

The Digital Plan for Irish

18th EFNIL Conference, Cavtat
06 October 2021

Neasa Ní Chiaráin
neasa.nichiarain@tcd.ie

The Irish language

Q-Celtic

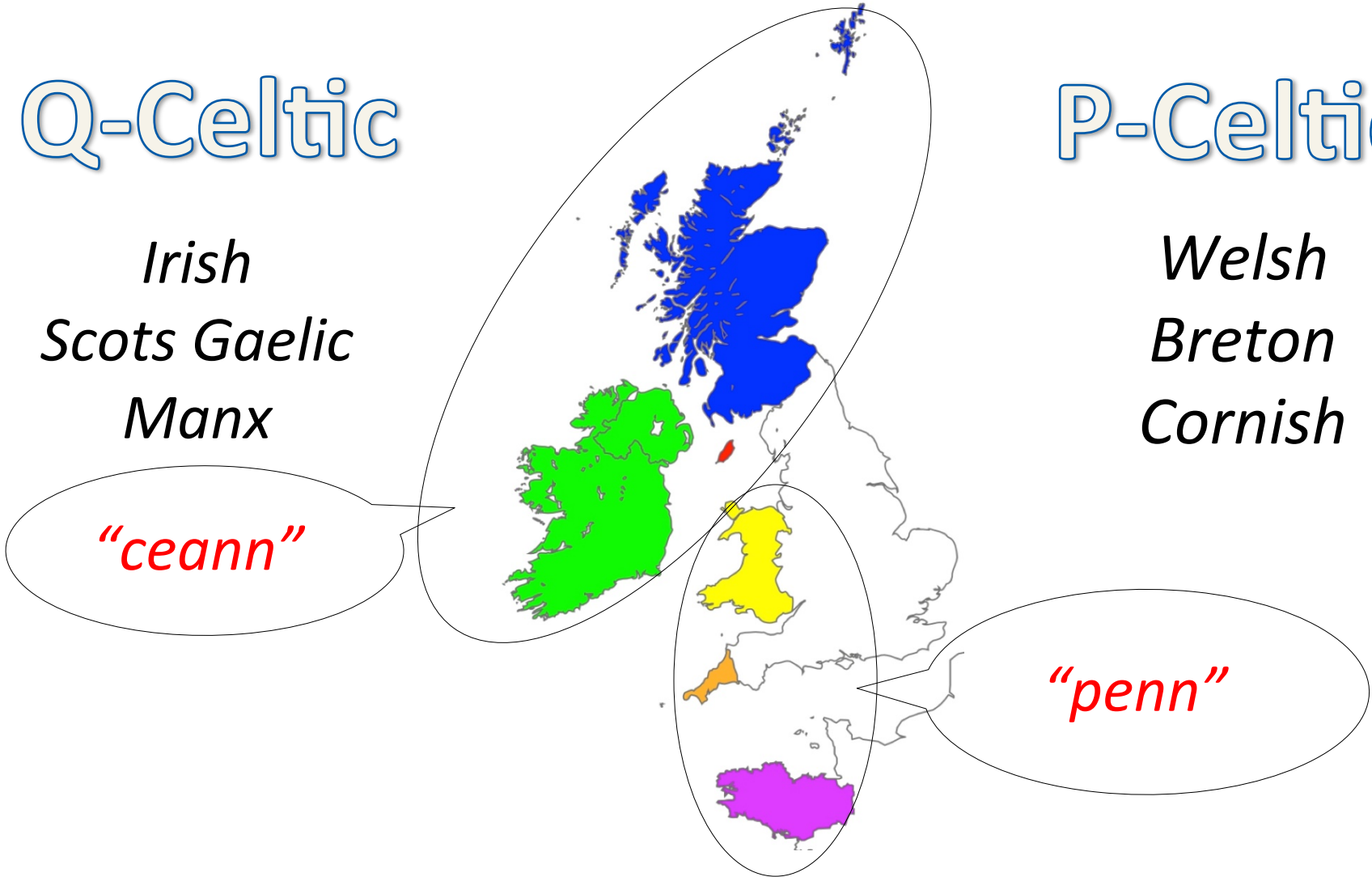
Irish
Scots Gaelic
Manx

“ceann”

P-Celtic

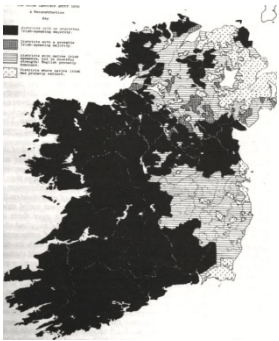
Welsh
Breton
Cornish

“penn”

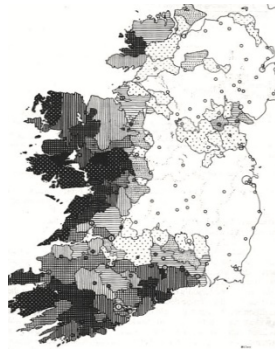


Decline over 150 years

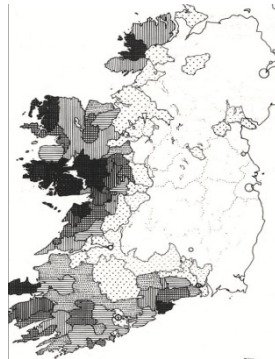
1800



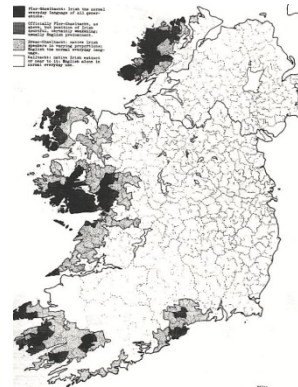
1851



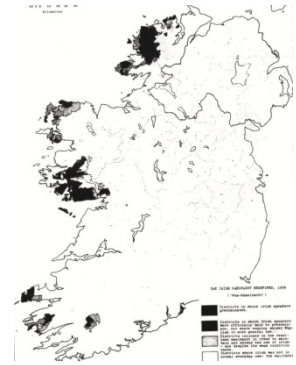
1891



1926



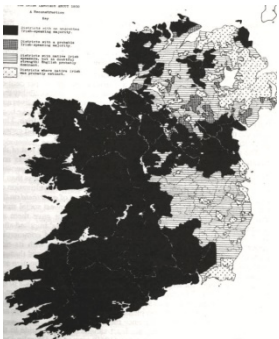
1956



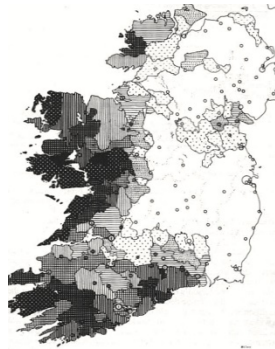
Decline over 150 years

Gaeltacht pockets:
3 main dialects

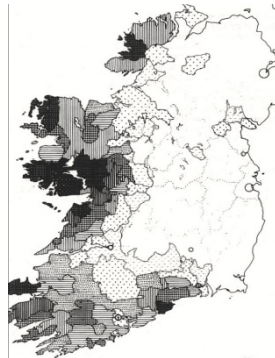
1800



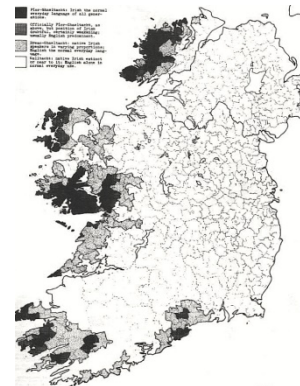
1851



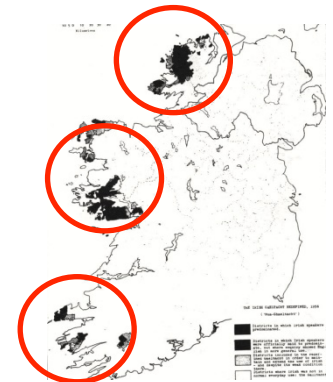
1891



1926



1956



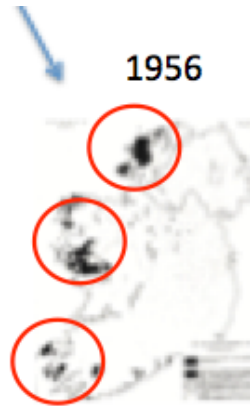
A Digital Plan

- Political support as a first national language is the crucial driving factor
- Revival and maintenance
- to do that the Digital Plan for Speech and Language Technology has to be firmly focused on the needs of the **Irish language community**

