learn Afaan Oromo or English, or you want to extend your knowledge, then the Oromo Dictionary is the ideal tool to help you understand and ...





English-Oromo-Amharic dictionary = Gaimee jechoota afaan ...
searchworks.stanford.edu/view/7942370 ▾
English-Oromo-Amharic dictionary = Gaimee jechoota afaan Ingilizi-Oromoofi-Amaaraa = ya'Englizefiā·Oromeñā·Amāreñā mazgaba-qālāt. Author/Creator ...

OROMO DICTIONARIES
oromodictionaries.com/ ▾
Afaanota 14, hiikkaa fi akkatti dubbifamu sagaleen caqasii barit English Amharic Arabic Chinese Dutch French German Hindi Italian Norwegian Somali Spanish ...

Oromo | Define Oromo at Dictionary.com
dictionary.reference.com/browse/oromo ▾ Dictionary.com ▾
Oromo definition, a member of a nomadic, pastoral people of Ethiopia and Kenya. See more.

English – Oromo Dictionary | Oromia & Horn of Africa News
https://oromianews.wordpress.com/english-oromo-borana-dictionary/ ▾
Full text of "English-Borana Word List" Compiled by Rev. Vincent G. Stanman, C.S.S.

Shop for dictionary of oromo or

 Oromo Children's Dictionary \$14.63 BookDeposito...	 Modern Oromo Dictionary: Or... \$13.87 eBay
 My First Oromo Dictionary: Co... \$13.62 Barnes & Noble	 Oromo Children's Dict... \$14.38 Barnes & Noble

Dangers of search engines, cont'd.

- Search algorithms rely on web presence
 - Some of the best materials may not be readily available, under restricted use (university, copyright, patent, license, etc.)

21. [MIDDLE VOICE VERB FORMS IN EASTERN CUSHITIC](#)
by [R. J. Hayward](#) Affiliation:
Article 1975 | Peer-reviewed
Transactions of the Philological Society v74 n1 (November 1975): 203-224
Database: WorldCat.org
Held by: USMAI Libraries

22. [Oromo dictionary](#)
by [Gene B. Gragg](#), [Michigan State University](#), [African Studies Center.](#), [Thomas Leiper Kane Collection \(Lib...](#)
Print book ©1982
Database: WorldCat
Held by: Committee on Institutional Cooperation (CIC)

23. [Hamid Muudee's Oromo dictionary : English-Oromo](#)
by [Mahdi Hamid Muudee](#)
Print book ©1995-
Database: WorldCat
Held by: Committee on Institutional Cooperation (CIC)

24. [Oromo-English dictionary](#)
by [Tilahun Gamta](#), [Thomas Leiper Kane Collection \(Library of Congress Hebraic Section\)](#)
Print book [1989]
Database: WorldCat
Held by: Committee on Institutional Cooperation (CIC)

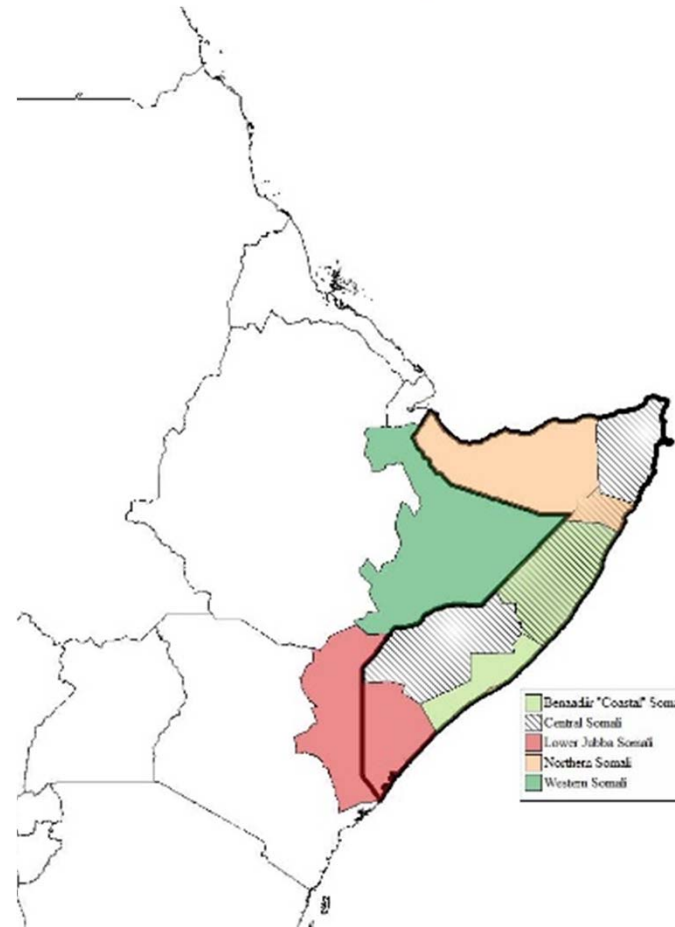
If materials ARE available:

