





AZRAEL

A-Z Reconocedor Automático de Español

Elena Álvarez Mellado

Licenciatura de Lingüística
Universidad Complutense de Madrid





AZRAEL

A-Z Reconocedor Automático de Español

¿Es esto español?

Escoba

Troglodita

Trasiego

Berenjena

Pespunte



AZRAEL

A-Z Reconocedor Automático de Español

¿Y esto?

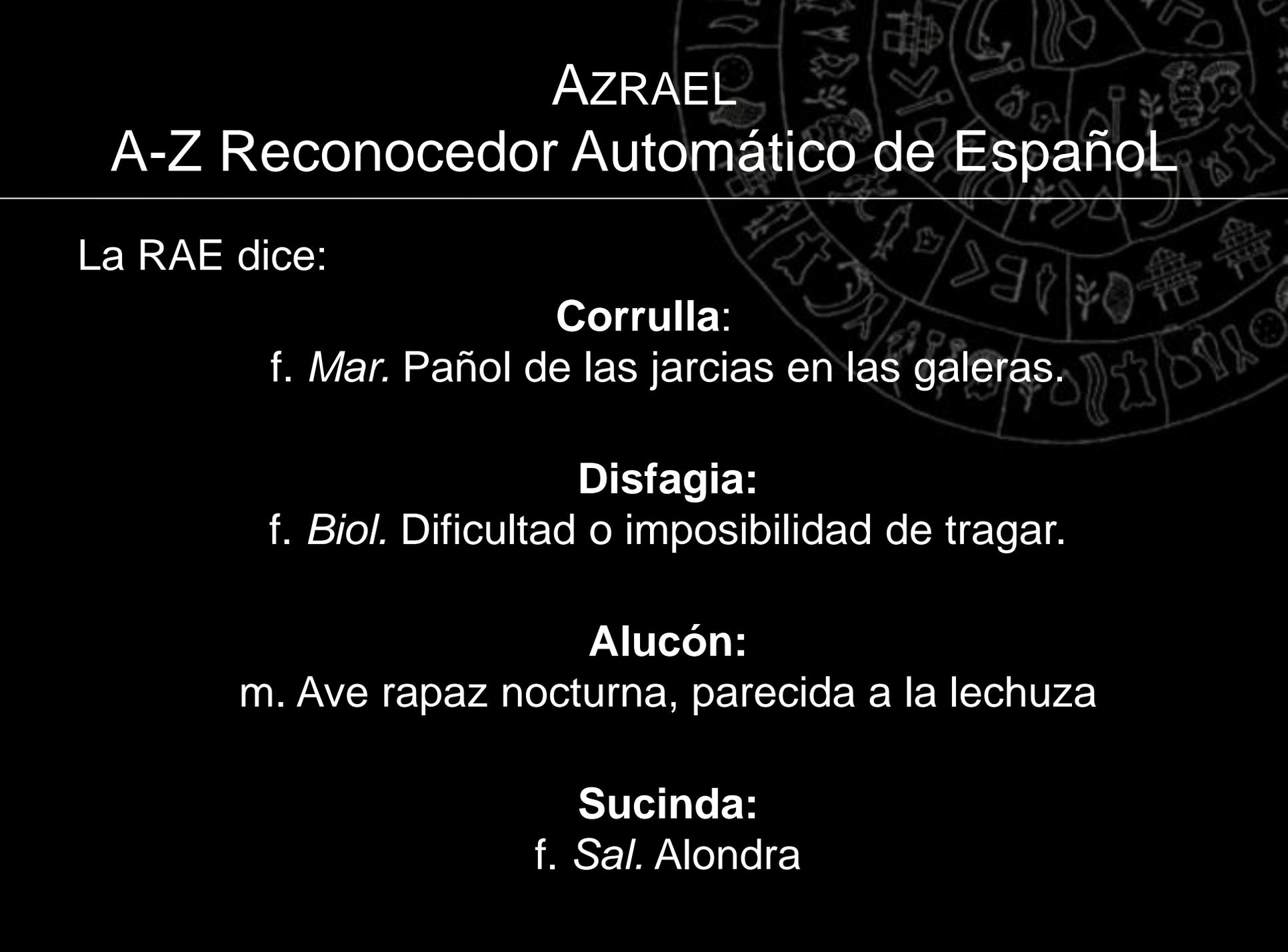
¿Corrulla?