Irish-language community

Gaeltacht

- Very different dialects
- No spoken standard
- A written standard (artificially constructed with element from the 3 main dialects)
- ❖ Speech technology has to be multi-dialect



Non-Gaeltacht

- Obligatory up to age 18
 - Proficiency dropping
 - Standards very variable
- Education is key
- Many challenges...
- access to native speaker models of the language
 - modern attractive digital materials, etc.

National Language Institutions

- Governmental: The Department of Tourism, Culture, Arts, **Gaeltacht**, Sports and Media

Foras na Gaeilge

- State Body responsible for promoting Irish North & South
- It funds the lead-organisations, with key responsibilities in domains relevant to the language/language community
- **Particular obligation: terminology & dictionaries**
 - ongoing development for many years
 - current expansion:
(Irish-Irish, Irish-English dictionaries)



National Language Institutions

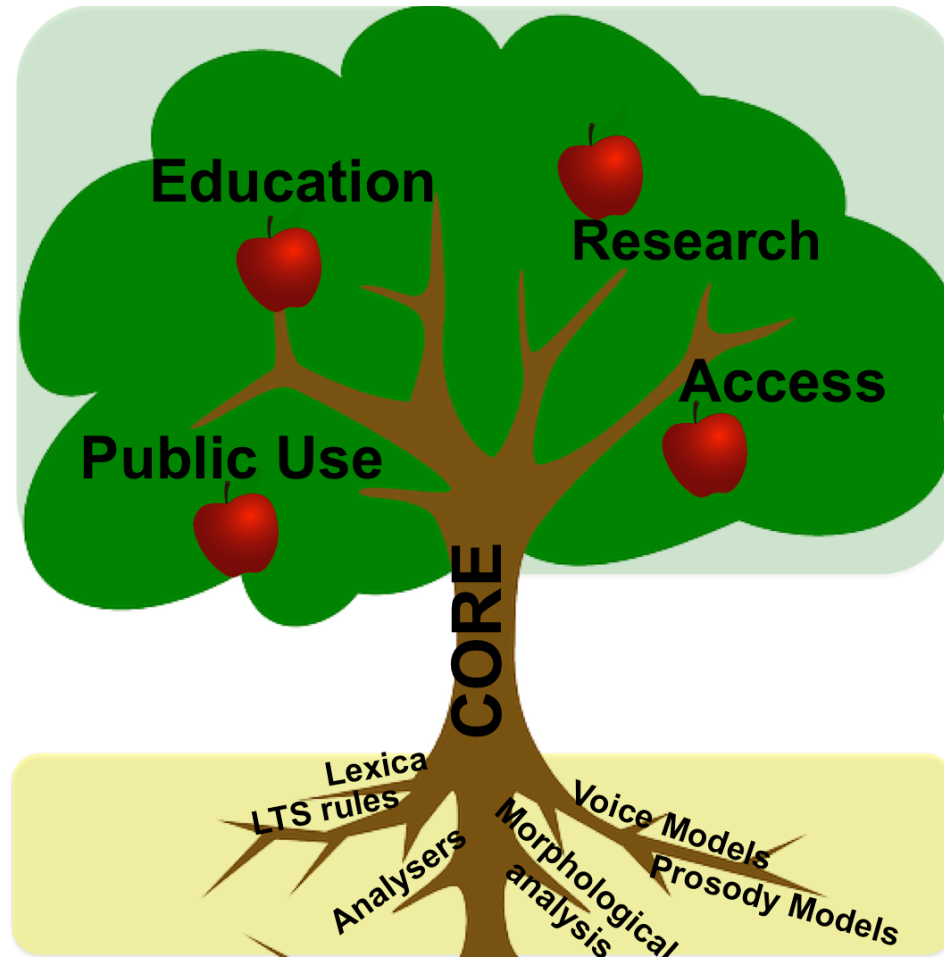
- Governmental: The Department of Tourism, Culture, Arts, **Gaeltacht**, Sports and Media

Digital Plan for Irish

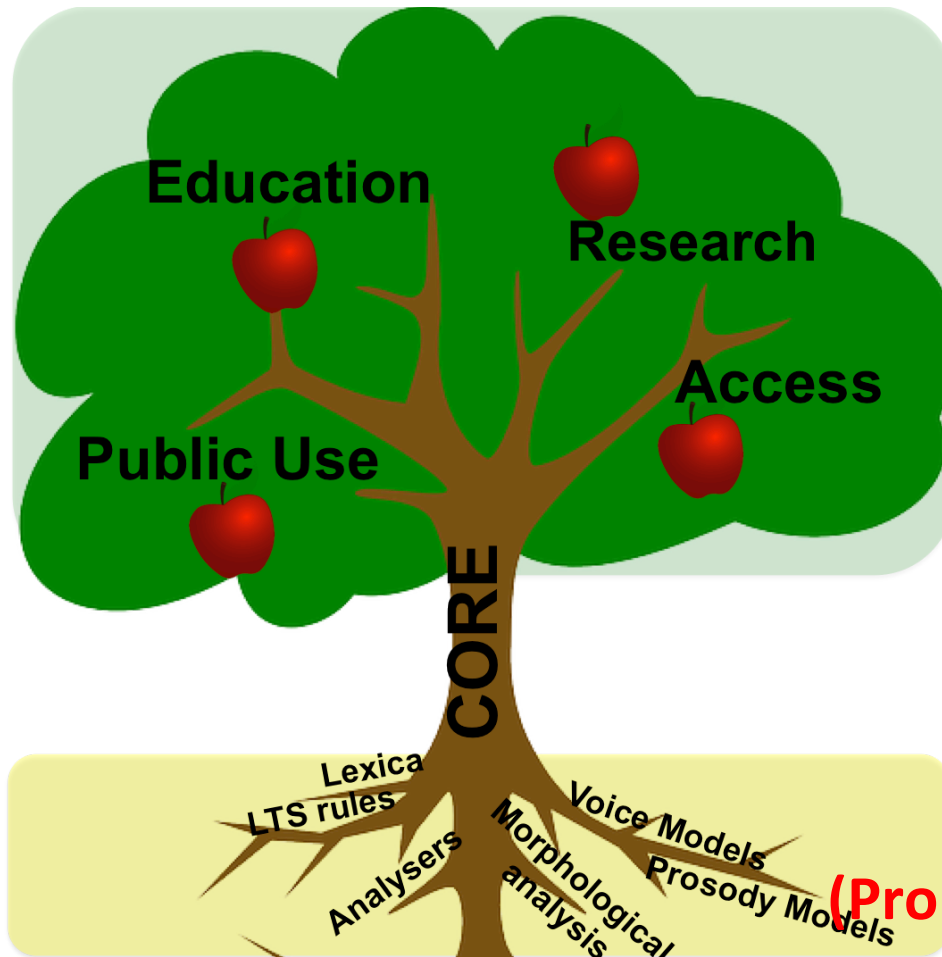
Towards the implementation of:

