# Accessibility for Foreign Language and Linguistics

Elizabeth J. Pyatt, Ph.D.
elizabeth.j.pyatt@gmail.com

Outline

■ About Me

■ Language Tagging

■ Unicode Fonts

■ Linguistic Notation and Tables
  - Tree Diagrams (this session)
  - Gloss Tables
  - Grammar Tables (no cell merges)
  - Vowel Diagrams
  - Phonetic Inventories
 - Phonetics ≠ MathML
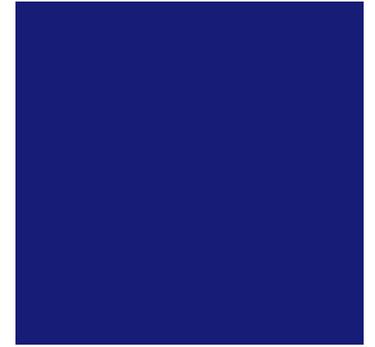
About Me

- Degree in Linguistics
  - Celtic language dissertation

- Instructional Designer
  - Interests in Unicode
  - Manager advocated accessibility (2000)

- Accessibility
  - Math, Spanish, STEM, PDF, PPT, Word, HTML, …

Language Tagging
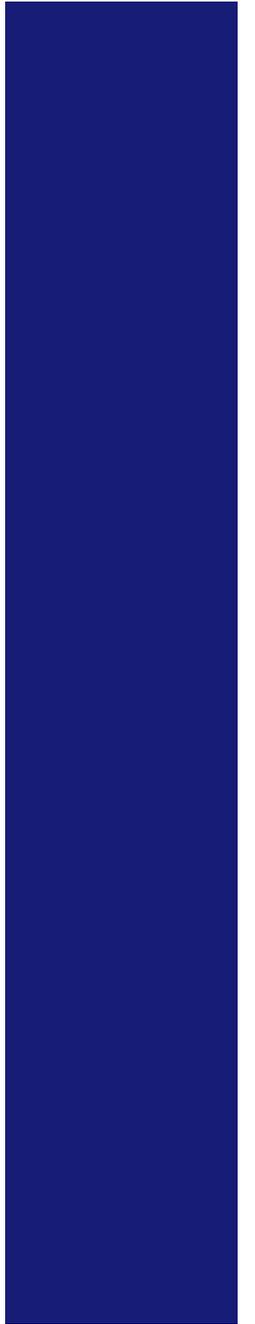
Language Tagging Controls
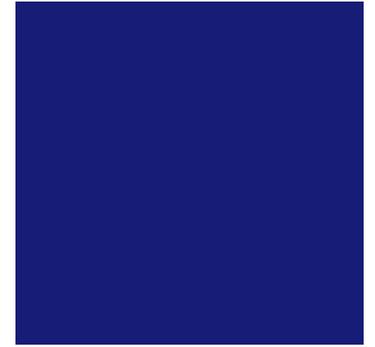
- Metadata on document language
  - Pronunciation engine in TTS
  - Cataloging
  - Controls spell checking in Microsoft
    - Le français est une langue indo-européenne de la famille des langues romanes dont les locuteurs sont appelés «francophones». –fr.wikipedia.org

Does it REALLY matter?
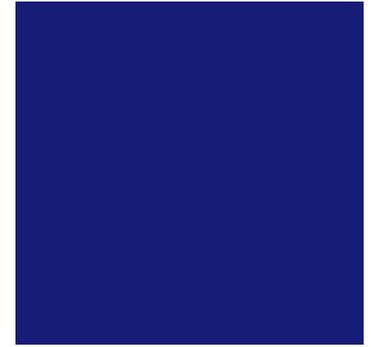
- Part of WCAG 2.2
  - A: Set document language
  - AA: Set section language

- Heteronyms (Span/English)
  - A, van, saber, once, quince, Chile
  - A, van, saber, once, quince, Chile

- International Screen Reader Audience
  - Need to switch languages

Audio Demo Spanish in JAWS
Audio reads part of text below

Browser window showing page E:\aLangTest.html

Language Tag Tests

Spanish Tagging

Only the second list item is tagged as being in Spanish.