¿Alucón?

¿Sucinda?

¿Clánide?

¿Disfagia?



AZRAEL

A-Z Reconocedor Automático de Español

La RAE dice:

Corrulla:

f. *Mar.* Pañol de las jarcias en las galeras.

Disfagia:

f. *Biol.* Dificultad o imposibilidad de tragar.

Alucón:

m. Ave rapaz nocturna, parecida a la lechuza

Sucinda:

f. *Sal.* Alondra



AZRAEL

A-Z Reconocedor Automático de Español

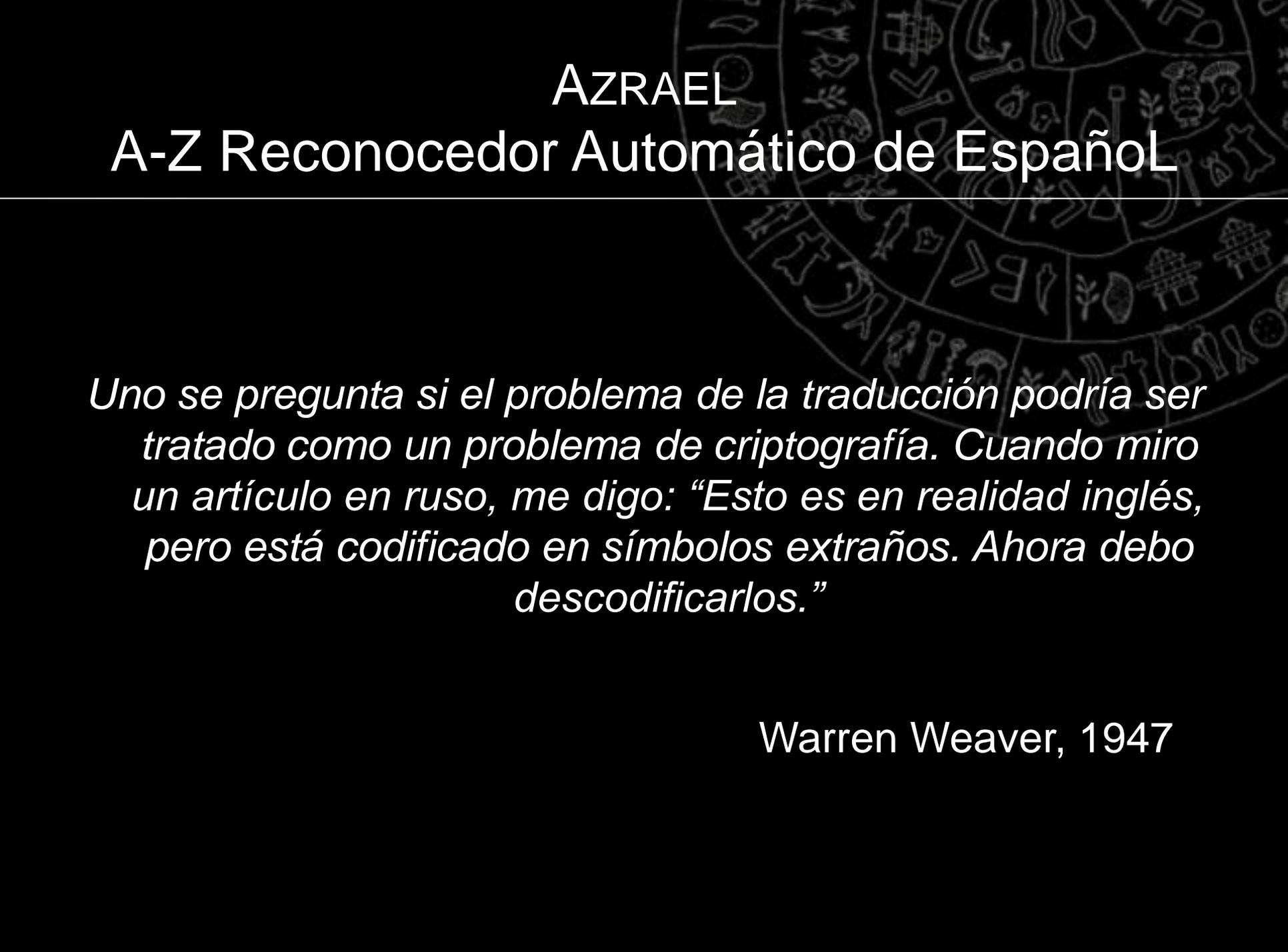
¿Y si en vez de *clánide* hubiese sido...

