



## Dubsar: A New Approach to Rendering TEI Documents and Non-Western Text

Kevin Graaf  
9 October 2024

Dubsar: Un nuevo acercamiento a renderizar  
documentos TEI y textos no occidentales

9 de octubre de 2024

# About Me

My name is Kevin Graaf.

I'm a computer scientist interested in...

- Computer Typography
- Emerging, minoritized, and ancient writing systems
- Programming Language Theory

# Sobre mi

Mi nombre es Kevin Graaf.

Soy informático interesado en...

- Tipografía informática
- Sistemas de escritura nuevos, minorizados y antiguos
- Teoría de lenguajes de programación

# About This Presentation

Generated from a TEI document (more on this later)

Rendered using a Dubsar font

No so-called "AI" was used to produce this work

# Sobre esta presentación

Generado a partir de un documento TEI (más información más adelante)

Renderizado con una fuente Dubsar

No se utilizó la llamada "IA" para producir esta obra

# The Status Quo

## OpenType

- Nearly universal font format
- Outline based
- Typically also contains metadata for kerning or ligatures

# El estatus quo

## OpenType

- Formato tipográfico casi universal
- Basado en el esquema
- Suele contener metadatos sobre interletrado o ligaduras

# The Status Quo (Continued)

Approximately three phases in a typical text rendering pipeline:

- Bidirectional Segmentation – e.g. GNU FriBiDi
- Shaping – e.g. HarfBuzz
- Rasterization – e.g. FreeType

## El estatus quo (continuación)

Aproximadamente tres fases en un proceso típico de renderizado de texto:

- Segmentación bidireccional – ej. GNU FriBiDi
- Modelado – ej. HarfBuzz
- Rasterización – ej. FreeType

# Limitations

- Vertical text (among other things) is hard
- Shaping model has to be built into OpenType
- Deeply complex (cumulatively hundreds of thousands of lines of code)
- Little to no abstraction

# Limitaciones

- El texto vertical (entre otras cosas) es difícil
- El modelo de configuración debe ser construido en OpenType
- Profundamente complejo (acumula cientos de miles de líneas de código)
- Poca o ninguna abstracción

# Three Myths...

...that Computer Scientists Believe About Text

# Tres mitos...

...que los informáticos creen acerca del texto

# Myth 1:

MYTH:

There are no new writing systems

FACT:

Dozens of new writing systems have been created over the past century

# Mito 1:

MITO:

No hay nuevos sistemas de escritura

VERDAD:

Se han creado docenas de nuevos sistemas de escritura en el último siglo



## Myth 2:

MYTH:

Writing is a linear sequence of symbols, written left to right  
(sometimes right to left), top to bottom

FACT:

Counterexamples exist for each of these assumptions

## Mito 2:

MITO:

La escritura es una secuencia lineal de símbolos, escritos de izquierda a derecha  
(a veces de derecha a izquierda) y de arriba abajo

VERDAD:

Existen contraejemplos para cada uno de estos supuestos

## Myth 3:

MYTH:

Computer text is just printed text or handwriting transposed to a screen

FACT:

Computer text is an entirely new modality of writing and language

## Mito 3:

MITO:

El texto informático no es más que texto impreso o manuscrito trasladado a una pantalla

VERDAD:

El texto informático es una modalidad totalmente nueva de escritura y lenguaje

# Introducing Dubsar

A new text and typography engine for rendering e.g. TEI documents

Named after the Sumerian word for scribe

Every font is a program

Every glyph is a function

Stroke rather than outline based

## Presentación de Dubsar

Un nuevo motor de texto y tipografía para renderizar ej. documentos TEI

Recibe su nombre de la palabra sumeria para escriba

Cada fuente es un programa

Cada glifo es una función

Trazos en lugar de contornos

# The Big Ideas

Each glyph function has access to its graphical context

This graphical state can be arbitrarily modified

Paths can be directly manipulated as first-class data

Glyphs can dynamically compute optimal spacing, i.e. kerning

Each glyph is responsible for placing the cursor when it returns

Conclusion: a Dubsar font program fully embodies  
the underlying logic of a writing system

# Las grandes ideas

Cada función de glifo tiene acceso a su contexto gráfico

Este estado gráfico puede ser modificado arbitrariamente

Las rutas pueden manipularse directamente como datos de primera clase

Los glifos pueden calcular dinámicamente el espacio óptimo, es decir, el interletraje

Cada glifo es responsable de ubicar el cursor cuando vuelve

Conclusión: Un programa de fuentes Dubsar representa  
plenamente la lógica subyacente de un sistema de escritura