- ¡Hola mundo! – bienvenidos a Denver. La Reina Letizia realizará este viernes, 7 de noviembre, su segundo viaje oficial en solitario fuera de nuestras fronteras, tras estrenarse en Viena la semana pasada, y esta vez su destino será un país que, aseguran, la tiene completamente fascinada: Portugal.
- ¡Hola mundo! – bienvenidos a Denver. La Reina Letizia realizará este viernes, 7 de noviembre, su segundo viaje oficial en solitario fuera de nuestras fronteras, tras estrenarse en Viena la semana pasada, y esta vez su destino será un país que, aseguran, la tiene completamente fascinada: Portugal.

Heteronyms (VoiceOver)
Audio reads text below

- Tacky...with a Twist
  - Letizia Ortiz Rocasolano (Oviedo, 15 de septiembre de 1972) es la reina consorte de España por su matrimonio con el rey Felipe VI.

- Incorrect
  - México, Argentina

- Problematic
  - No dice, once, saber, van, hay, a vs.
  - No dice, once, saber, van, hay, a

Implementation

- Microsoft Word
  - Set (Proofing) Language
  - Exports LANG tag to PDF

- HTML
  - <html, div, p, span, i...>
    - lang="en, es, fr, de,..."

- ISO-639 Standards
  - ISO-639 (2 letters),
  - ISO-639-3 (3 letters, ext)
    - sgn = Signed languages (some families)
    - ase = American Sign Language
  - ISO-693-5 for language families
    - Romance = roa or ine:itc:roa

Multilingual and Code Switching

- "The present indicative is used to express actions or states of being in a present time frame. For example:"
  - *Soy alto* (I am tall). (Subject pronoun "*yo*" not required and not routinely used).
  - *Ella canta en el club* (She sings in the club).

- juuyeah RT @whatsupleticia*: Sinhá Moça* is a brazilian novela, que the white people maltrata the black people, rialto

Minority Languages?

- Language codes exist
  - cy (Welsh), ang (Old English), sw (Swahili), tt (Tatar), chr (Cherokee)

- Dialects (en-US, en-GB, en-CA)

- Macrolanguage
  - Chinese (zh) or "Sinitic Languages" (zhx)
    - Mandarin (cmn), Cantonese (yue), Wu (wuu),…
  - sr (Serbian), hr (Croatian), bs (Bosnian), cnr (Montenegrin)
    - sh (Serbo-Croatian) or hbs ("Serbo-Croatian" the family?)

- Specify script (ISO 15924)
  - tt-Latn vs tt-Cyrl
  - hi-Latn, ru-Latn (?), en-Brai (in BRF files)
  - IPA?

Localization Standards
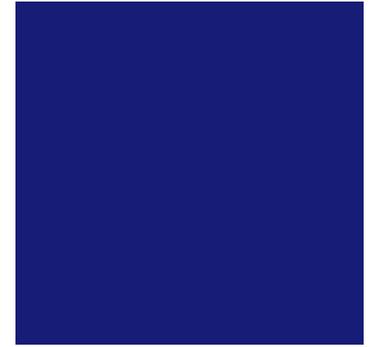
■ Keyboards for different scripts

■ Voice Recognition

■ Voice Synthesis (for screen readers and more)

■ Optical Character Recognition

■ Localization (X in language Y)

■ Welsh Examples
  - [GALLU Speech Recognition](#)
  - [Welsh TTS Voices](#)