... *trock*?

... *shippet*?

... *whert*?

¿Qué pasa con *whisky*?

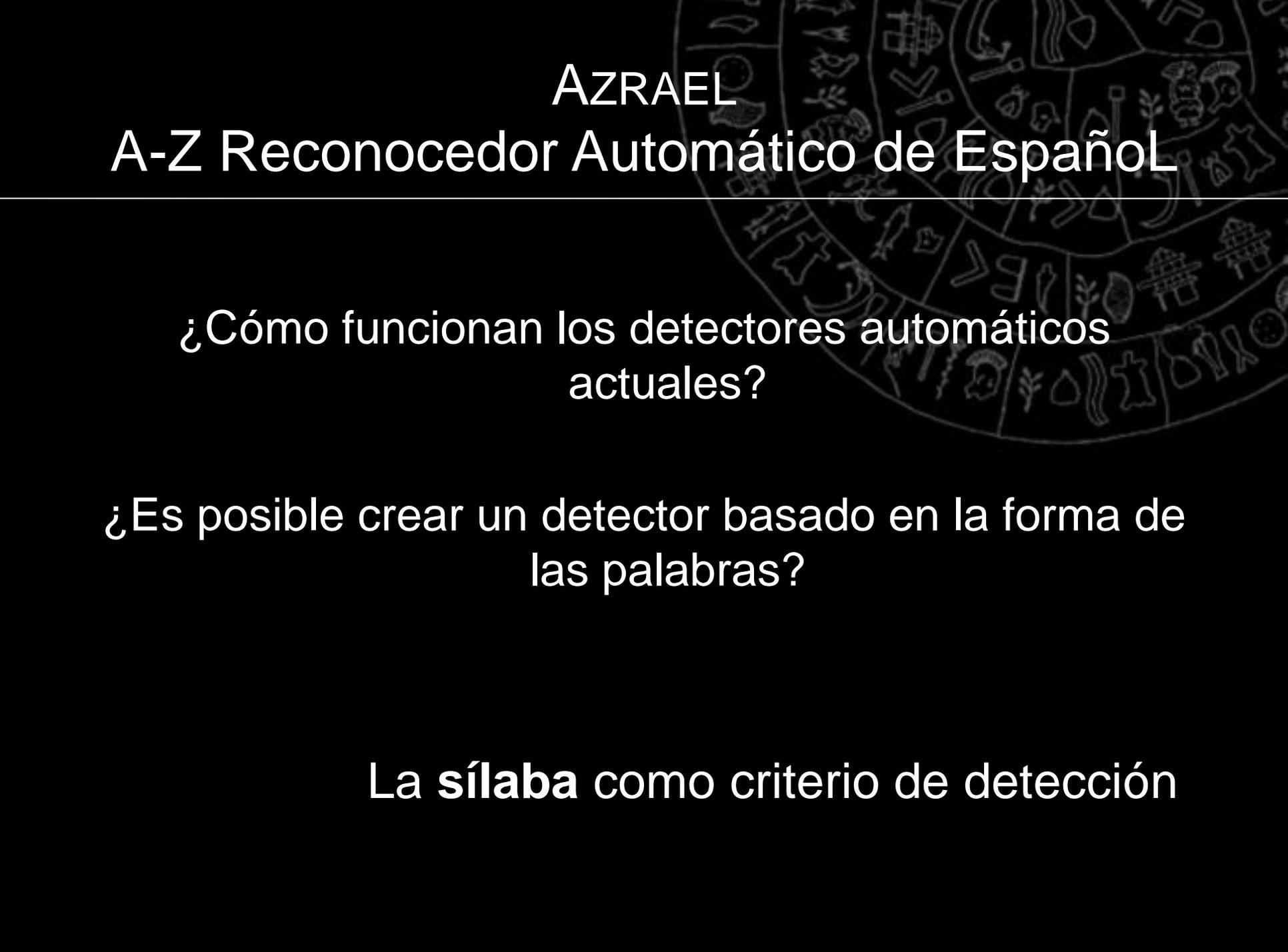


AZRAEL

A-Z Reconocedor Automático de Español

Uno se pregunta si el problema de la traducción podría ser tratado como un problema de criptografía. Cuando miro un artículo en ruso, me digo: “Esto es en realidad inglés, pero está codificado en símbolos extraños. Ahora debo descodificarlos.”

Warren Weaver, 1947



AZRAEL

A-Z Reconocedor Automático de Español

¿Cómo funcionan los detectores automáticos actuales?

¿Es posible crear un detector basado en la forma de las palabras?

La **sílaba** como criterio de detección

AZRAEL

A-Z Reconocedor Automático de Español

Ventajas de la sílaba:

- a. No dependemos de un diccionario
- b. Detección de jerga, apreciativos, palabras inventadas, etc.
- c. Estabilidad
- d. Válido para otros idiomas

scanner → *escáner*

ski → *esquí*

jerez → *sherry*



AZRAEL

A-Z Reconocedor Automático de Español

Estudio de la sílaba en español

- No existen estudios exhaustivos sobre el tema
- Hemos realizado nuestro propio estudio a partir de un corpus del español de 600.000 palabras.
- ¿Qué combinaciones están permitidas? ¿Cuáles están prohibidas?

AZRAEL

A-Z Reconocedor Automático de Español

Tipos de consonantes

Emparejadas:

B C F G C P → R, L

Ejemplos: *bri-llo blan-co*

T D → R

Ejemplos: *tren, drui-da*

Otras parejas: CH, LL, QU, RR



AZRAEL

A-Z Reconocedor Automático de Español

Consonantes finales

R S L N D Z

Ejemplos: *tos* *sol* *re-ír*
ver-dad *achu-chón* *a-rroz*

Consonante final absoluta: S

Ejemplo: *ins-truir*

AZRAEL

A-Z Reconocedor Automático de Español

Consonantes pseudofinales

CONSONANTE FINAL DE LA SÍLABA	CONTEXTO CONSONÁNTICO	NÚMERO DE CASOS EN EL CORPUS	EJEMPLOS
C	t	8627	ac-tuar
	c	2189	ac-ce-der
G	n	2487	ig-no-rar
M	p	21054	pom-pa
	b	16549	bom-ba
	n	564	a-lum-no
	t	2972	a-cep-tar
P	c	413	op-ción

AZRAEL

A-Z Reconocedor Automático de Español

Extranjerismos: cultismos y tecnicismos

CONSONANTE FINAL DE LA SÍLABA	CONTEXTO CONSONÁNTICO	NÚMERO DE CASOS RECOGIDOS EN EL CORPUS	EJEMPLOS
C	n	275	a-rác-ni-do
	s	19	fuc-sia
F	t	332	dif-te-ria
	g	18	af-ga-no
G	m	696	dia-frag-ma
	d	42	a-míg-da-las
T	m	194	at-mós-fe-ra
	l	115	at-le-ta
	n	72	et-nia
P	n	121	hip-no-sis