## Example/Ejemplo ✱

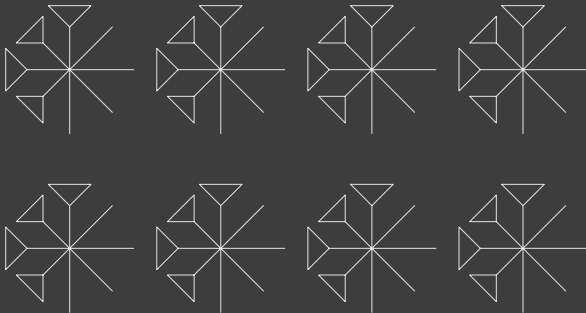
```
function wedge(rot: int, size: real)
tip: path := (0,5) lineto (4,5) lineto (2,4) close
len: path := (2,4) lineto (2,5)
wed: path := tip .. len
```

```
return scale(rotate(wed, rot), size)
end
```

```
glyph U+1202D
draw wedge(0, 1.0)
draw wedge(-45, 0.8)
draw wedge(-90, 1.0)
draw edge(-135, 0.8)
```

```
return 16, 16, 0
end
```

# Result/Resultado



# Case Study: TEI

These slides were generated from a TEI document

Inspired by "difficult" TEI features such as writing direction and custom glyphs

All text you see has been rendered with Dubsar

Latin glyphs are based on public-domain Hershey Fonts

Non-Latin glyphs are all my own design

## Estudio de caso: TEI

Estas diapositivas se generaron a partir de un documento TEI

Inspirado en características "difíciles" de TEI, como la dirección y glifos personalizados

Todo el texto que ves ha sido renderizado con Dubsar

Los glifos latinos se basan en las fuentes de Hershey de dominio público

Los glifos no latinos son de diseño propio

## Case Study: Maya

Yaxchilan Lintel 14 (Transcription, transliteration, and translation from mesoweb.com)

(On the day) 4 Imix (on) the 4th of Mol, K'awiil Kab(?) Muwaan Chanal  
Chak Bay Kaan, the wahy(?) of ...1 Ixik K'uh Ix Chak Joloom Sajal  
and mother of the king, was conjured.

## Estudio de caso: Maya

Yaxchilán Dintel 14 (Transcripción, transliteración y traducción de mesoweb.com)

(El día) 4 Imix, (el) 4º día de Mol, se conjuró a K'awiil Kab(¿) Muwaan Chanal  
Chak Bay Kaan, wahy(?) de ...1 Ixik K'uh Ix Chak Joloom sajál y madre del rey.



# Transcription/transcripción

A1 B1 C1

A2 B2 C2

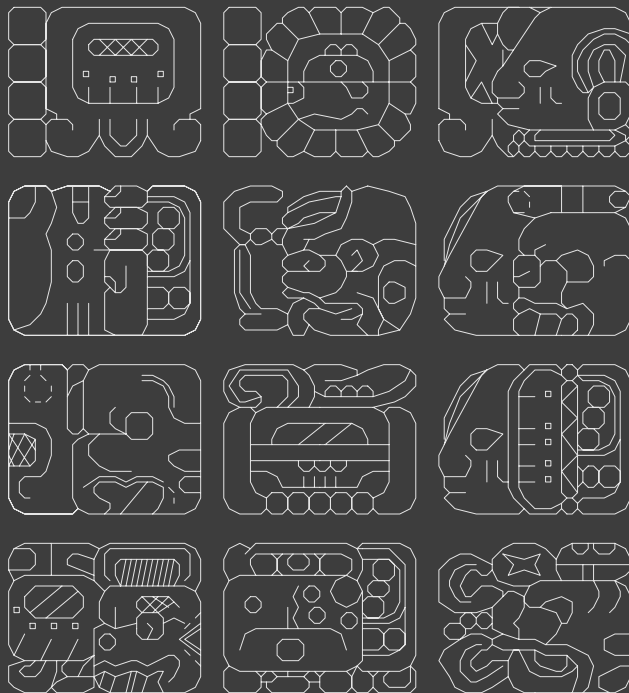
A3 B3 C3

A4 B4 C4

(A1) 4-''IMIX'' (B1) 4-mo[lo] (A2) TZAK-ja (B2) K'AWIIL (A3) KAB?-MUWAAN  
(B3) [CHAN] NAL-la (A4) CHAK-ba-ya ka-KAAN (B4) u-WAY-ya-ja-la  
(C1) T1001-IXIK-K'UH (C2) IX-CHAK-JOL (C3) IX-sa-ja (C4) ya-ANATZ'?'-AJAW chan

# Transliteration/transliteración

imix chan mol tza[h]kaj k'awiil kab? muwaan chanal chak bay kaan  
uwayjal?...[l] ixik k'uh ix chak jol[oom] ix saj[al] yanatz'? ajaw



# Future Work

Formalization of the virtual machine

Improved graphics model

Public release in 6 months to 1 year  
(depending on funding, etc)

## Trabajo a futuro

Formalización de la máquina virtual

Modelo gráfico mejorado

Publicación en un plazo de 6 meses a 1 año  
(dependiendo de la financiación, etc)

# Questions?

<https://kbg.sh/Projects/Dubsar/>

# ¿Preguntas?