

Departamento de Computación, FCEyN, UBA

Procesamiento del Habla

Agustín Gravano

1er Cuatrimestre 2017

Sistema text-to-speech (TTS)

- **Front end**

- **Input:** Texto en algún formato: HTML, email, etc.
 - Procesamiento del texto de entrada: tokenización, segmentación, normalización.
 - Análisis lingüístico: clase de palabra, parsing.
 - Análisis fonético: grafemas → fonemas.
 - Análisis prosódico: asignación de prosodia (F0, dur, int).
- **Output:** Secuencia de fonemas + prosodia deseada.

- **Back end**

- **Input:** Secuencia de fonemas + prosodia deseada.
 - Síntesis del habla propiamente dicha: Articulatoria, de formantes, concatenativa, etc.
- **Output:** Archivo de audio.

Sistema text-to-speech (TTS)

- **Front end**

- **Input:** Texto en algún formato: HTML, email, etc.

*Clase
de hoy*

- **Procesamiento del texto de entrada:**
tokenización, segmentación, normalización.
- Análisis lingüístico: clase de palabra, parsing.
- Análisis fonético: grafemas → fonemas.
- Análisis prosódico: asignación de prosodia (F0, dur, int).

- **Output:** Secuencia de fonemas + prosodia deseada.

- **Back end**

- **Input:** Secuencia de fonemas + prosodia deseada.
 - Síntesis del habla propiamente dicha:
Articulatoria, de formantes, concatenativa, etc.
- **Output:** Archivo de audio.



Posibles textos de entrada

LAGOS Y LA OMC.- El Presidente Lagos saluda al director general de la OMC, Mike Moore, quien visita América Latina. Posteriormente partió viaje a EE.UU. para iniciar su visita a Silicon Valley.

“Si vamos a hacer un centro empresarial de alta tecnología, tenemos que buscar que exista esa interacción entre universidades, centros de investigación, empresas, etc. En Silicon Valley hay 6.000 empresas. Hace 10 años había menos de 2.500”.

LA CALERA (Alfredo Umaña).- El director del Hospital “Dr. Mario Sánchez Vergara” de esta ciudad, doctor Julio Trigo, afirmó ayer [...]

El llamado lo hizo en el 222º natalicio de O'Higgins. Pdte. Lagos, dijo que todas las instituciones y personas deben estar a la altura de los tiempos.

(Textos sacados del diario El Mercurio de Valparaíso del año 2000.)

Posibles textos de entrada

LAGOS Y LA OMC.- El Presidente Lagos saluda al director general de la **OMC**, **Mike Moore**, quien visita América Latina. Posteriormente partió viaje a **EE.UU.** para iniciar su visita a **Silicon Valley**.

“Si vamos a hacer un centro empresarial de alta tecnología, tenemos que buscar que exista esa interacción entre universidades, centros de investigación, empresas, **etc.** En **Silicon Valley** hay **6.000** empresas. Hace **10** años había menos de **2.500**”.

LA CALERA (Alfredo **Umaña**).- El director del Hospital “**Dr. Mario Sánchez Vergara**” de esta ciudad, doctor Julio Trigo, afirmó ayer [...]

El llamado lo hizo en el **222º** natalicio de **O’Higgins. Pdte.** Lagos, dijo que todas las instituciones y personas deben estar a la altura de los tiempos.

(Textos sacados del diario El Mercurio de Valparaíso del año 2000.)

Tokenización:

¿Qué es una palabra?

- ¿Cuáles son los delimitadores?

*deberán ser expendidos bajo receta médica o **receta-cheque**.*

*el jurado estará compuesto por [...] tres **afro-americanos** [...]*

*El llamado lo hizo en el 222º natalicio de **O'Higgins**.*

*“Toy **enamora'o**” (Cuba) de Joel David Rodríguez*

*¿Y **pa'qué**?...*

- Chino, japonés, coreano y otros: **problema no trivial**.

– Chino: 马路上生病了

[mǎ] [lù] [shàng] [shēng bìng] le (1)

[mǎ lù] [shàng] [shēng bìng] le (2)

(1) el caballo se enfermó en el camino

(2) alguien se enfermó en el camino

Segmentación:

¿Qué es una oración?