AZRAEL

A-Z Reconocedor Automático de Español

ORIGEN ETIMOLÓGICO DE LAS COMBINACIONES DE LETRAS

COMBINACIONES	ORIGEN	EJEMPLOS
TL	GRIEGO O NÁHUATL	atleta, tlacoyo
TM	GRIEGO	atmósfera
TN	GRIEGO	etnia
PN	GRIEGO	hipnosis
FT	GRIEGO O ÁRABE	difteria, muftí
GM	GRIEGO	diafragma
GD	GRIEGO	amígdalas
CN	GRIEGO	arácnido
IT	LATINISMO	déficit

AZRAEL

A-Z Reconocedor Automático de Español

Prefijos

CONSONANTE FINAL DE LA SÍLABA	CONTEXTO CONSONÁNTICO	NÚMERO DE CASOS RECOGIDOS EN EL CORPUS	EJEMPLOS
B	s	1024	ob-ser-var
	t	413	sub-te-rráneo
	v	367	ob-vio
	d	327	ab-di-car
	j	269	ob-je-to
	c	250	ob-ce-car
	y	175	ab-yec-to
	n	89	ob-nu-bi-lar
	m	66	sub-ma-ri-no
X	p	2569	ex-pan-dir
	t	2459	ex-tra-er
	c	1573	ex-cep-ción
	h	397	ex-ha-lar
	q	53	ex-qui-si-to