Unicode Fonts
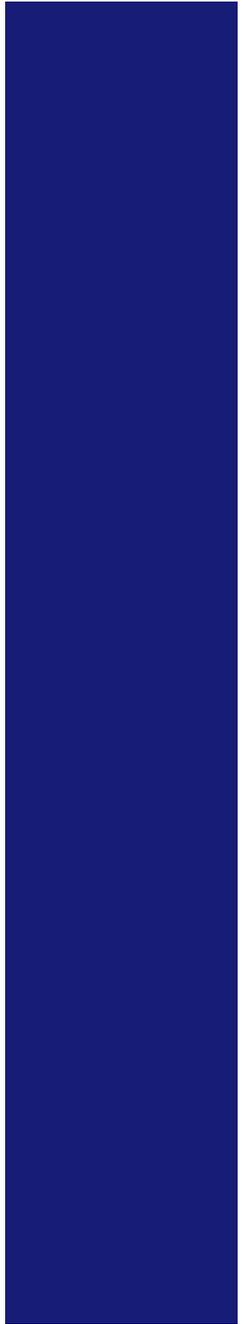
Unicode Standard

- ASCII...but for every character
  - Assigns a number to all characters in all scripts...including emojis, phonetic symbols, math
    - 0-127 = ASCII
    - 128-255 = Latin 1
    - Up to millions
- Enables email, Wikipedia, social media...

Tatar (/ˈtɑːtər/ "TAH-tər" **татар теле**, tatar tele or **татарча**, tatarça) is a Turkic language spoken by the Volga Tatars mainly located in modern Tatarstan (European Russia), as well as Siberia.

–*Wikipedia*

3 Scripts

- **Татар теле** (Cyrillic) tatar tele (Latin) تاتار تلی • تاتار تئلئ (Persian)

Before Unicode

Including many PDFs

- Just ASCII or "8-bit"

- Specialty ASCII Fonts
  - Regular – **a b c d**
    - = a b c d (ASCII/screen reader)
  - Greek ("Symbol") – **α β γ δ θ ξ**
    - = a b g d q j
  - Phonetic - **ə ʃ ʒ ʔ**
    - = @ S Z ?
  - Braille ASCII (BRF) - **,,abc #abc**
   - = ,, abc #abc (=123)
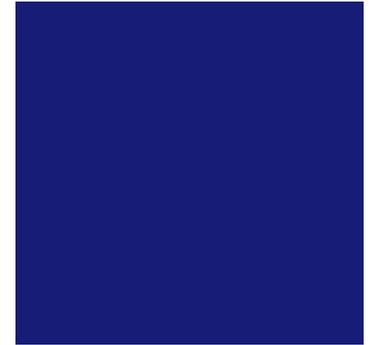
Input Non-English Unicode

- Keyboard Codes
  - E.g. ¿ = ALT+0191 (Win) or OPT+? (Mac)

- Keyboard Utilities
  - International Keyboard (Win) or other lang
  - ABC Extended (Mac)
  - Russian QWERTY/Phonetic or "native"
    - Most major scripts

- Keyboard Shorteners (e.g. aText)
  -]d = δ \d = ð ;all = ∀

- Stray Characters
  - Character Map (Win)
  - Emoji & Symbols (Mac)
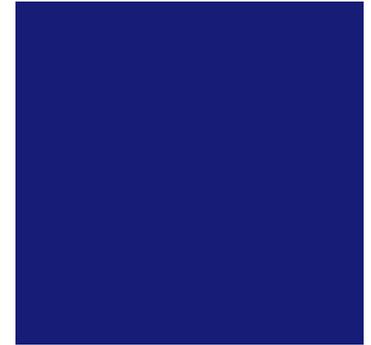
Keyboard Setup

- Available to ALL users
  - [Windows keyboard utilities](#)
  - [Mac keyboard utilities](#)

- QWERTY/Phonetic vs. Native layouts
  - Arabic Qwerty, etc.
    - E.g. Arabic letters in English keyboard positions
    - Even East Asian languages
  - Native
   - Keyboard mimics Arabic typewriter
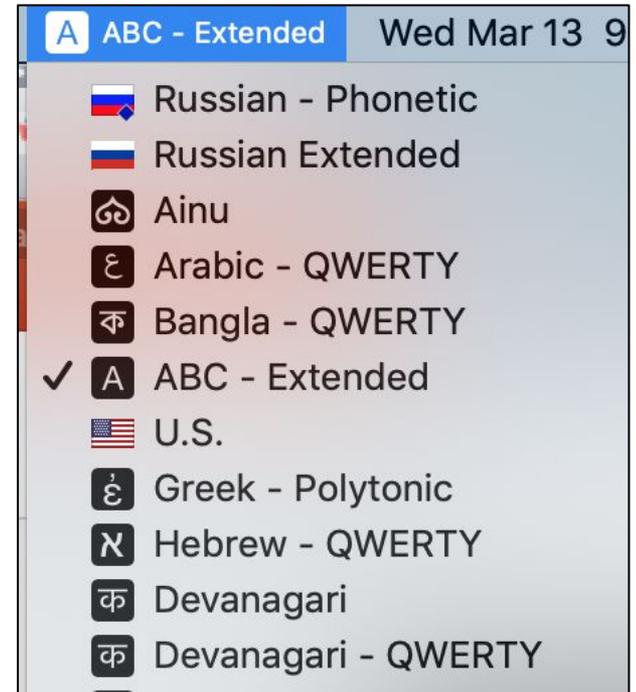
Is it Unicode? (PDF)