- *20 Year Strategy for the Irish Language 2010-2030*
- *The Action Plan for the Irish Language, 2018-2022*

The Digital Plan



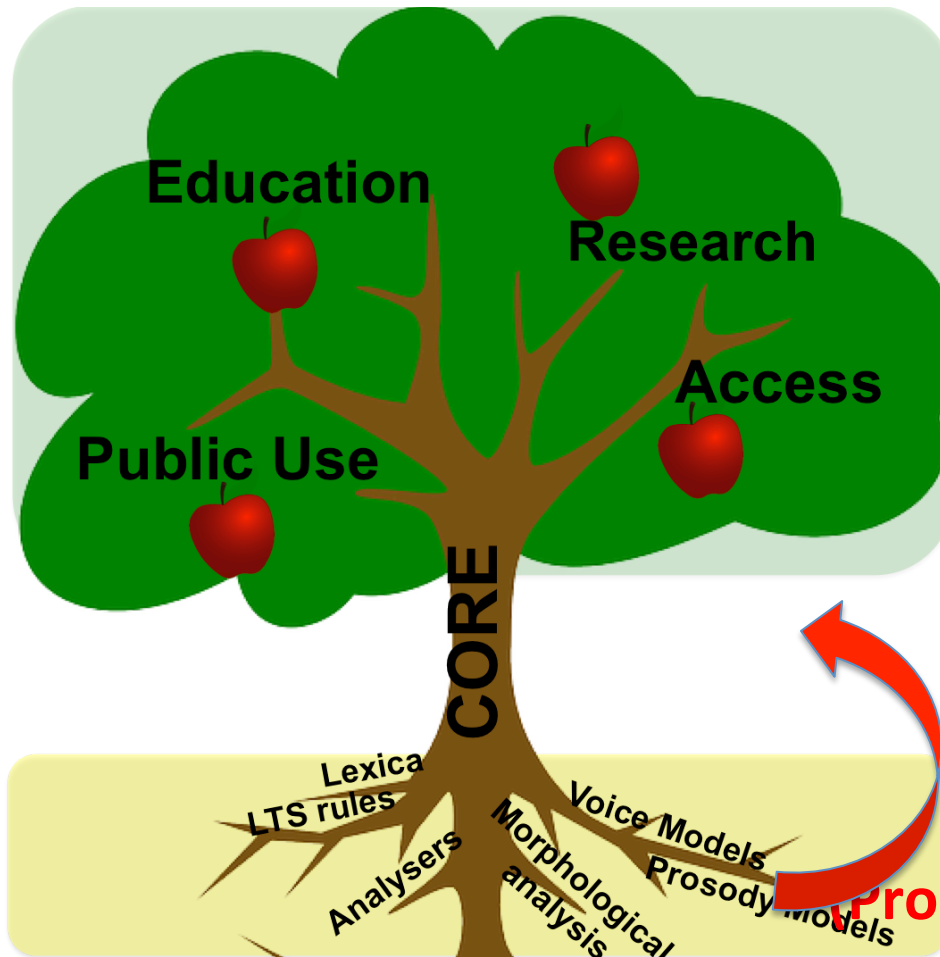
The Digital Plan



Speech & text corpora
(Pronunciation) dictionaries
Language models
Acoustic models