AZRAEL

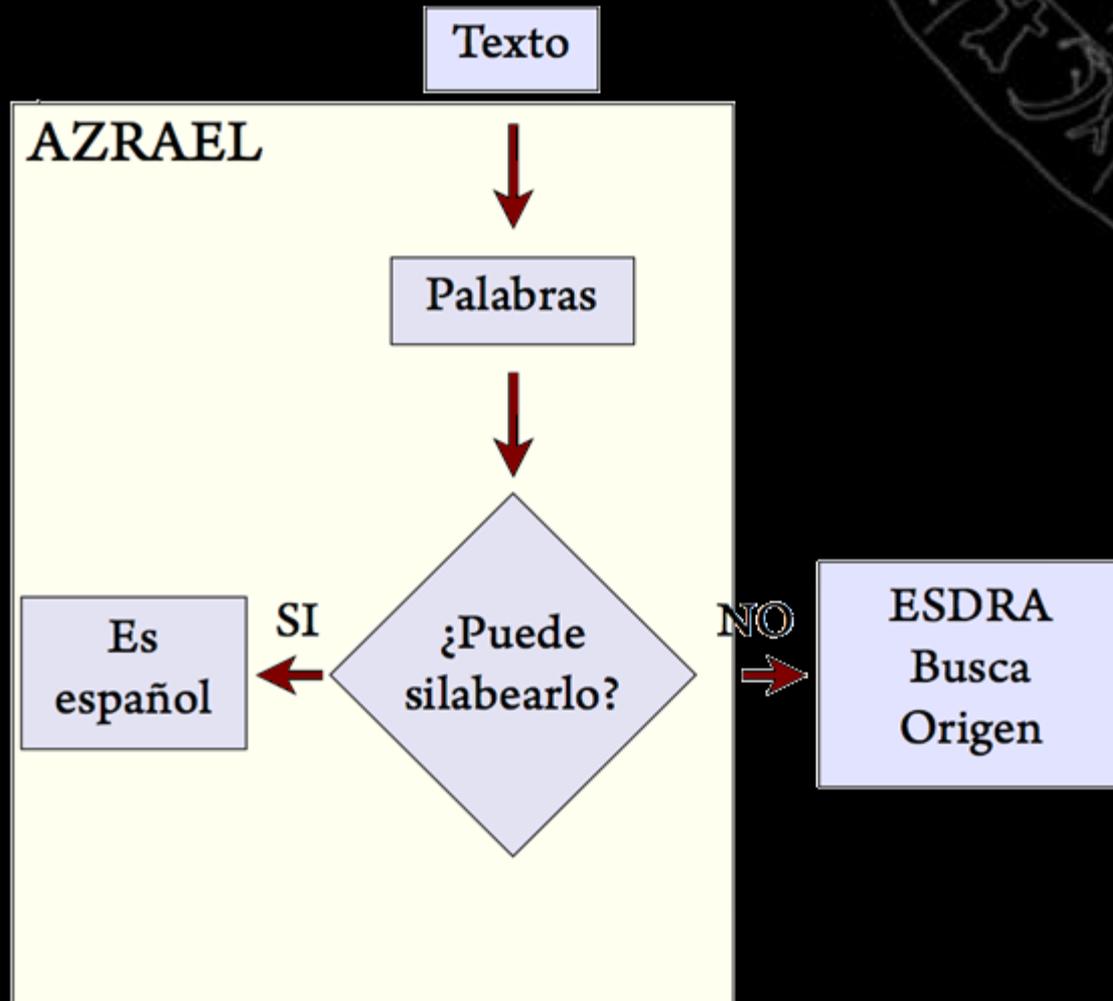
A-Z Reconocedor Automático de Español

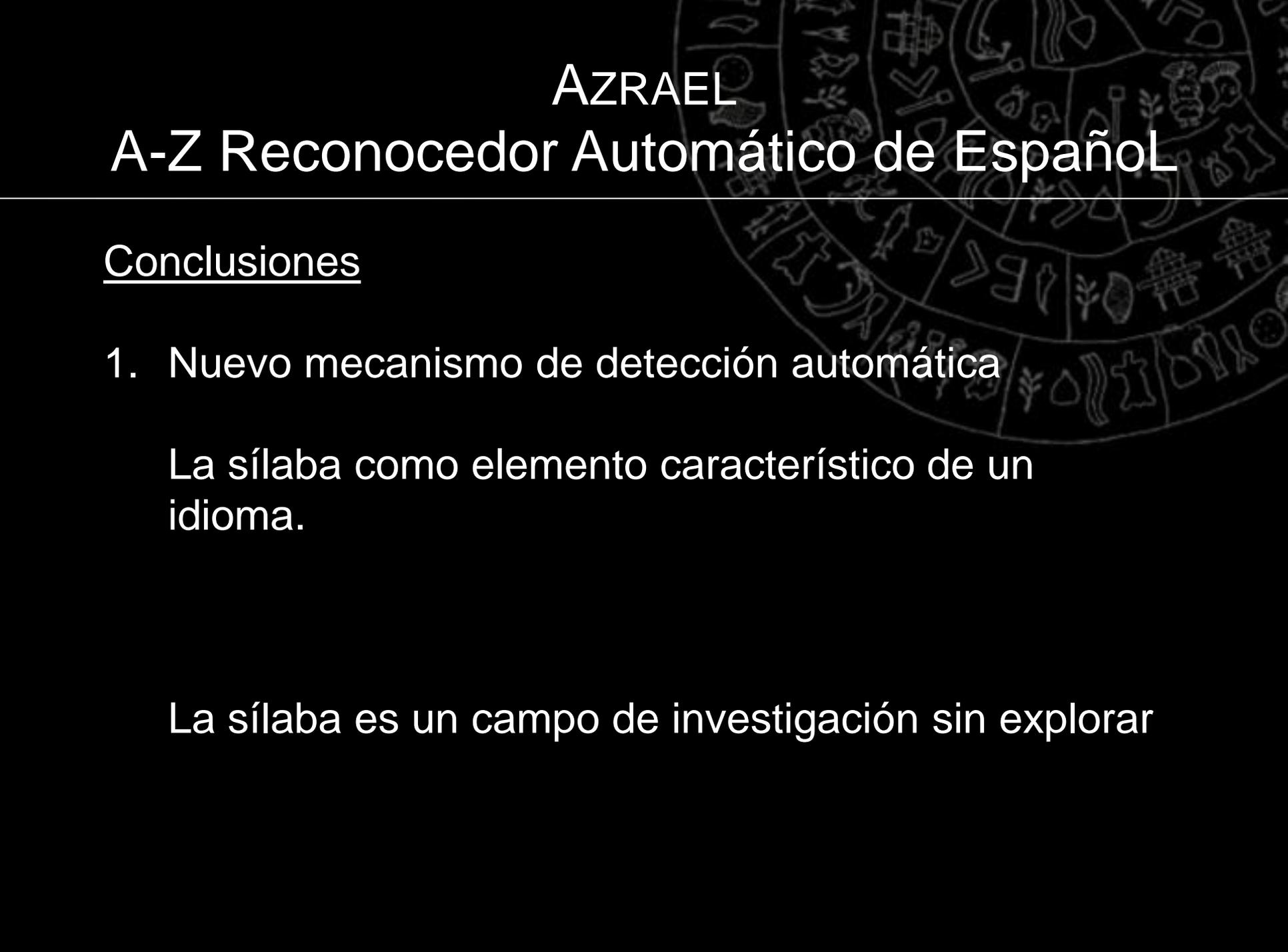
¿Cómo funciona AZRAEL?

```
español (X, Restos) :-  
    transcribir (X, W) ,  
    ! ,  
    lista (W, Y) ,  
    comprobar_palabras (Y, N) ,  
    recuperar_descartes (N, Restos) .  
  
transcribir ([W|Y], [W|Z]) :-  
    W < 65 ,  
    transcribir (Y, Z) .  
origen_etimológico ([Restos|MasRestos], [Origen|MasOrigen]) :-  
  
atom_chars (Restos, Deletreo) ,
```