- Test:
  - Copy sample text and paste it into another file
  - It should have the same characters

- Gotchas
  - Differ, influence → "di2er, in¯uence"
  - Coptic (Biblical)
    - ⲬⲢⲎⲤⲦⲞⲤ (Χριστός) → CrHstos (Χριστός)
    - ⲤⲈⲐⲚⲞⲤ (ἔθνος)→ heqnos (ἔθνος)
 - ʃ ʒ θ ? → "Z S T 6"
  - ∀ Ǝ → "Ð ÷"

Is it even text?


FIGURE 2
New Jersey Counties and Communities Surveyed

It used to be...

FIGURE 2
New Jersey Counties and Communities Surveyed
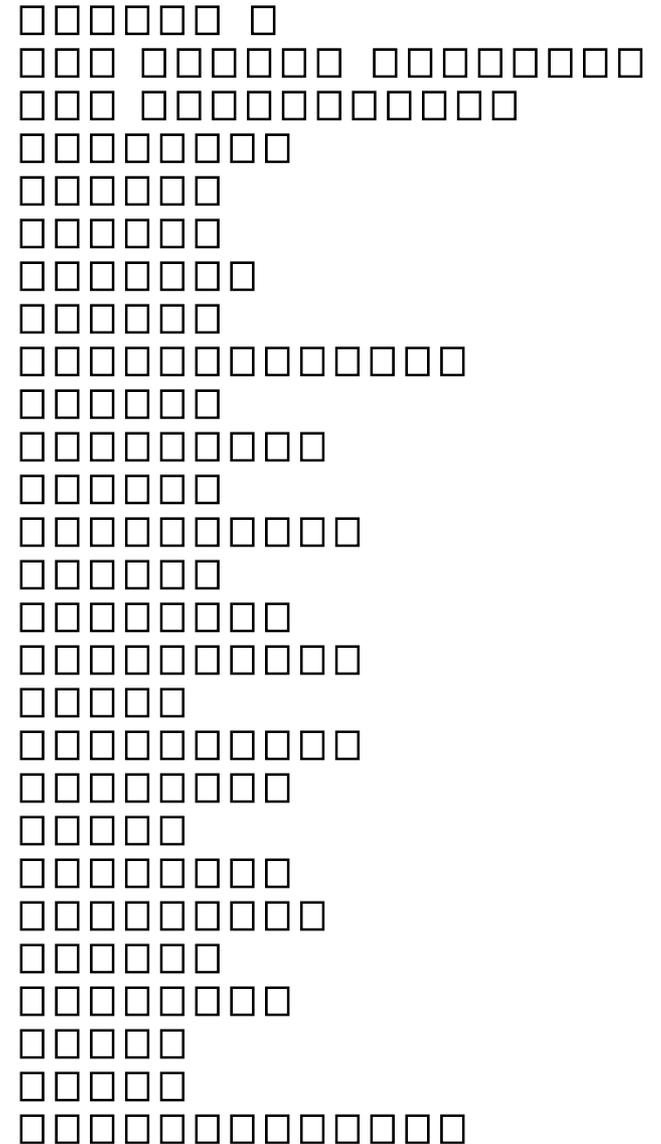
Some Encoding Sites

■ Unicode.org

■ Script Encoding Initiative (Berkeley)

■ MUFI.info (Medieval Unicode Font Initiative)

■ (Google Search)