I
Resources

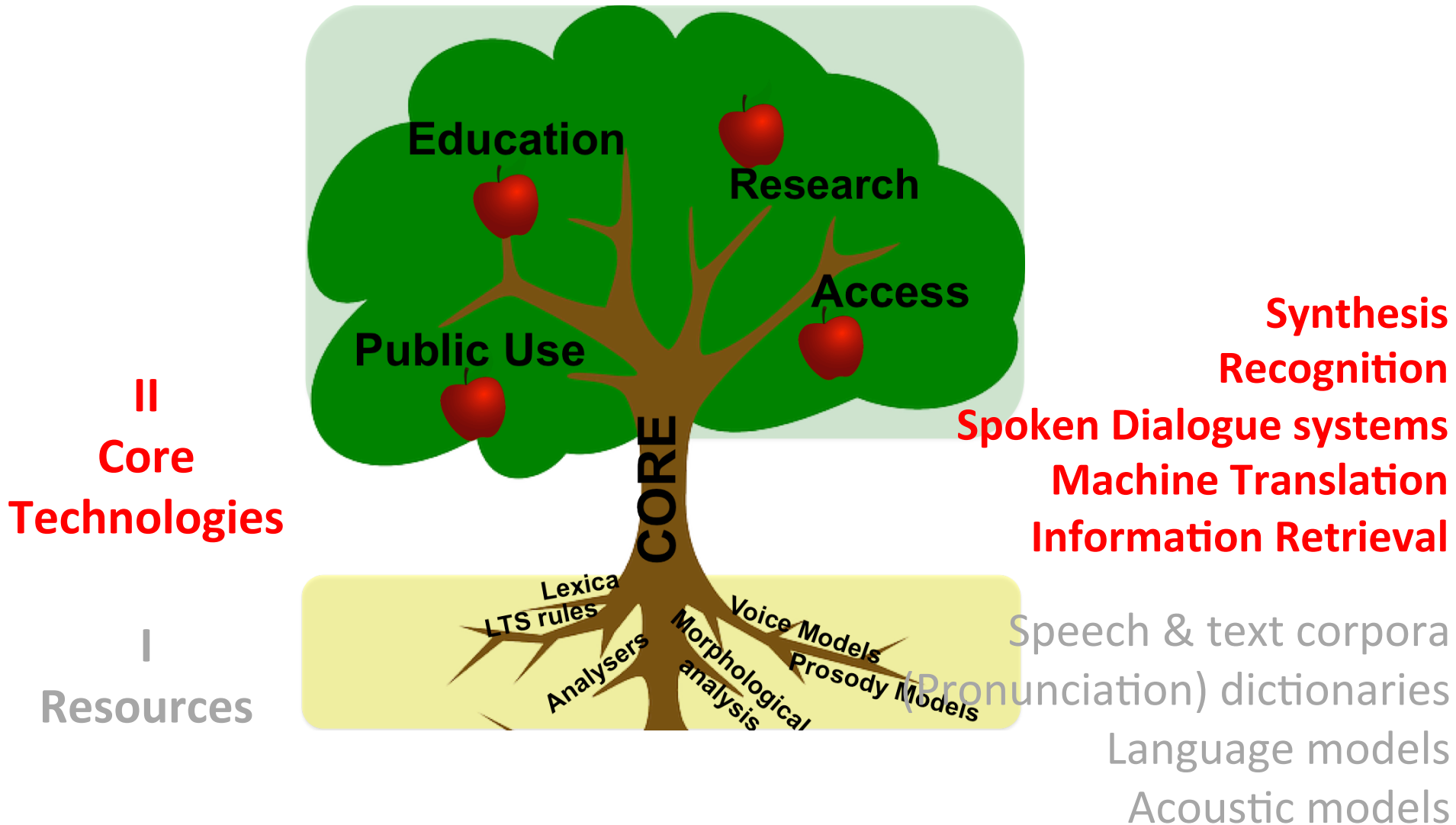
The Digital Plan



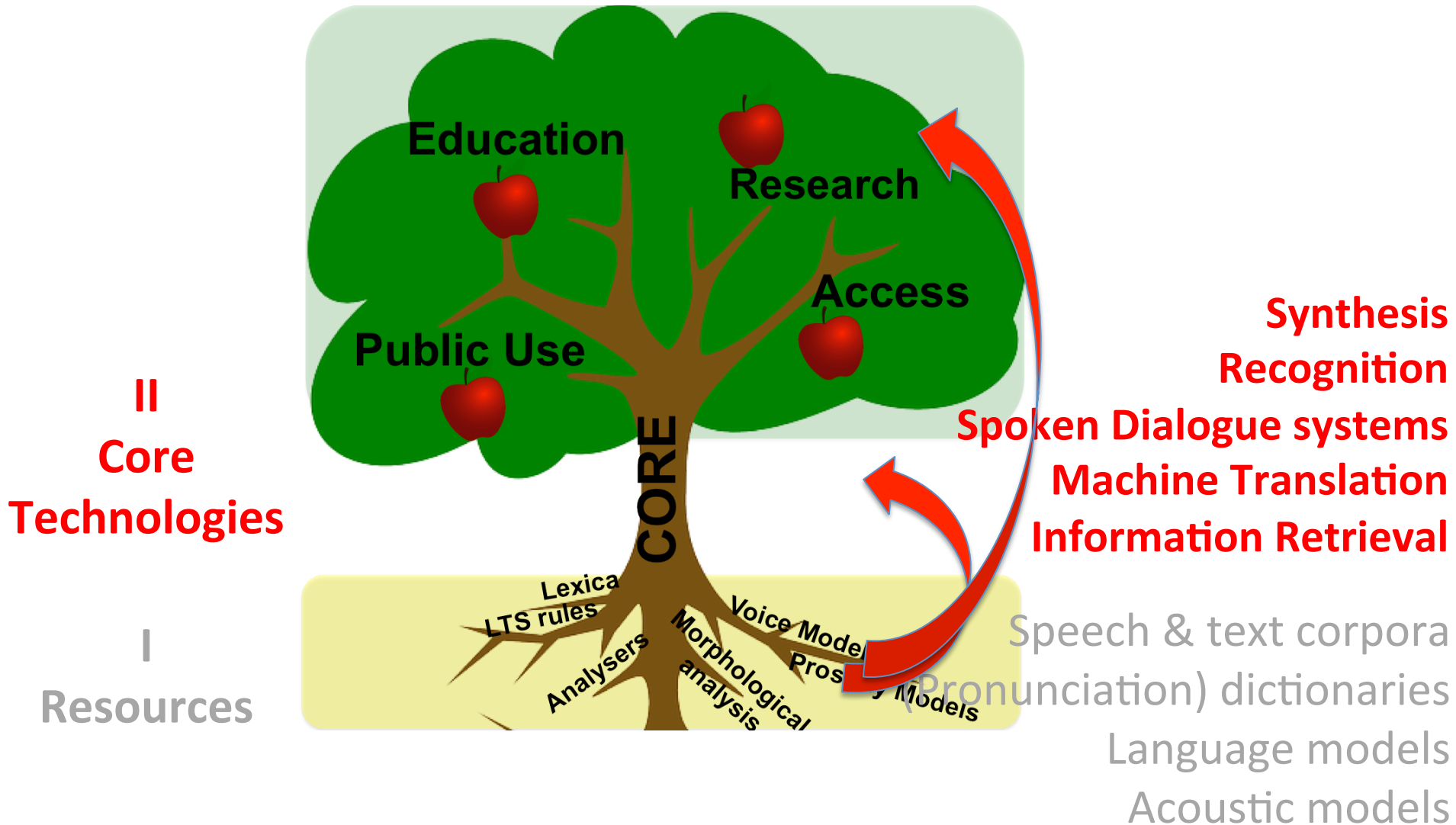
Speech & text corpora
(pronunciation) dictionaries
Language models
Acoustic models