AZRAEL

A-Z Reconocedor Automático de Español





AZRAEL

A-Z Reconocedor Automático de Español

Conclusiones

1. Nuevo mecanismo de detección automática

La sílaba como elemento característico de un idioma.

La sílaba es un campo de investigación sin explorar



AZRAEL

A-Z Reconocedor Automático de Español

Conclusiones

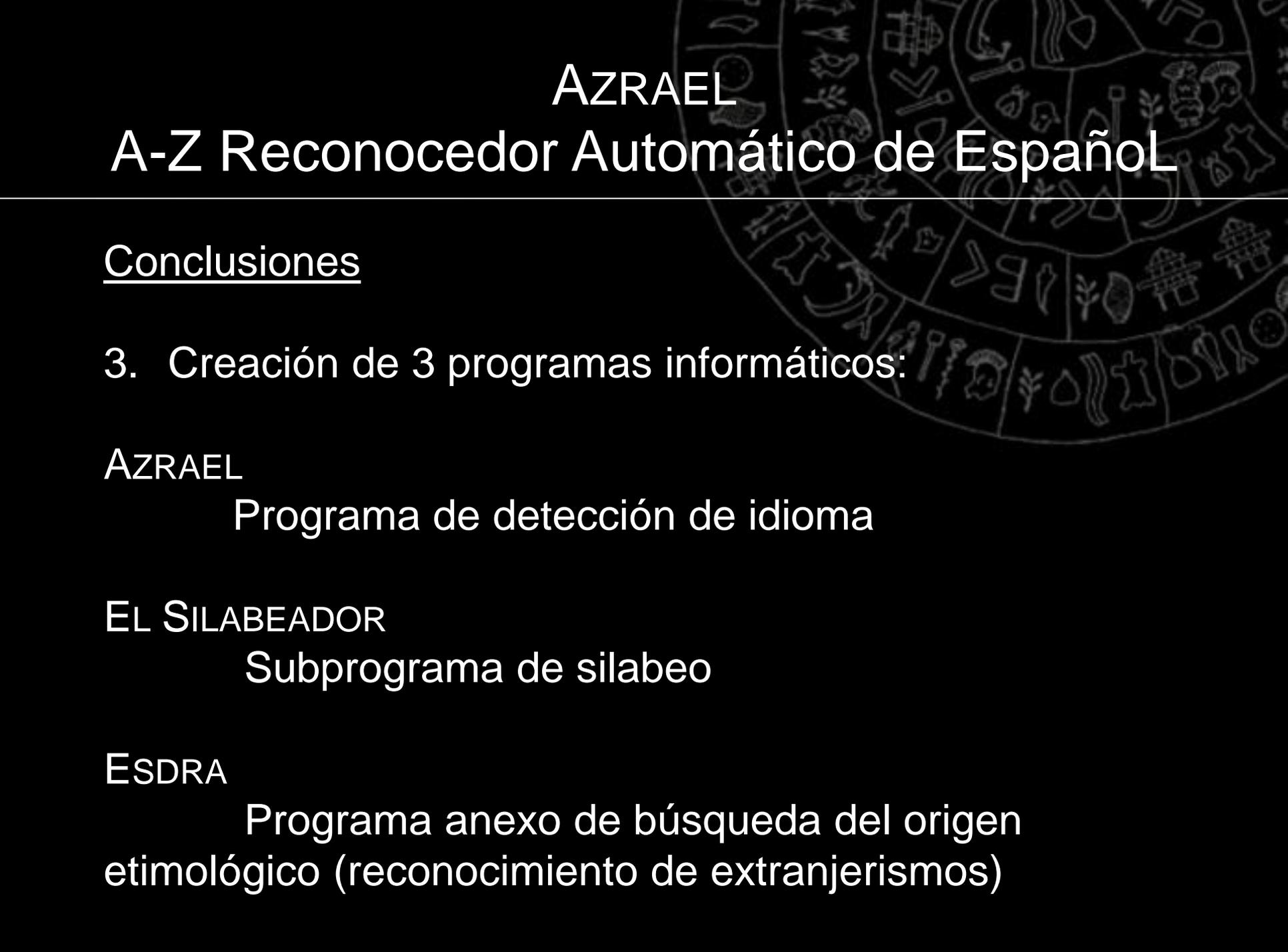
2. Estudio exhaustivo de la sílaba en español