■ Many offer free fonts
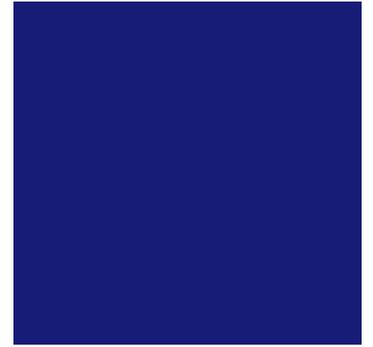
Linguistic Tables and Trees
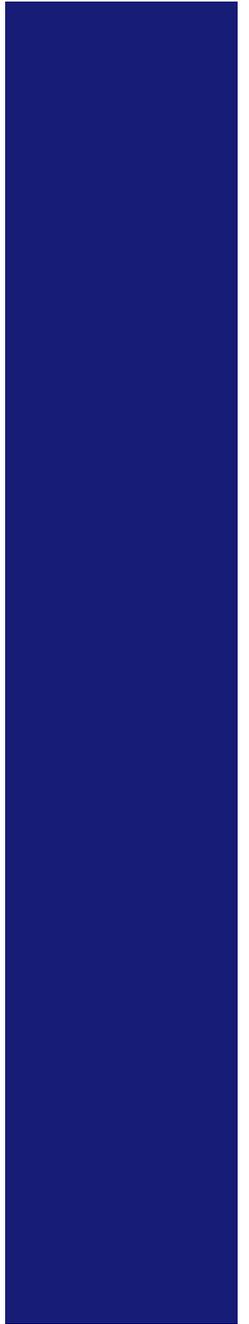
Tree Diagram
Long Description Required
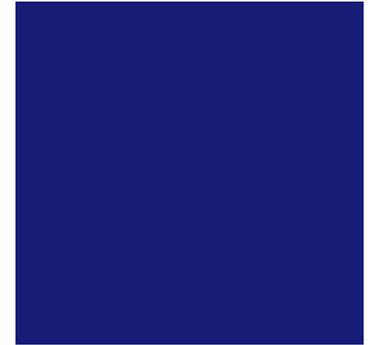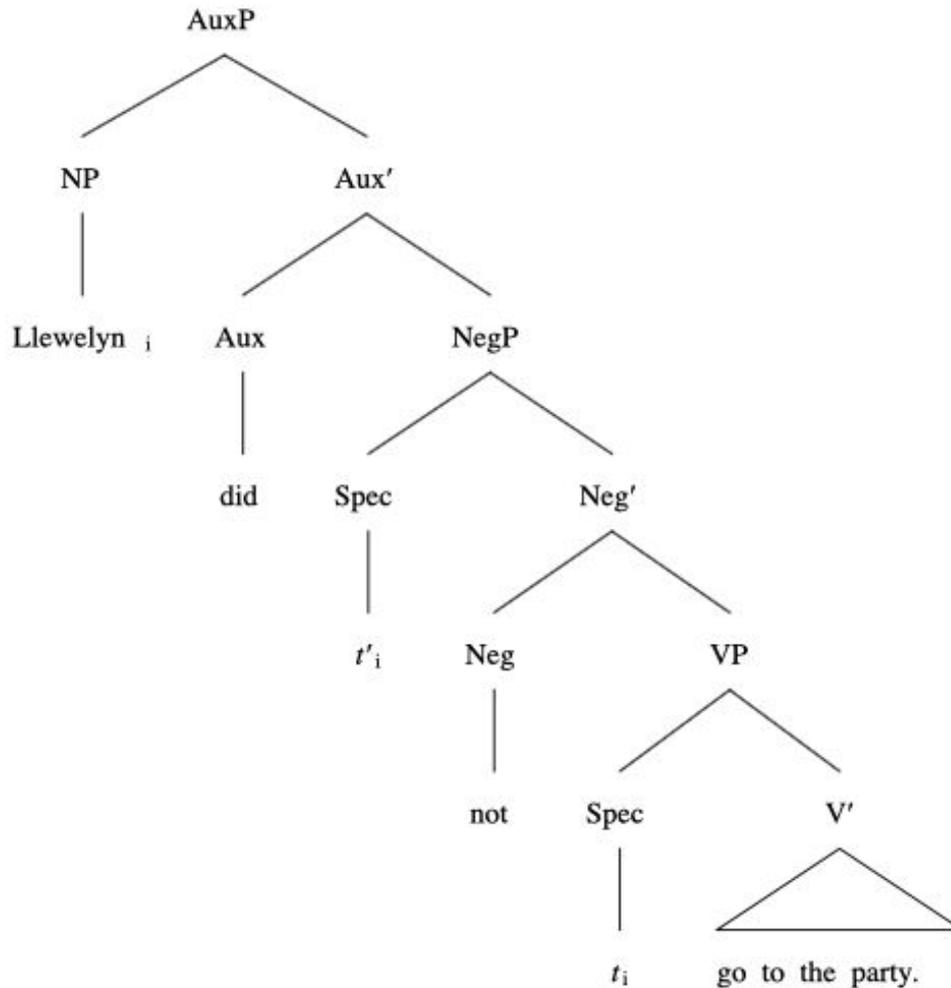
Alternate Formats