- El punto como fin de oración:

“Si vamos a hacer un centro empresarial de alta tecnología, tenemos que buscar que exista esa interacción entre universidades, centros de investigación, empresas, etc 😊 En Silicon Valley hay 6 😊 000 empresas 😊 Hace 10 años había menos de 2 😊 500” 😊

El llamado lo hizo en el 222º natalicio de O’Higgins 😊 Pdte 😊 Lagos, dijo que todas las instituciones y personas deben estar a la altura de los tiempos 😊

- Otros símbolos: ¿ ? ¡ ! ...

¿ Va a implicar mayores costos para la medicina? Existen borradores y prevenciones acerca del fallo, o sea [...]

Bajo el título “Comercio y Medioambiente en América del Sur: ¿Quo Vadis?”, se desarrolló en la costera localidad de Zapallar un encuentro que reunió a especialistas de [...]

Abreviaturas

LAGOS Y LA OMC.- El Presidente Lagos saluda al director general de la OMC, Mike Moore, quien visita América Latina. Posteriormente partió viaje a EE.UU. para iniciar su visita a Silicon Valley.

“Si vamos a hacer un centro empresarial de alta tecnología, tenemos que buscar que exista esa interacción entre universidades, centros de investigación, empresas, etc. En Silicon Valley hay 6.000 empresas. Hace 10 años había menos de 2.500”.

LA CALERA (Alfredo Umaña).- El director del Hospital “Dr. Mario Sánchez Vergara” de esta ciudad, doctor Julio Trigo, afirmó ayer [...]

El llamado lo hizo en el 222º natalicio de O’Higgins. Pdte. Lagos, dijo que todas las instituciones y personas deben estar a la altura de los tiempos.

(Textos sacados del diario El Mercurio de Valparaíso del año 2000.)

Este problema cambia continuamente. Ejemplos:

- Mensajes: xfa sbdo dnd kdms?
- Tecnologías: Tcl, SCSI, JPEG, DVD, HDMI, MySQL.
- Clasificados: 3amb,coc,blcn-corr,coch,sum,muy lum,estr,D\$125.
- Recetas: lb, lt, cda, cdta, tz, oz, kg, gr.
- Emoticones: :-) ;-P \o/ ~(_8°(l)

Expansión de abreviaturas

- **Abreviatura**: Secuencia de letras usadas para representar de forma breve una palabra o una frase.
 - *cap., pág., Cía., km, srta.*
- **Sigla**: Palabra formada por las letras iniciales de una expresión compuesta.
 - *ONU, IVA, DGI, IBM, AFIP, EEUU, RRPP.*
- **Acrónimo**: Palabra formada por las letras iniciales de una expresión compuesta, pero que suele ajustarse a las reglas fonológicas de la lengua.
 - *sida, radar, ovni, laser.*

Expansión de abreviaturas

- **Tablas** de abreviaturas conocidas.
 - Cuidado con el *número*: *1 kg* vs. *5 kg*
- Para abreviaturas ambiguas:
 - *5 m* → minutos o metros?
 - Guiarse por el dominio/tópico del contexto.
 - *Carlos Zbinden, 400 m con obstáculos.*
 - *5 m* → texto de cocina: “5 minutos”
 texto de arquitectura: “5 metros”
 texto de fútbol: ??
- Para abreviaturas desconocidas:
 - Idear reglas para **crear** abreviaturas: “*terraza*” → *trz?* *terr?*

Expresiones numéricas

LAGOS Y LA OMC.- El Presidente Lagos saluda al director general de la **OMC**, **Mike Moore**, quien visita América Latina. Posteriormente partió viaje a **EE.UU.** para iniciar su visita a **Silicon Valley**.

“Si vamos a hacer un centro empresarial de alta tecnología, tenemos que buscar que exista esa interacción entre universidades, centros de investigación, empresas, **etc.** En **Silicon Valley** hay **6.000** empresas. Hace **10** años había menos de **2.500**”.