Distinción entre distintos tipos de consonantes según sus posibilidades de combinación

El comportamiento particular de los prefijos

Los extranjerismos:

- * dónde: lengua de origen
- * cuándo: época de introducción al español
- * cómo: vía de entrada



AZRAEL

A-Z Reconocedor Automático de Español

Conclusiones

3. Creación de 3 programas informáticos:

AZRAEL

Programa de detección de idioma

EL SILABEADOR

Subprograma de silabeo

ESDRA

Programa anexo de búsqueda del origen etimológico (reconocimiento de extranjerismos)



AZRAEL

A-Z Reconocedor Automático de Español

Conclusiones

Los tres programas funcionan y pueden seguir desarrollándose

Aplicaciones en Procesamiento de Lenguaje Natural,
e Inteligencia Artificial

- * reconocimiento automático del lenguaje
- * traducción automática
- * procesadores de texto
- * predicción de escritura en sms
- * motores de búsqueda (Internet)

AZRAEL

A-Z Reconocedor Automático de Español

Apenas él le amalaba el noema, a ella se le agolpaba el clémiso y caían en hidromurias, en salvajes ambonios, en sustalos exasperantes. Cada vez que él procuraba relamar las incopelusas, se enredaba en un grimado quejumbroso y tenía que envulsionarse de cara al nóvalo, sintiendo cómo poco a poco las arnillas se espejunaban, se iban apeltronando, reduplicando, hasta quedar tendido como el trimalciato de ergomanina al que se le han dejado caer unas fílulas de cariaconcia. Y sin embargo era apenas el principio, porque en un momento dado ella se tordulaba los hurgalios, consintiendo en que él aproximara suavemente su orfelunios. Apenas se entreplumaban, algo como un ulucordio los encrestoriaba, los extrayuxtaba y paramovía, de pronto era el clinón, las esterfurosa convulcante de las mátricas, la jadehollante embocapluvia del orgumio, los esproemios del merpasmo en una sobrehumítica agopausa. ¡Evohé! ¡Evohé! Volposados en la cresta del murelio, se sentía balparamar, perlinos y márulos. Temblaba el troc, se vencían las marioplumas, y todo se resolviraba en un profundo pínice, en niolamas de argutendidas gasas, en carinias casi crueles que los ordopenaban hasta el límite de las gunfias.

Julio Cortázar, *Rayuela*