- Common concerns:
 - Too limited in scope (e.g. missionary materials, travel guides, only address the 'standard')
 - Too far in depth for the average student (e.g. academic articles, dissertations)
 - Questionable source (e.g. student project)
 - Outdated/inconsistent (e.g. no work on the particular language for years)



Case in point: Somali

- An enduring, geopolitically important language
- Comparatively well-documented for an African language
- BUT...resources are severely limited in scope, coverage, and contemporaneity



The bad news...

- ‘One’ out-of-print reference grammar with errors and major gaps
- Somali-English dictionary with outdated entries, minimal dialect info
- No complete English-Somali dictionary
- Two largely grammar-driven textbooks
- Relatively little contemporary research on the language



More bad news...

- Somali is only semi-standardized, creating difficulties for Human Language Technology development
- Grammatical and pedagogical resources also adopt different descriptive terminology due to research biases



The better news...

- Fantastic Somali-Somali dictionary (but requires that Somali is already known)
- Rich academic literature (but requires deep knowledge of linguistics)
- Strong internet presence (but, again, requires that Somali is already known)
- Pockets of new research are emerging in US and Europe



One response...LanguageNation

- Federally-funded joint venture between the IBM Watson Research Group and University of Maryland – CASL
- Create a language learning platform for bringing students from ILR 0 to 3 in Somali
- Informed by research expertise in linguistics, second language acquisition, psycholinguistics, computer science, computational linguistics



CASL contributions

- Research on Somali grammar and dialectology
- Written and audio corpus development
- Electronic dictionary repair, fuzzy search dictionary, morphological parsing
- Learning plan and activities informed by and grounded in SLA principles