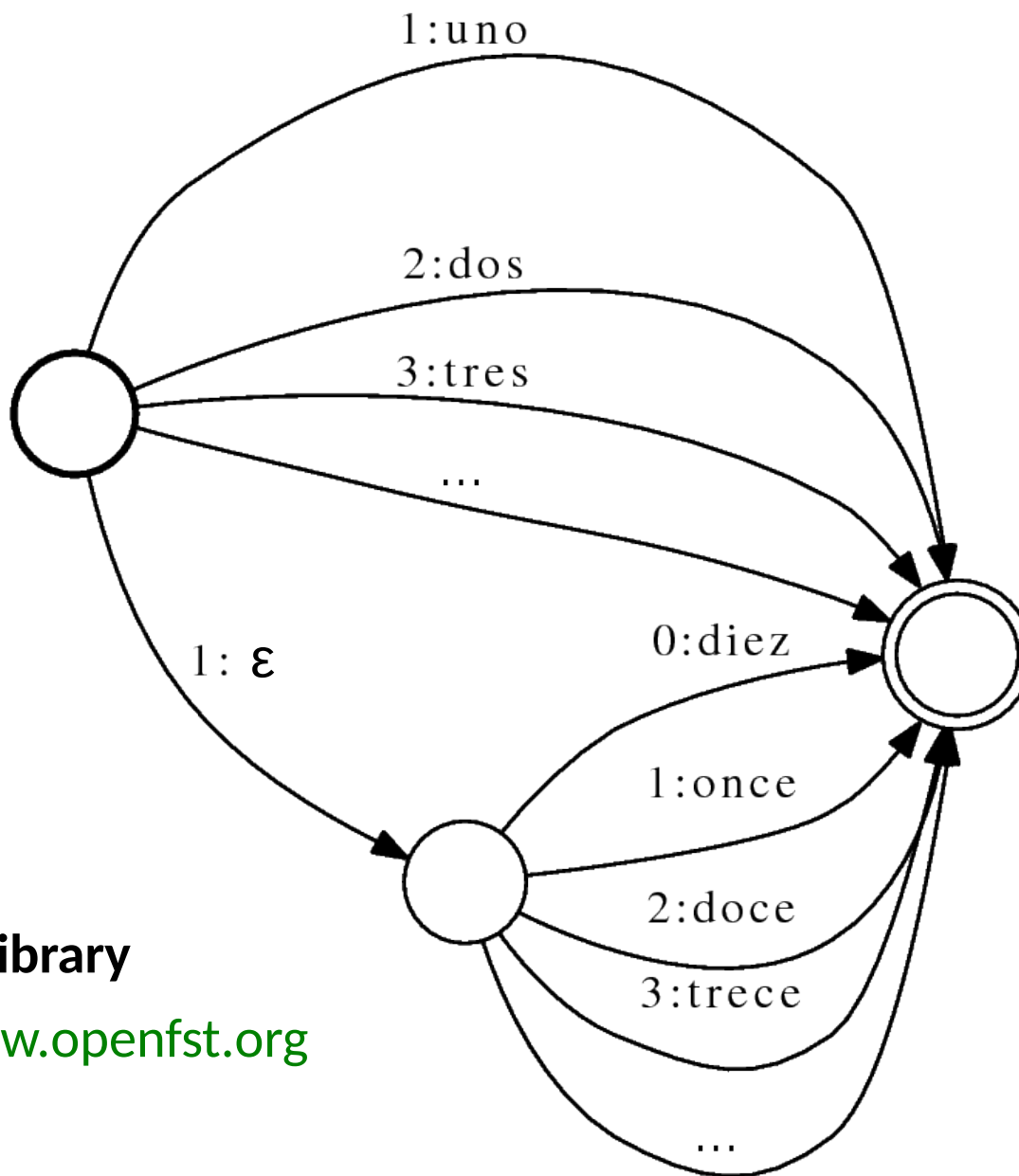
LA CALERA (Alfredo **Umaña**).- El director del Hospital “**Dr. Mario Sánchez Vergara**” de esta ciudad, doctor Julio Trigo, afirmó ayer [...]

El llamado lo hizo en el **222º** natalicio de **O’Higgins. Pdte.** Lagos, dijo que todas las instituciones y personas deben estar a la altura de los tiempos.

(Textos sacados del diario El Mercurio de Valparaíso del año 2000.)

Es importante tener en cuenta las **convenciones regionales/culturales** para: Fecha, hora, dinero, teléfono, DNI, pasaporte, tarjeta de crédito, cuenta de banco, dirección, código postal, ... Cuidado también con el género: “**500 millas**”

Expansión de números: WFST



OpenFst Library

<http://www.openfst.org>

Preprocesamiento del Texto

- Problemas:

- Tokenización (identificar palabras).
- Segmentación (identificar oraciones).
- Expansión de abreviaturas.
- Expansión de expresiones numéricas.