Linear

[$_{AuxP}$ [$_{NP}$ Llewelyn$_i$][$_{Aux'}$[$_{Aux}$ did]
[$_{NegP}$ [$_{Spec}$ t'$_i$] [$_{Neg'}$[$_{Neg}$ not]
[$_{VP}$ [$_{Spec}$ t$_i$—] [$_{V'}$ go to the party.]]]]]]

Stacked

[$_{AuxP}$ [$_{NP}$ Llewelyn$_i$]
  [$_{Aux'}$[$_{Aux}$ did]
    [$_{NegP}$ [$_{Spec}$ t'$_i$]
      [$_{Neg'}$ [$_{Neg}$ not]
        [$_{VP}$ [$_{Spec}$ t$_i$—]
         [$_{V'}$ go to the party.]]]]]]

R Syntax Tree
yohasebe.com/rsyntaxtree
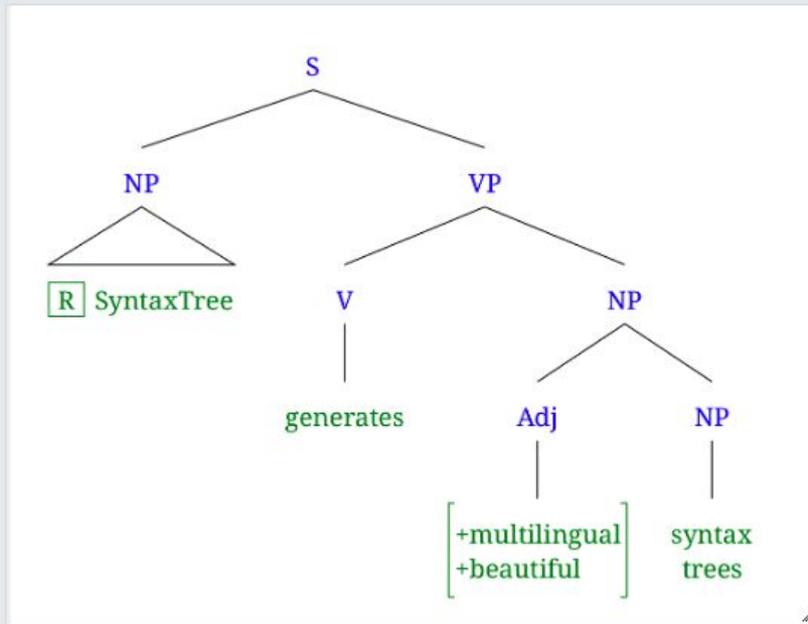
Gloss Tables

Don't use Tabs

- Wnaethon ni ddim mynd i'r parti. (Welsh)
  - We didn't go to the party.

Wnaethon	ni	ddim	mynd	i'r	parti
Did.1p	we	not	go.VN	to'the	party

'We didn't go to the party.'

Gloss Tables Reformatted

Hide the borders

- Wnaethon ni ddim mynd i'r parti. (Welsh)
  - We didn't go to the party.

Wnaethon	**ni**	**ddim**	**mynd**	**i'r**	**parti**
Did.1p	we	not	go.VN	to'the	party

'We didn't go to the party.'