IBM contributions

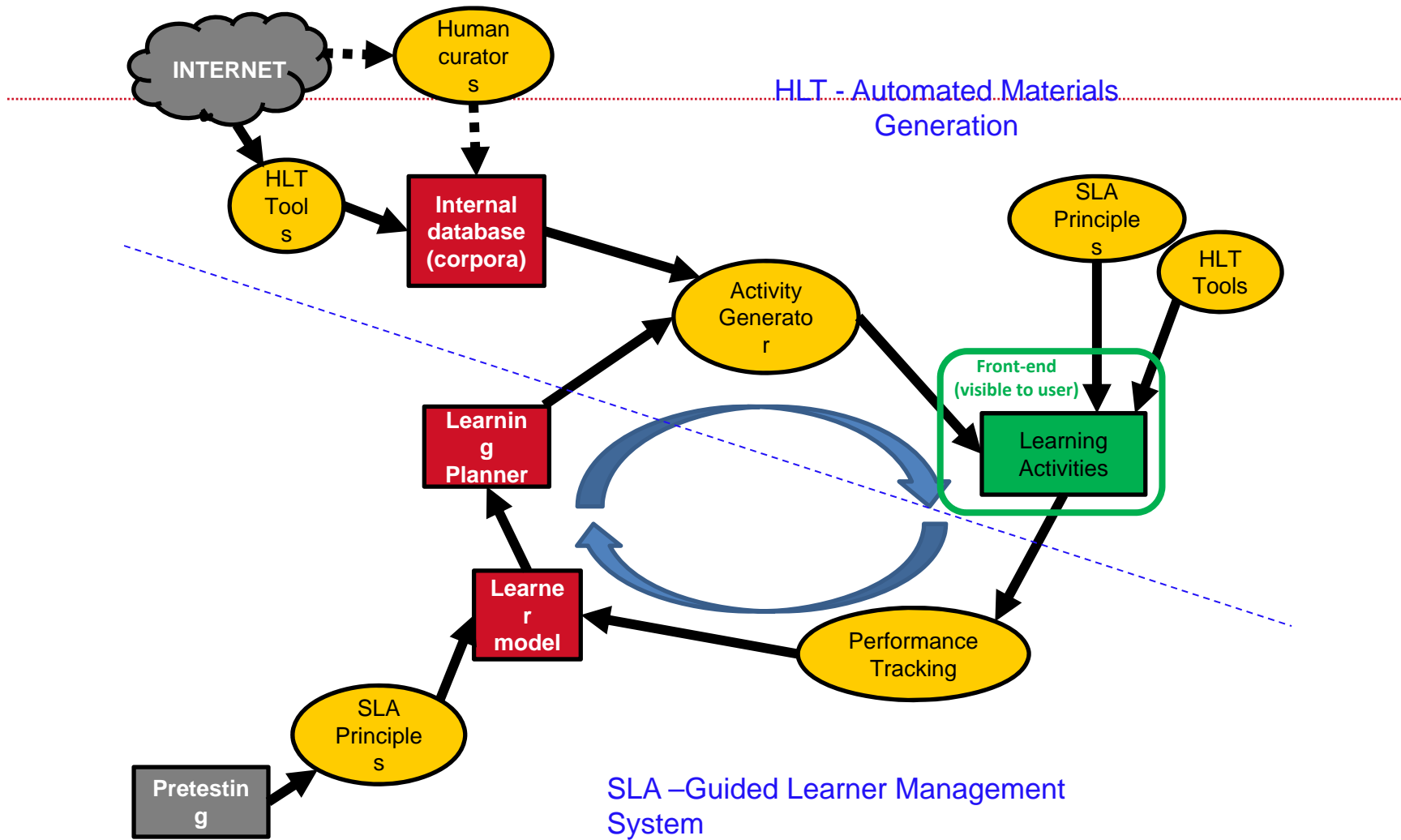
- Under the hood software architecture, human/computer interface, statistics
- Machine translation, part-of-speech tagging, word alignment, text leveling, topic identification
- Fetching and data analytics



CASL-IBM partnership

- Learning plan/activity implementation
- Vetting machine outputs and providing feedback
- Guidelines and best practices for native-speaker annotators
- Content evaluation
- Outcomes extensible to other languages





Back to reality

- A LanguageNation-like response to resource needs is not likely or feasible in most instances
- Constraints on time, money, expertise, etc.

What else can be done?



How do I find the best resources?

- Let your internet search be a first pass
- Cast a wide net - reach out to academic experts, teachers, those with insider access
- Try think-tanks who may be actively working on languages you care about
 - E.g., Center for Applied Linguistics, Linguistic Data Consortium, Center for Advanced Study of Language, among others



How do I evaluate the resources?

- Needs analysis
- Consider the source and scope
- Consider the audience
- Find a native speaker or other expert to review materials
- Evaluate the resource yourself!
 - Would it be helpful to **you**? What would need to be changed, adapted?



How do I adapt the resources?

- Use older works as a baseline to expand and/or build something new
- Develop new partnerships with colleagues and experts
- Work with a native speaker
- Make use of open source software
- Develop a workflow
- Use the internet to find new content/vocabulary/usage/context



Thank you!

Contact information:

Christopher Green

UMD-CASL

cgreen@umd.edu

301-226-8954

www.casl.umd.edu

