

Strategies towards the development of LTs and LRs for Ethiopian languages

Michael Gasser
School of Informatics
Indiana University



-
- Prioritize

-
- Prioritize
 - Collaborate

-
- Prioritize
 - Collaborate
 - Agree on standards

-
- Prioritize
 - Collaborate
 - Agree on standards
 - Learn from work in similar settings

-
- Prioritize
 - Collaborate
 - Agree on standards
 - Learn from work in similar settings
 - Take advantage of similarities among languages

-
- Prioritize
 - Collaborate
 - Agree on standards
 - Learn from work in similar settings
 - Take advantage of similarities among languages
 - Focus on practical projects

-
- Prioritize
 - Collaborate
 - Agree on standards
 - Learn from work in similar settings
 - Take advantage of similarities among languages
 - Focus on practical projects
 - Stimulate the growth of corpora

Prioritize

Prioritize

- Practical considerations

Prioritize

- Practical considerations
 - Tradeoff between achieving the full range of basic resources for Amharic and achieving a more limited range for multiple languages

Prioritize

- Practical considerations
 - Tradeoff between achieving the full range of basic resources for Amharic and achieving a more limited range for multiple languages
- Political considerations

Prioritize

- Practical considerations
 - Tradeoff between achieving the full range of basic resources for Amharic and achieving a more limited range for multiple languages
- Political considerations
- National needs and goals

Prioritize

- Practical considerations
 - Tradeoff between achieving the full range of basic resources for Amharic and achieving a more limited range for multiple languages
- Political considerations
- National needs and goals
- Ethiopian regional needs and goals

Prioritize

- Practical considerations
 - Tradeoff between achieving the full range of basic resources for Amharic and achieving a more limited range for multiple languages
- Political considerations
- National needs and goals
- Ethiopian regional needs and goals
- International needs and goals

Collaborate

Collaborate

- Becoming aware of what's already going on

Collaborate

- Becoming aware of what's already going on
- Avoiding duplication of efforts

Collaborate

- Becoming aware of what's already going on
- Avoiding duplication of efforts
- Dividing and conquering

Collaborate

- Becoming aware of what's already going on
- Avoiding duplication of efforts
- Dividing and conquering
- Communicating within the community

Collaborate

- Becoming aware of what's already going on
- Avoiding duplication of efforts
- Dividing and conquering
- Communicating within the community
- Role of new IT PhD: centralization of computational work on Ethiopian languages

Collaborate

- Becoming aware of what's already going on
- Avoiding duplication of efforts
- Dividing and conquering
- Communicating within the community
- Role of new IT PhD: centralization of computational work on Ethiopian languages
- Common data repository (DY's site?)

Collaborate

- Becoming aware of what's already going on
- Avoiding duplication of efforts
- Dividing and conquering
- Communicating within the community
- Role of new IT PhD: centralization of computational work on Ethiopian languages
- Common data repository (DY's site?)
- Coordinating committee

Agree on standards

Agree on standards

- Grammars

Agree on standards

- Grammars
- Morphological and POS categories

Agree on standards

- Grammars
- Morphological and POS categories
- Data format

Agree on standards

- Grammars
- Morphological and POS categories
- Data format
- Romanization

Learn from work in similar settings

Learn from work in similar settings

- Countries with multiple “disadvantaged” languages

Learn from work in similar settings

- Countries with multiple “disadvantaged” languages
 - South Africa

Learn from work in similar settings

- Countries with multiple “disadvantaged” languages
 - South Africa
 - India

Take advantage of similarities among the languages

Take advantage of similarities among the languages

- Ethiopia as a language area

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations
- Lexicon

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations
- Lexicon
 - ያስፈልገኛል። የድልዎኒ። *na barbaachisa*. 'I need (it).'

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations
- Lexicon
 - ያስፈልገኛል። የድልየኒ። *na barbaachisa*. 'I need (it).'
 - ሳይመጣ አይቀርም። ... አይተርፍን። ... *hinhafu*. 'He'll probably come.'

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations
- Lexicon
 - ያስፈልገኛል። የድልየኒ። *na barbaachisa*. 'I need (it).'
 - ሳይመጣ አይቀርም። ... አይተርፍን። ... *hinhafu*. 'He'll probably come.'
 - መጥቶ አያውቅም። ... አይፈልጥን። ... *hinbeeku*. 'He has never come.'

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations
- Lexicon
 - ያስፈልገኛል። የድልዩኒ። *na barbaachisa*. 'I need (it).'
 - ሳይመጣ አይቀርም። ... አይተርፍን። ... *hinhafu*. 'He'll probably come.'
 - መጥቶ አያውቅም። ... አይፈልጥን። ... *hinbeeku*. 'He has never come.'
- Sharing algorithms, code

Take advantage of similarities among the languages

- Ethiopia as a language area
- Morphology
 - Semitic: passive, causative, frequentative, combinations
 - Oromo: passive, causative, frequentative, combinations
- Lexicon
 - ያስፈልገኛል። የድልየኒ። *na barbaachisa*. 'I need (it).'
 - ሳይመጣ አይቀርም። ... አይተርፍን። ... *hinhafu*. 'He'll probably come.'
 - መጥቶ አያውቅም። ... አይፈልጥን። ... *hinbeeku*. 'He has never come.'
- Sharing algorithms, code
- Using a language with more resources to bootstrap work on other languages

Focus on large-scale practical projects

Focus on large-scale practical projects

- Provide focus and encourage collaboration across groups

Focus on large-scale practical projects

- Provide focus and encourage collaboration across groups
- Health applications

Focus on large-scale practical projects

- Provide focus and encourage collaboration across groups
- Health applications
- Rural development applications

Focus on large-scale practical projects

- Provide focus and encourage collaboration across groups
- Health applications
- Rural development applications
- Small business applications

Focus on large-scale practical projects

- Provide focus and encourage collaboration across groups
- Health applications
- Rural development applications
- Small business applications
- Educational applications: computer-assisted language instruction, intelligent tutoring

Stimulate the growth of corpora

Stimulate the growth of corpora

- Utilizing the speech community

Stimulate the growth of corpora

- Utilizing the speech community
- Creation of bilingual corpora through human-machine collaboration, with increasing machine involvement as the MT systems learn to become more sophisticated

Stimulate the growth of corpora

- Utilizing the speech community
- Creation of bilingual corpora through human-machine collaboration, with increasing machine involvement as the MT systems learn to become more sophisticated
- Wikis where language experts can edit and discuss translations, providing feedback to MT system, as well as more bilingual texts