I
Resources

The Digital Plan



The Digital Plan

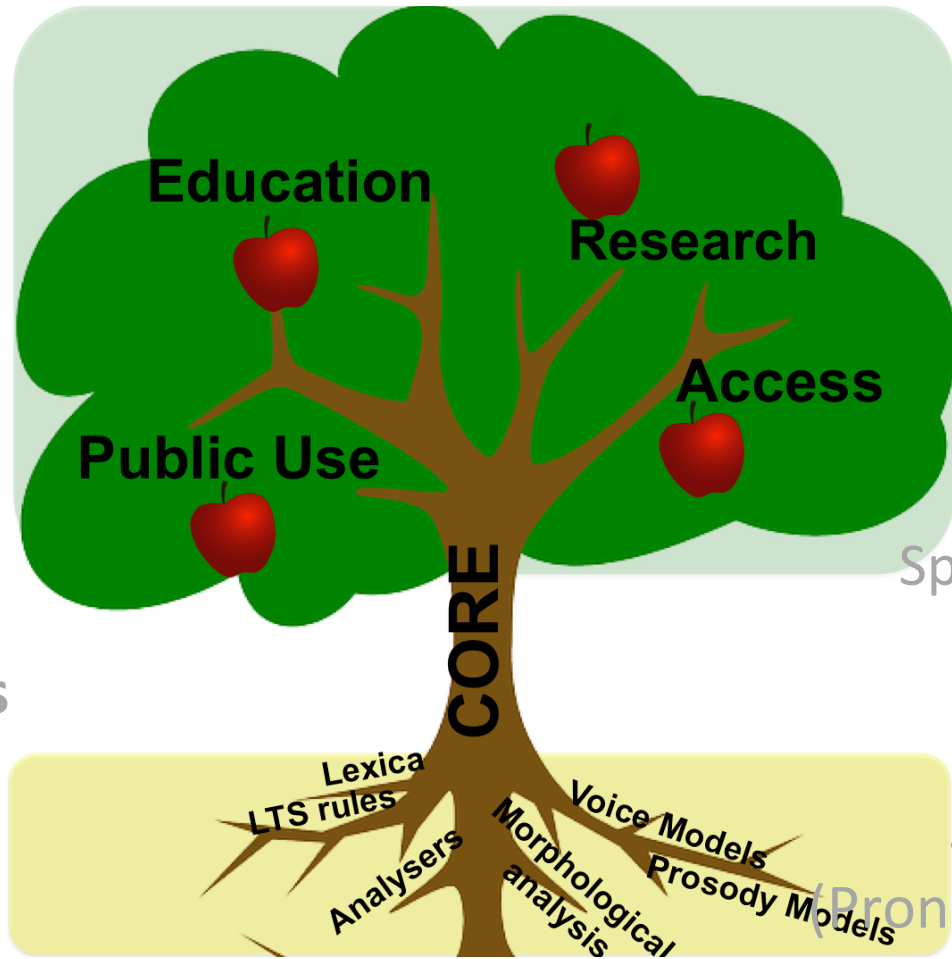


The Digital Plan

III
Applications

II
Core Technologies

I
Resources

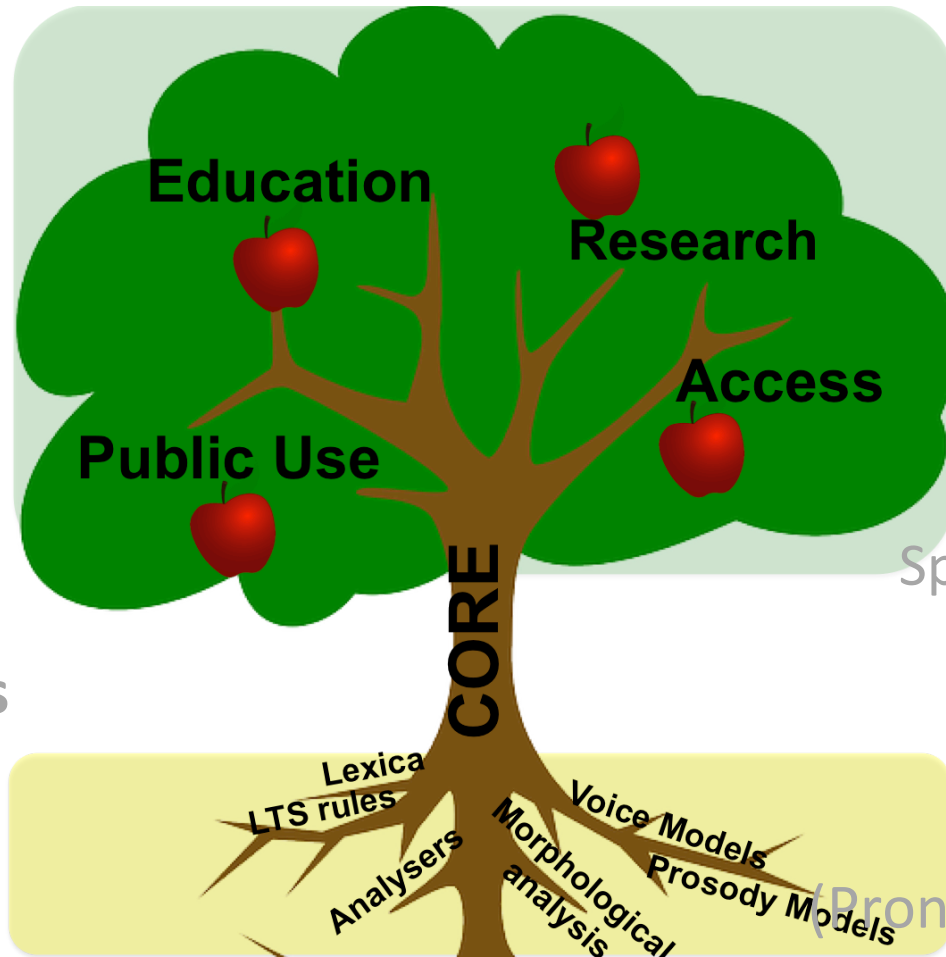


General Public
Education
Access/Disability

- Synthesis
- Recognition
- Spoken Dialogue systems
- Machine Translation
- Information Retrieval
- Speech & text corpora
- (Pronunciation) dictionaries
- Language models
- Acoustic models