- Soluciones:

- Reglas hechas a mano, tablas, etc.
 - Desarrollo rápido y sencillo; buenos resultados.
 - Atado al dominio; mantenimiento lento.
- Enfoques de machine learning.
 - Mejores resultados; independiente del dominio; mantenimiento rápido.
 - Desarrollo más costoso y sofisticado (pero no tanto!).
 - Principal obstáculo: DATOS.

Sistema text-to-speech (TTS)

- **Front end**

- **Input:** Texto en algún formato: HTML, email, etc.

- Procesamiento del texto de entrada: tokenización, segmentación, normalización.
- Análisis lingüístico: clase de palabra, parsing.
- **Análisis fonético: grafemas → fonemas.**
- Análisis prosódico: asignación de prosodia (F0, dur, int).

- **Output:** Secuencia de fonemas + prosodia deseada.

- **Back end**

- **Input:** Secuencia de fonemas + prosodia deseada.
 - Síntesis del habla propiamente dicha:
Articulatoria, de formantes, concatenativa, etc.
- **Output:** Archivo de audio.

¿Qué codifica el lenguaje escrito?

- Para un front-end, necesitamos una función:
 - Texto arbitrario \rightarrow Secuencia de fonemas con prosodia
- ¿Codifica el lenguaje escrito un mensaje oral?
 - ¿Existe $f : \Sigma^+ \rightarrow \text{habla}$?
 - ¿Provee el lenguaje escrito toda la información necesaria para reconstruir el mensaje oral?
 - Si no, ¿cómo hacemos para leer un texto en voz alta?
- ¿Qué codifica el lenguaje escrito?
 - Alfabeto $\Sigma \rightarrow$ ¿Pronunciación? ¿Significado?

¿Qué codifica el lenguaje escrito?

- Alfabeto chino moderno: codifica **pronunciación** y algo de **significado**.

木	árbol	(mù)
木 _{x2} = 林	arboleda	(lín)
木 _{x3} = 森	bosque	(sēn)
人	persona	(rén)
人 + 木 = 休	descansar	(xiū)
火	fuego	(huo)
木 _{x2} + 火 = 焚	arder	(fén)

¿Qué codifica el lenguaje escrito?

- Alfabetos derivados del griego: codifican sólo **pronunciación**.
 - Algunos de manera bastante directa...
 - Español, finlandés, alemán.
 - Letras \leftrightarrow fonemas (aproximadamente)
 - Y otros no tanto...
 - Inglés!!! *blood* vs. *broom*, *though* vs. *cough*, etc....
- ¿Alcanza con codificar sólo la pronunciación?
 - ¿Existe $f : \Sigma^+ \rightarrow \text{habla}$?

Análisis fonético

- Reglas letra-a-sonido (o grafema-a-fonema):
 - Mapean ortografía → representación fonémica.
 - $c : _ [aou] \rightarrow /k/$
 - $c : _ [ei] \rightarrow /s/$ (en español argentino)
 - $qu : _ \rightarrow /k/$
 - Sirven para procesar cualquier entrada.
 - Solución útil en **español**.
 - Muchas excepciones en **inglés**. Ejemplo:
 - $a : _ \{C\} e \rightarrow /ey/$ es incorrecta para *water*
- Enfoques basados en diccionarios
 - Usar diccionarios con la ortografía y pronunciación de cada palabra.
 - CMU Pronouncing Dictionary (1993): 120k palabras.
 - UNISYN Dictionary (2002): 110k palabras.

Palabras extranjeras y palabras fuera de vocabulario

LAGOS Y LA OMC.- El Presidente Lagos saluda al director general de la OMC, Mike Moore, quien visita América Latina. Posteriormente partió viaje a EE.UU. para iniciar su visita a Silicon Valley.

“Si vamos a hacer un centro empresarial de alta tecnología, tenemos que buscar que exista esa interacción entre universidades, centros de investigación, empresas, etc. En Silicon Valley hay 6.000 empresas. Hace 10 años había menos de 2.500”.

LA CALERA (Alfredo Umaña).- El director del Hospital “Dr. Mario Sánchez Vergara” de esta ciudad, doctor Julio Trigo, afirmó ayer [...]

El llamado lo hizo en el 222º natalicio de O’Higgins. Pate. Lagos, dijo que todas las instituciones y personas deben estar a la altura de los tiempos.

(Textos sacados del diario El Mercurio de Valparaíso del año 2000.)

