



# **Multilingualism in the cyberspace Technological challenges and solutions**

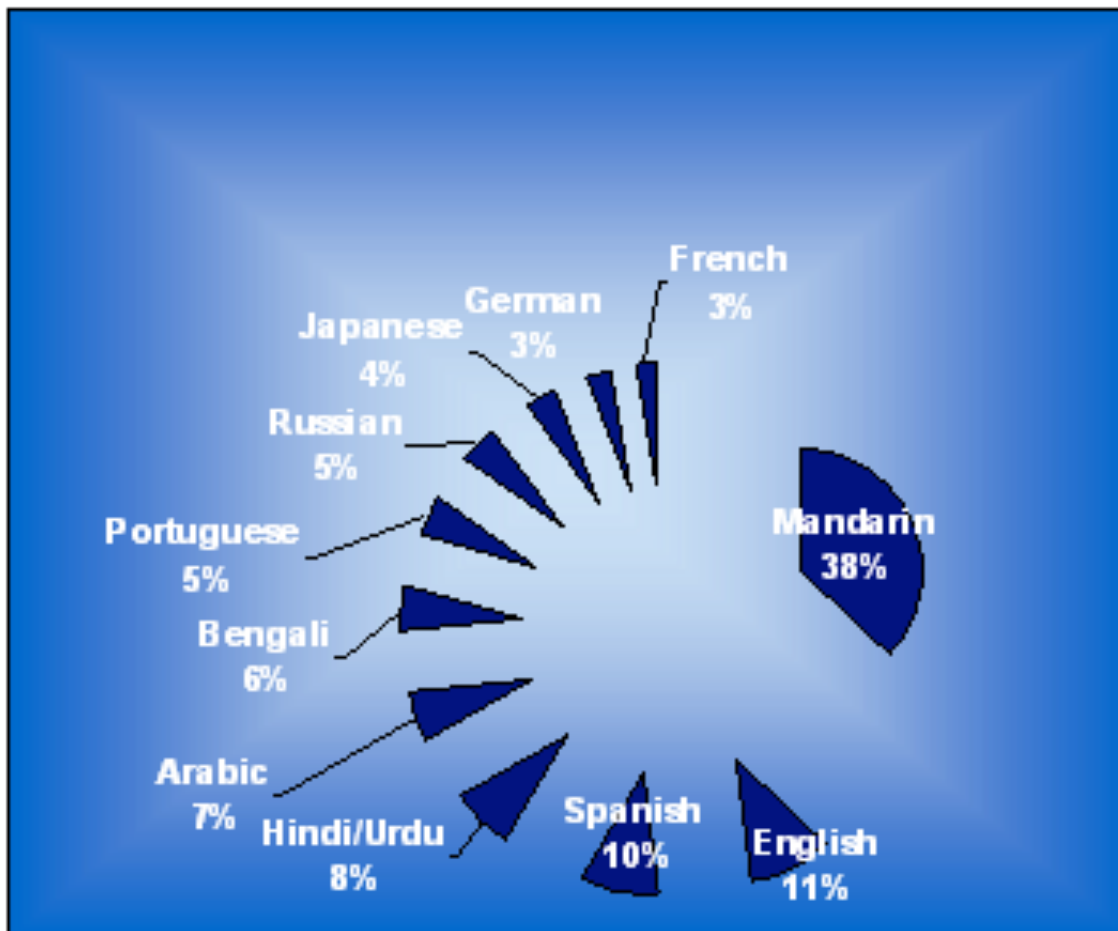
**Dawit Bekele**

**Assistant Professor, Addis Ababa  
University**

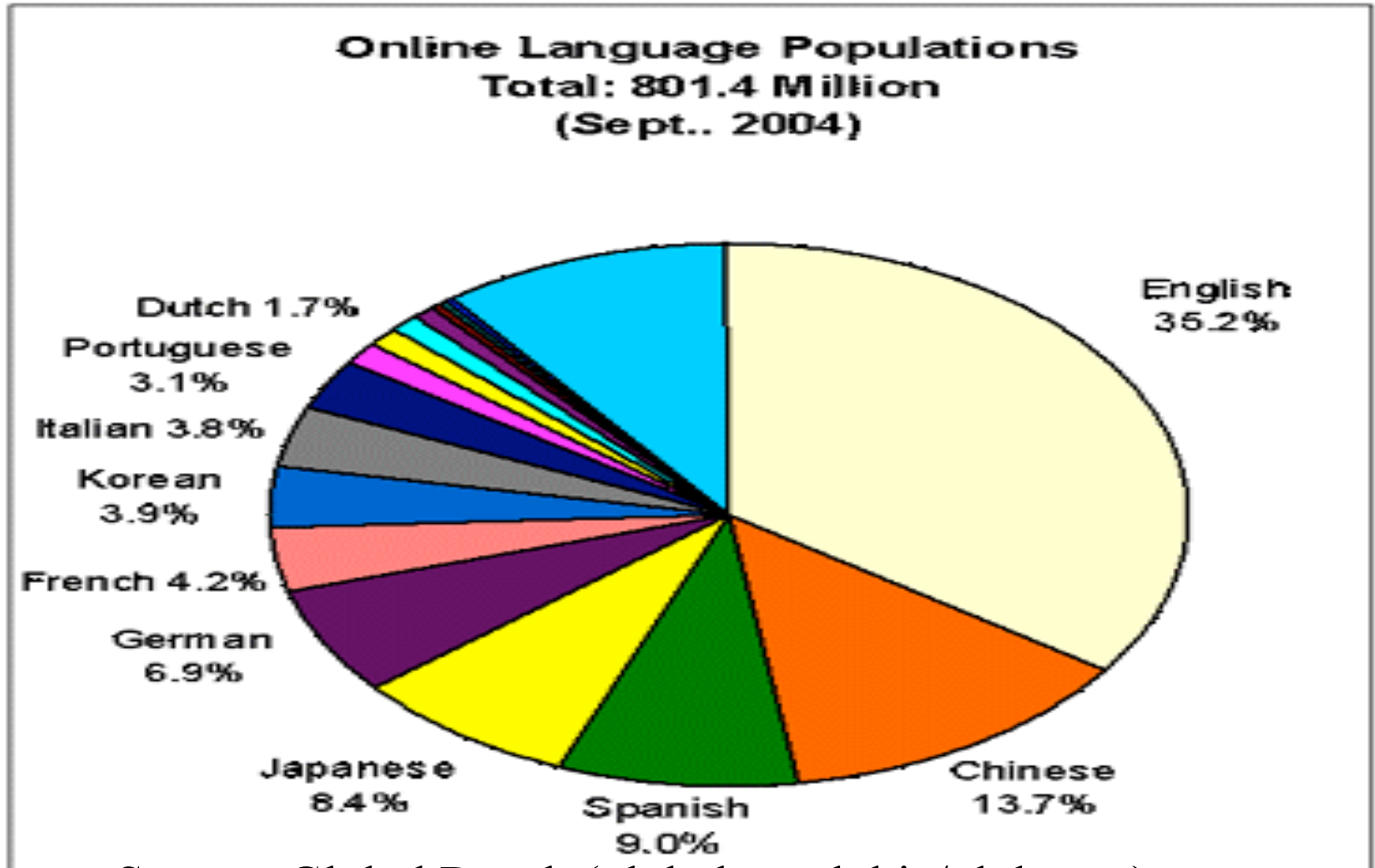
## Introduction

- Multilingualism is increasingly considered as very important for the stability, peace and development in the world
- It should be nurtured in the real world as well as in the cyberspace
- This presentation tries to
  - show the discrepancy that exists in terms of the languages in the world and the cyberspace
  - Identify hardware and software causes
  - Propose solutions