The Digital Plan

III
Applications



General Public

Education

Access/Disability

Synthesis

Recognition

Spoken Dialogue systems

Machine Translation

Information Retrieval

Speech & text corpora

(Pronunciation) dictionaries

Language models

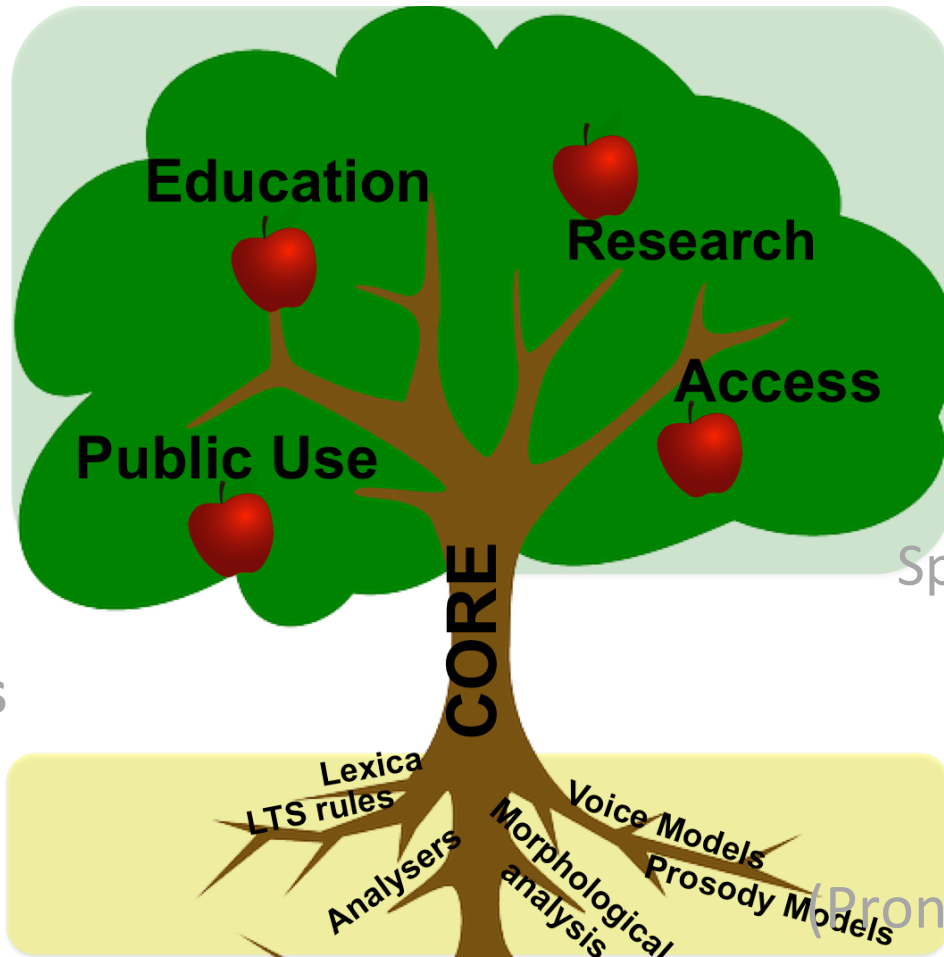
Acoustic models

II
Core
Technologies

I
Resources

The Digital Plan

III
Applications



General Public
Education

Access/Disability

Synthesis

Recognition

Spoken Dialogue systems

Machine Translation

Information Retrieval

Speech & text corpora

(Pronunciation) dictionaries

Language models

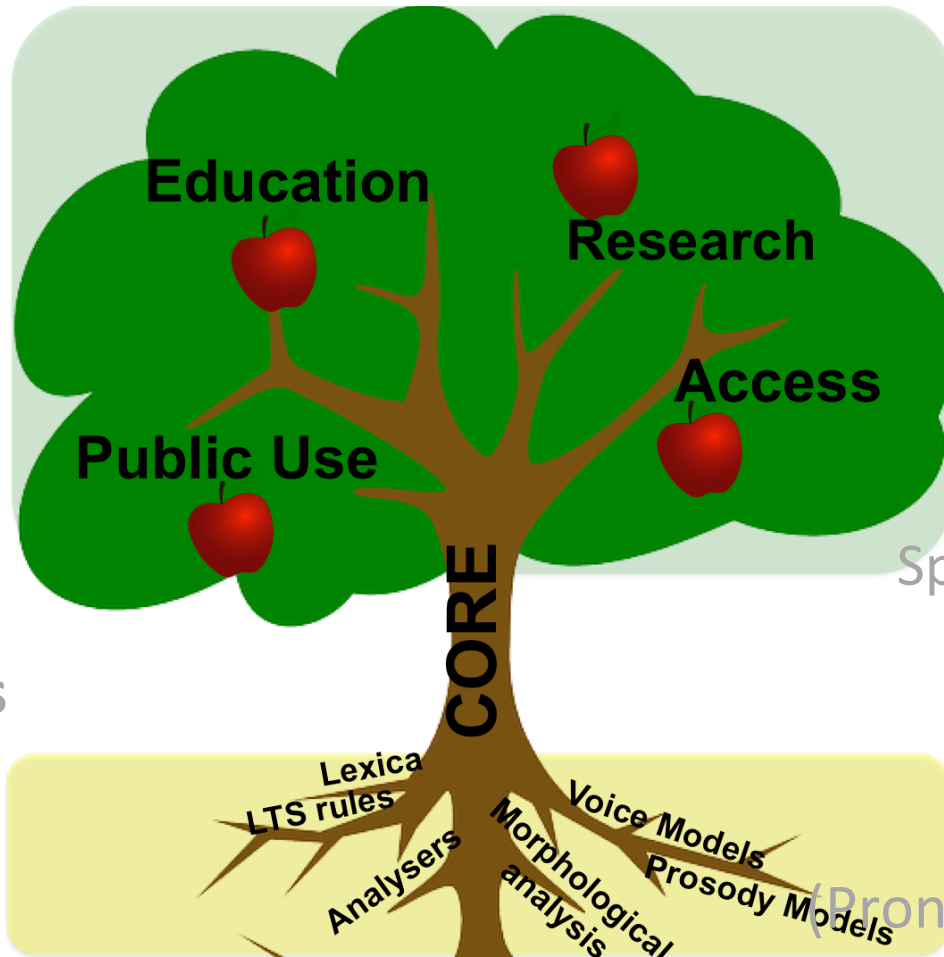
Acoustic models

II
Core
Technologies

I
Resources

The Digital Plan

III
Applications



General Public
Education

Access/Disability

Synthesis

Recognition

Spoken Dialogue systems

Machine Translation

Information Retrieval

Speech & text corpora

(Pronunciation) dictionaries

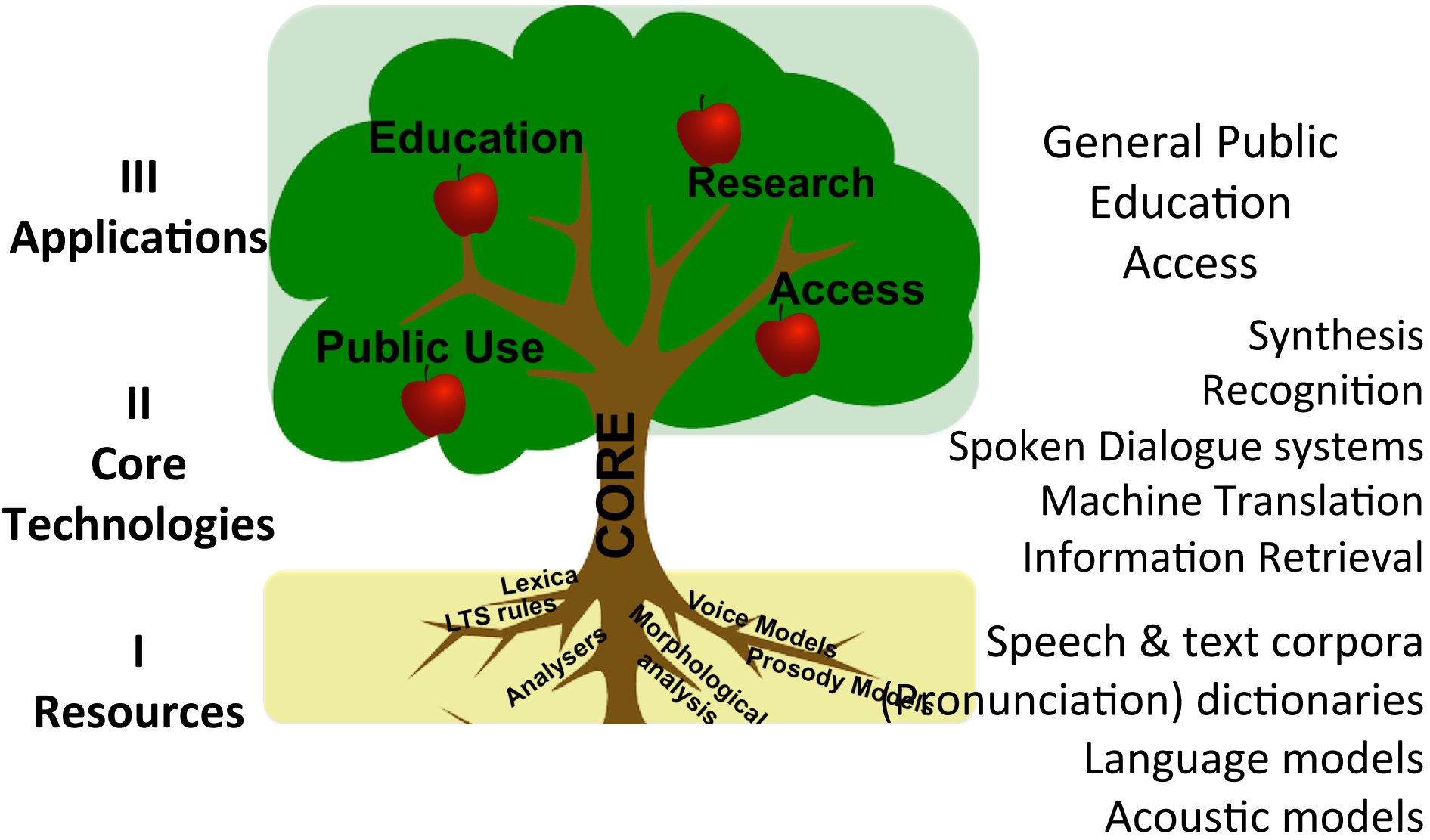
Language models

Acoustic models

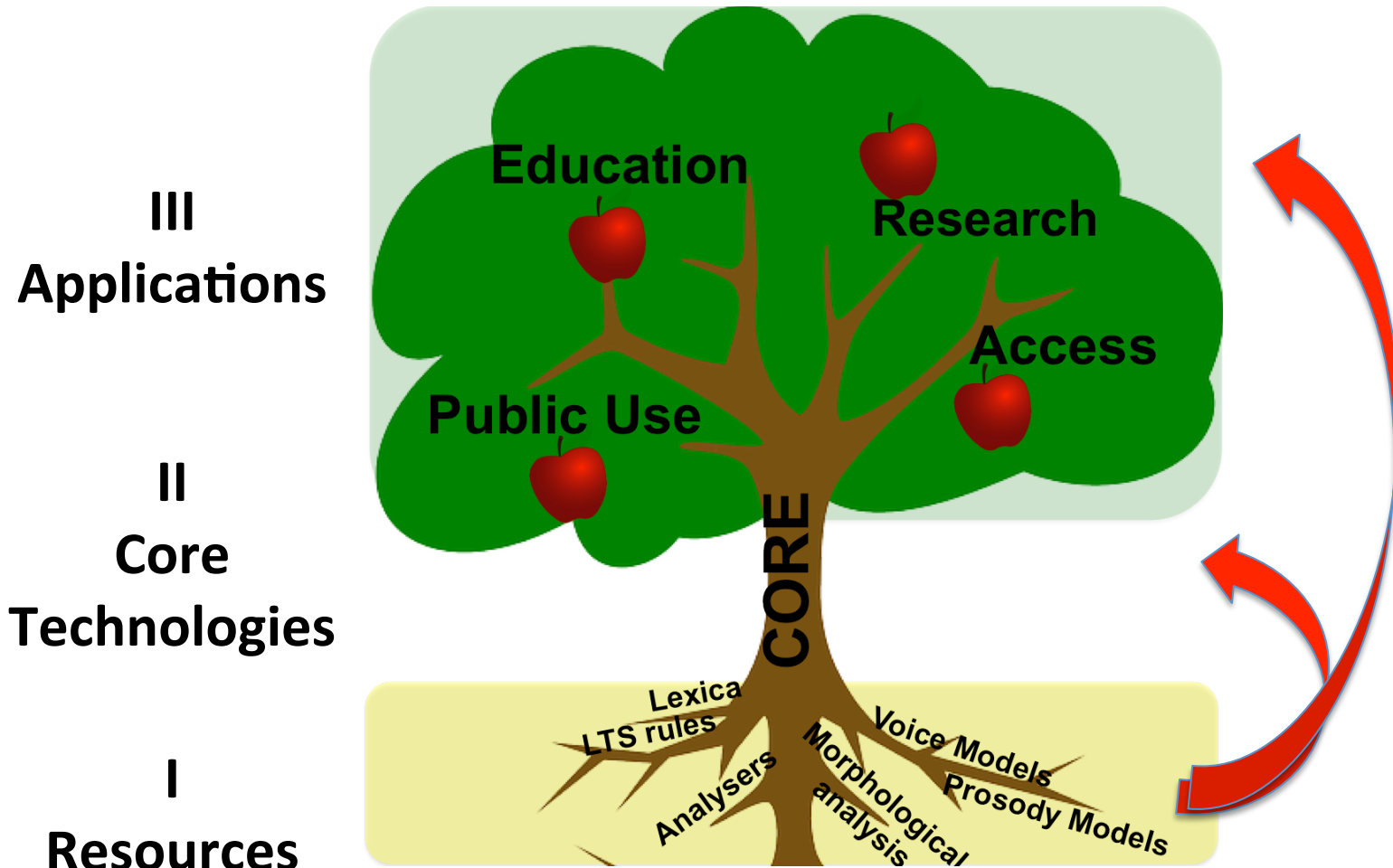
II
Core
Technologies

I
Resources

The Digital Plan



The Digital Plan



The Digital Plan

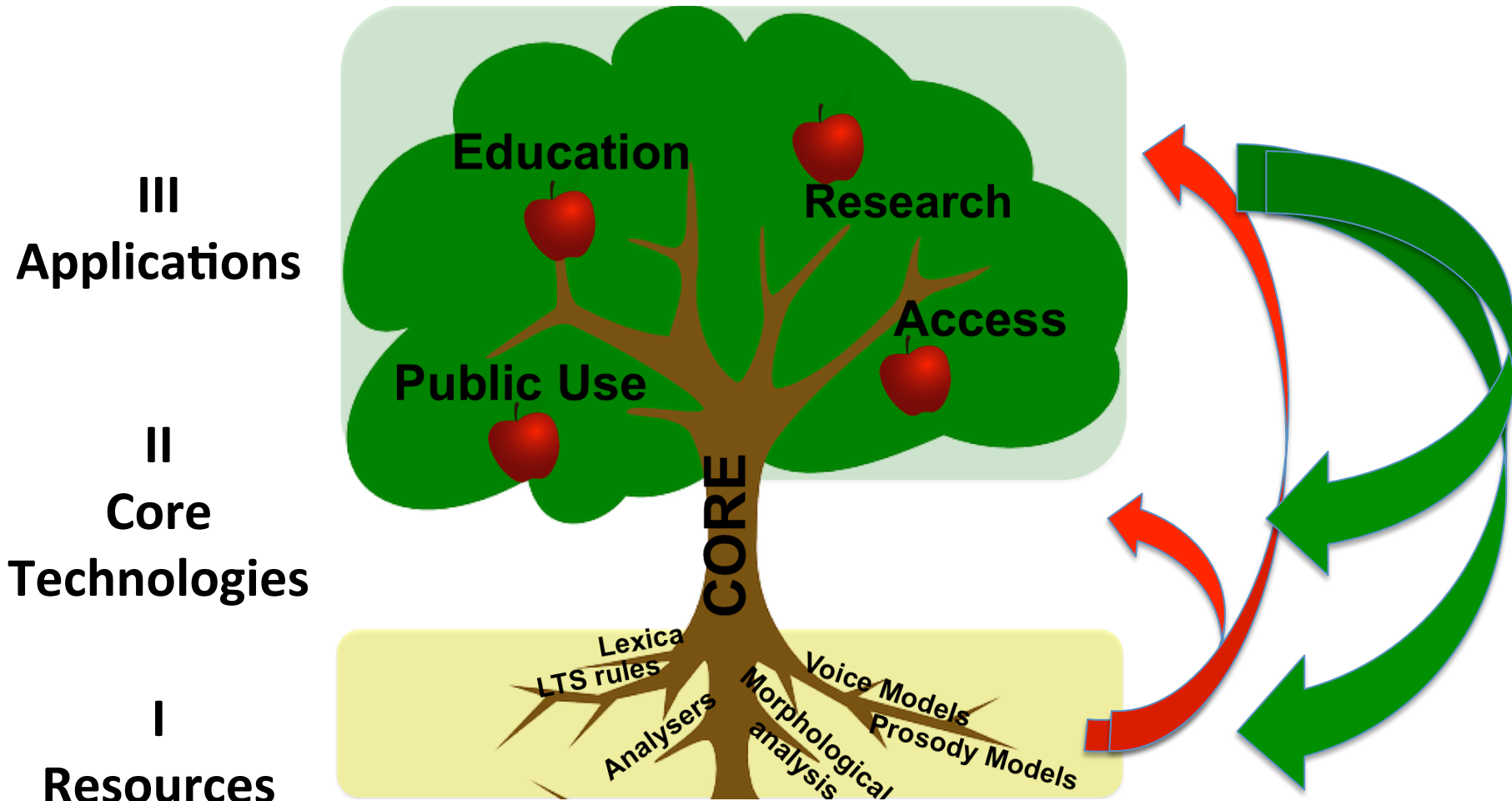


Table of contents

Part I

Part II

1. Introduction: Irish in the Digital Age
2. Linguistic Analysis and Digital Documentation of Spoken and Written Irish
3. Language Resources: Corpora
4. Linguistic Knowledge Bases
5. Natural Language Processing Tools
6. Natural Language Generation
7. Speech Models for Irish

8. Speech Synthesis
9. Speech recognition
10. Dialogue systems
11. Machine Translation
12. Information Retrieval
13. Applications for Public Use
14. Computer-Assisted Language Learning