Vowel Inventory: Add Labels

No tabs, and don't merge cells

Latin Vowels (Tabs)

i   u

e   o

    a

Latin Vowels (Table)

Place	Front	Central	Back
High	i	--	u
Mid	e	--	o
Low	--	a	--

Inaccessible Grammar Table
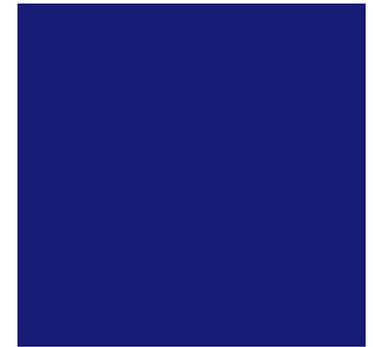
Case	feminine		masculine		neuter	
	1 sg.	1 pl.	2 sg.	2 pl.	2n sg.	2n pl.
Nominative	puella	puellae	dominus	dominī	bellum	bella
Vocative	puella	puellae	domine	dominī	bellum	bella
Accusative	puellam	puellās	dominum	dominōs	bellum	bella
Genitive	puellae	puellārum	dominī	dominōrum	bellī	bellōrum
Dative	puellae	puellīs	dominō	dominīs	bellō	bellīs
Ablative	puellā	puellīs	dominō	dominīs	bellō	bellīs

Another Inaccessible Table
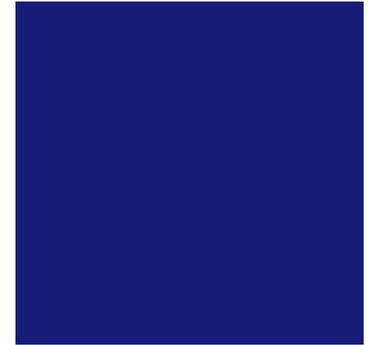
|			
1st sg.	*I*	**amō**	I love
2nd sg.	*you sg.*	**amās**	you sg. love
3rd sg.	*he, she, it*	**amat**	he, she, it loves
1st pl.	*we*	**amāmus**	we love
2nd pl.	*you pl.*	**amātis**	you pl. love
3rd pl.	*they*	**amant**	they love

Ideally...

- Split large table into separate entities

- Recode labels

- Include headers

- Ensure coherent data reading order

- Don't use blank rows/columns for spacing

- DO NOT MERGE CELLS!!
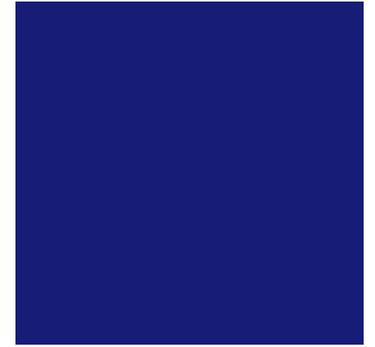
MathML <> Phonetics

- Phonetic characters not recognized

- Language tagging manual

- Need to use <mtext> or \text{} functions for [−nasal], etc.

- Recommend plain text or images instead

- Insert true Unicode minus sign as needed
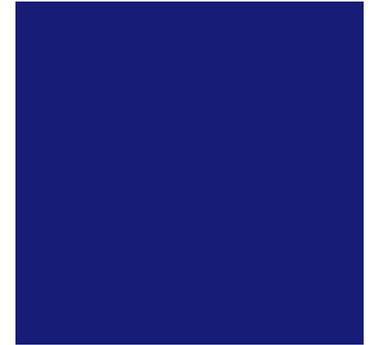
Plain Text: English Plurals

1. $\emptyset \rightarrow \text{ə} / \text{C}_{[+\text{strident}]} \quad /\text{z}/_{[+\text{plural}]}$

   a. cruises, vices, judges, churches, massages, foxes, ashes

2. $/\text{z}/_{[+\text{plural}]} \rightarrow [-\text{voice}] / \text{C}_{[-\text{voice}]} \quad \#$
   $/\text{z}/_{[+\text{plural}]} \rightarrow [\text{s}] / \text{C}_{[-\text{voice}]} \quad \#$

   a. kicks, pats, taps, cliffs, sixths, Bachs

The End

- Questions?