## Languages of the world

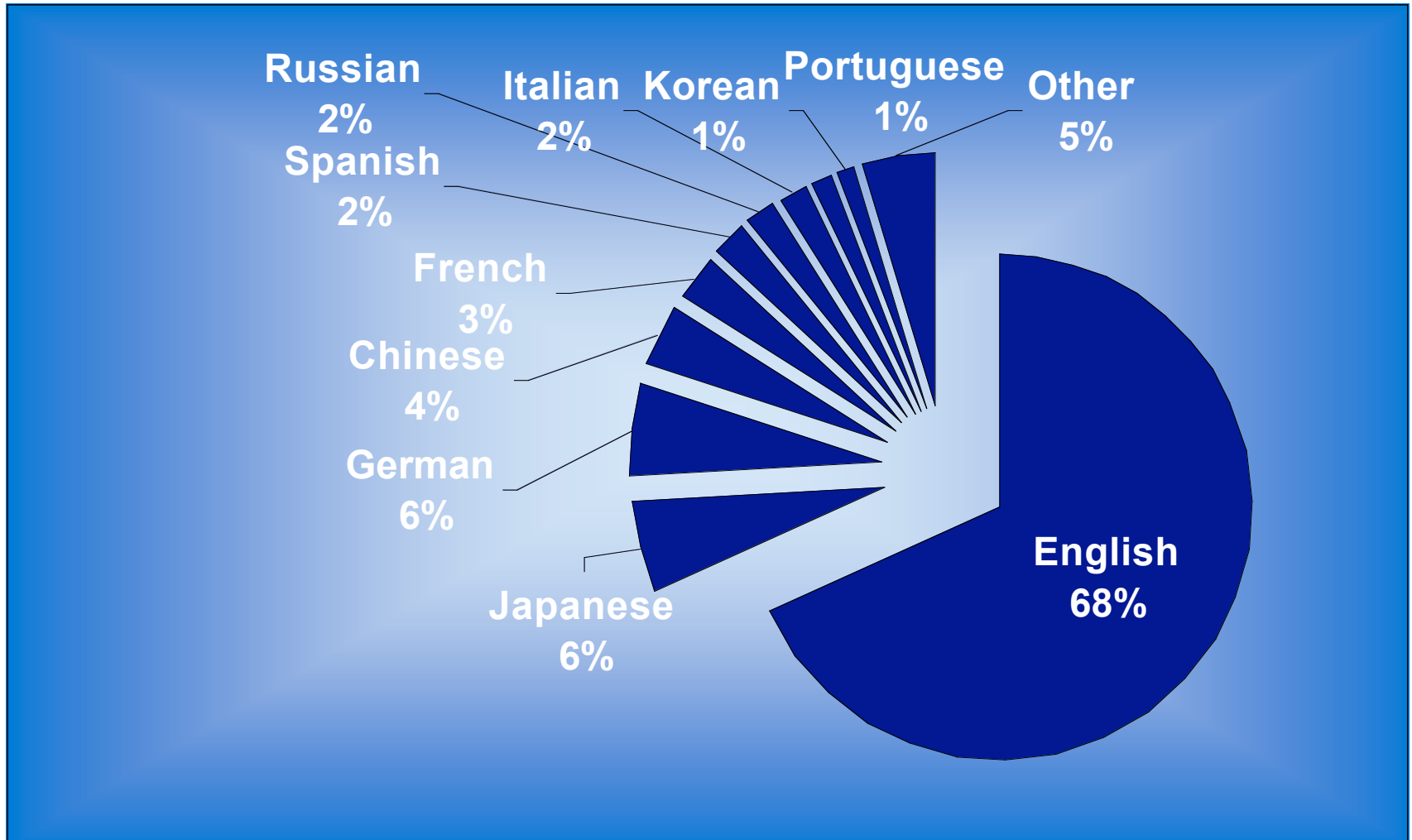


## Languages on the cyberspace – On-line population



Source: Global Reach ([global-reach.biz/globstats](http://global-reach.biz/globstats))

## Languages on the cyberspace - Content



Source: Global Reach ([global-reach.biz/globstats](http://global-reach.biz/globstats))

## Problems

- Many languages do not exist in the cyberspace
  - 90% of languages are not found on the Internet [Source: <http://alumni.indiana.edu/lostlanguages/stats.html>]

## Problems ...

- The proportion of language of the content on the cyberspace is very different from that of the population
  - English
    - 11% of the population of the world speaks English
    - 35% of on-line population has English as a native language
    - 68% of the cyberspace content is in English
  - Mandarin
    - 38% of the population of the world speaks Mandarin
    - 13.7% of on-line population has Mandarin as a native language
    - 4% of the cyberspace content is in Mandarin
  - There are many languages of the world where the situation is much worse than Mandarin

## Why this discrepancy?

- Economic
  - Digital divide
- Technological
  - The focus of this presentation
- Two main technological problems
  - Problem to develop the digital content in the languages using computer hardware and software
  - Problem of representation of content

## Problem to develop the digital content

- Difficulty to find hardware that supports some languages' scripts (ex. Keyboard)
- Difficulty to find the software to develop content in the language of the content
  - Example: Ethiopic content has been developed for decades with software with English interface
- Limits the number of people who can develop content in that language

## Problem to develop the digital content ...

- What is the solution?
  - Localization - change the interface of the software to local languages, culture and tradition
  - For decades, proprietary software developers didn't want to localize for languages that do not have economic power
  - Recently, many localizations are being done using Free and Open Source Software

## FOSS gives the freedom to localize

- No need of authorization to localize a FOSS software
- Economic reasons not to localize are much less important than with proprietary software
  - The developers are not necessarily the localizers
  - Localizers have other reasons (pride, political will, technical interest, etc.)
- Since the source code is available all sorts of localization are possible
  - Time and date localization
  - Customization

## Problem of representation of content

- Until the 1990s ASCII, was the standard of Internet and was adequate only for Latin based languages
- It was necessary to use complex methods just to represent the content of other scripts and alphabets
  - Ex: Amharic: Image, Specific downloadable font, etc.
- UNICODE is solving the problem of encoding since it is the CODE of the world scripts and alphabets
- But there are other standards that need to consider multilingualism (ex. XML) in order to be able to develop a content in any language with the same ease as in English
- That's why Africans should involve in standardization works

## Conclusion

- The first step towards multilingualism in the cyberspace is to have, in the cyberspace, content in all languages of the world in a proportion that respects the population of the world
- Localization is crucial for multilingualism