15. Applications for Disability and Access
16. Conclusions: Considerations for Implementation

Part III

Two illustrative samples


Dictionaries

- Already well on the road
- Extensively used
- > 2.7million users
- Two flagship projects
 - teanglann.ie
 - focloir.ie
- Need for Irish-Irish and new Irish-English dictionary


ABAIR.ie



- Speech technology for Irish
- Extensively used
- 500,000 visits: first 3-months of 2019
- A global community materialises
 - 50% from abroad
- TTS for the main dialects
- Research towards ASR ongoing

Dictionaries: teanglann.ie

Baile  Foclóir Nua Béarla–Gaeilge » Eolas ▼ Gaeilge English

Cuardaigh focal Gaeilge nó Béarla.

 LEABHARLANN TEANGA AGUS FOCLÓIREACHTA

AN DRAOI GRAMADAÍ

AINMFHOCAL + AIDIACHT ▼

1 ROGHNAIGH AINMFHOCAL

2 ROGHNAIGH AIDIACHT

+

Leabharlann Teanga agus Foclóireachta

Fáilte go dtí an *Leabharlann Teanga agus Foclóireachta* ar líne atá á forbairt ag Foras na Gaeilge mar thionscadal comhthreomhar le tionscadal an *Fhoclóra Nua Béarla–Gaeilge*. Is é cuspóir an tsuimh ná rochtain shaor agus réidh a thabhairt do lucht úsáidte na teanga ar shonraí foclóireachta, gramadaí agus foghraíochta a bhaineann le focail sa Ghaeilge. [Breis eolais...](#)

Ar an suíomh seo

 [Foclóir Gaeilge–Béarla \(Ó Dónaill, 1977\) »](#)



Dictionaries: focloir.ie

English > Irish



[Advanced Search](#)
English-Irish Dictionary (1959)
Foclóir Gaeilge-Béarla (1977)



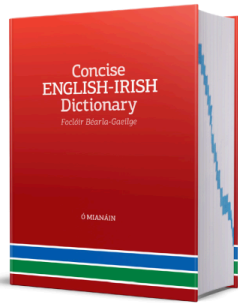
New English-Irish Dictionary



eNEID widgets

Welcome

Welcome to [Foras na Gaeilge's](#) New English-Irish Dictionary. [More information...](#)



Buy the book

Concise English-Irish Dictionary now available!

The brand new *Concise English-Irish Dictionary* is now available. For more information, [click here](#).

Index

- [Welcome](#)
- [Browse Dictionary](#)
- [Tools](#)
- [About the Dictionary](#)
- [FAQ](#)
- [Accessibility](#)
- [Our App](#)
- [Book](#)
- [Education](#)

Word of the Day

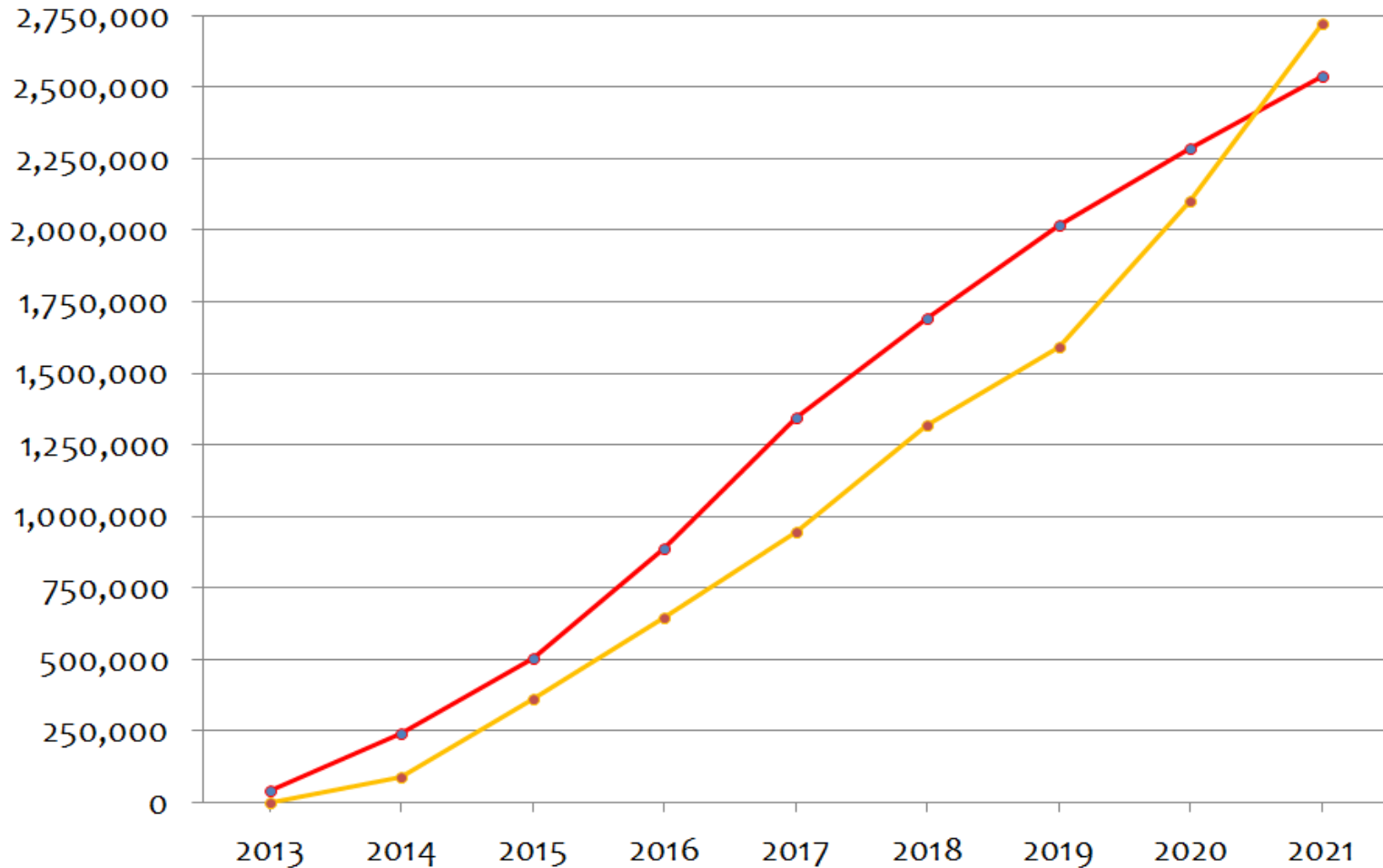
beer garden
noun: [gairdín beorach](#)

Get our app for free!

The New English-Irish Dictionary app is available free of charge for both Apple and Android devices.

Number of Users

Foclóir.ie Teanglann.ie



- 30% overseas (16% USA, 8% Britain)
- 7000+ copies of the published English-Irish dictionary sold since Sept. 2020

Speech Technology: abair.ie

Multidialect synthetic voices

The screenshot displays the abair.ie website interface. At the top, a dark navigation bar contains the 'abair.ie' logo on the left and menu items: 'Home', 'About', 'Resources', 'Accessibility', 'Contact', and 'Mile Glór'. Below this is a secondary navigation bar with tabs for 'Synthesis', 'Dialects', 'Voices', and 'Info'. The 'Synthesis' tab is active, leading to a page titled 'Synthesis'. This page features a dialect selector with 'Gweedore', 'Connemara', and 'Dingle' options. A large text input area is provided with the instruction 'Type here or click on 'Say it!'' and a keyboard layout showing Irish vowels: 'Á á É é Í í Ó ó Ú ú'. Below the input area are four dropdown menus for 'Mode' (set to 'DNN'), 'Speed' (set to 'normal'), 'Pitch' (set to 'normal'), and 'Speaker' (set to 'female'). A prominent green 'Say it!' button is located at the bottom of the interface.

Abair: speech technologies for Irish

- places the living, spoken language centre stage in the digital world
 - *a public utility for the wider Irish language community, including the diaspora*
 - *disability/access*
 - *education*
 - *language maintenance and documentation*
 - *Linguistic research on the Irish dialects*
- by embedding synthesis into applications, we extend the everyday domains where we use and are exposed to spoken Irish

Summary

Community: development in close collaboration with language community

- needs prioritisation, user-guided design, community input (e.g. crowdsourcing), application testing & evaluation, etc.
- embracing Gaeltacht and non-Gaeltacht communities

Interdisciplinary teams, with

- high proficiency in Irish and Irish linguistic structure
- expertise in speech and language technology building
- expertise in domains of application (education, disability, etc.)

A long term vision...

Goal

- to build an indigenous Irish speech and language technology sector into the future
- need for infrastructure building, including the education of interdisciplinary research and development groups with high competence in Irish
- collaborating with similar groups in other countries

Go raibh míle maith agaibh!
neasa.nichiarain@tcd.ie



Foras na Gaeilge



An Roinn Turasóireachta, Cultúir,
Ealaíon, Gaeltachta, Spóirt agus Meán
Department of Tourism, Culture,
Arts, Gaelacht, Sport and Media

S20

PLEAN GNÍOMHAÍOCHTA 2018-2022

Ár dTeanga, Ár bPobal

S20

STRAITÉIS 20 BLIAIN DON
GHAELIGE 2010-2030

Ár dTeanga, Ár bPobal