Análisis fonético: Palabras OOV

- Problema: **Palabras fuera de vocabulario (OOV)**.
 - Black et al. (1998). De ~40k palabras en un corpus de noticias, 1775 (4.4%) no figuraban en el diccionario:
 - 1360 → Nombres propios
 - 351 → Palabras desconocidas (ej: nuevas, extranjeras)
 - 64 → Typos
- Algunas soluciones:
 - Agrandar el diccionario... :-\
 - Usar reglas letra-a-sonido cuando no se encuentre una palabra en el diccionario.
 - Palabras extranjeras:
 - Inferir idioma de origen; usar reglas del idioma (si están disponibles)
 - Jackson, Tchebicheff, Ahmadinejad, Infiniti.
 - (sigue...)

Análisis fonético: Palabras OOV

- Reglas de reescritura (Fackrell and Skut 2004)
- Mejorar la cobertura del diccionario.
- Muchos nombres propios se escriben de varias maneras:
 - *Britney/Brittany; Smith/Smythe; More/Moore*
 - Homófonos
- Buscar un mapeo difuso entre palabras OOV y palabras que sí figuran en el léxico.
- Identificar alteraciones de escritura que sean neutrales respecto a la pronunciación, para así producir reglas de reescritura para palabras OOV.

Análisis fonético: Palabras OOV

- Derivar pronunciaciones de la web (Ghoshal et al. 2009)
- Extraer pares candidatos de **ortografía-pronunciación** de la web (con notación IPA o ad-hoc)
 - Ej: “*bruschetta (pronounced broo-SKET-uh)*”
- **Validar** los candidatos: ¿cuán probable es que estos pares representen una palabra y su correspondiente pronunciación?
 - *pronounced dead*
 - *pronounced swing rhythm*

Análisis fonético: Palabras OOV

- Análisis morfológico
- Morfología: Estudia la estructura de la formación de las palabras.
- Morfema: Unidad minimal portadora de significado.
 - *sol* → un morfema: *sol*
 - *gatas* → tres morfemas: *gat-*, *-a*, *-s*.
- Dos clases de morfemas: raíces y afijos.
 - Raíz: parte “principal” de la palabra (*sol*, *gat-*).
 - Afijos: agregan significados adicionales (*-a*: femenino, *-s*: plural).
 - Prefijos, sufijos, infijos/interfijos, circunfijos.
- Puede servir para inferir la pronunciación de palabras OOV:
 - Inglés: *demagnetizability* → *de-* + *magnet-* + *-ize-* + *-able* + *-ity* (inglés)
 - Español: *subregión* → *sub-* *región* (y no “*su bre gión*”)

Análisis morfológico

- **Prefijos**: preceden a la raíz.
 - *inesperado*: in-, *superhéroe*: super-
- **Sufijos**: van después de la raíz.
 - *esperable*: -able, *simplemente*: -mente
- **Infijos e interfijos**: Se insertan en medio de la palabra.
 - infijos: tienen significado.
 - *Man-fucking-hattan*, *abso-bloody-lutely*.
 - Tagalog (Filipinas):

<i>hingi</i>	(tomar prestado)
+ <i>um</i>	(agente de la acción)
= <i>humingi</i>	(quien tomó prestado)
 - interfijos: no tienen significado.
 - *en-s-anchar*, *cafe-c-ito/cafe-l-ito*, *te-t-era*.
- **Circunfijos**: Rodean a la raíz.
 - Alemán: los participios se crean con *ge-* _ -t
 - Ejemplo: *sagen* (decir) → *gesagt* (dicho).

Análisis morfológico

- Cómo combinar morfemas para crear palabras:
 - **Flexión**: cambios puramente gramaticales
 - Nominal: *gat- -a -s*
 - Verbal: *caminábamos* → *caminar*, pretérito imperfecto, primera persona, plural.
 - **Derivación**: cambios semánticos referenciales
 - *cuchillo* → *cuchillada*
 - **Composición**: juntar dos o más raíces
 - *sordomudo, medianoche, quienquiera, boquiabierto, malpensado, salvoconducto, baloncesto, pasamontañas, bocacalle, telaraña, correveidile, sabelotodo, nomeolvides, bajorrelieve, pelirrojo, ...*

Palabras homógrafas

- Se escriben igual, pero se pronuncian distinto:
 - Inglés: *lead*, *live*, *desert*, *bass*.
 - Español: “como acá”.
- Desambiguarlas es otro desafío para TTS.
- Técnicas de desambiguación del sentido ([word sense disambiguation](#)).
 - También sirven para saber cómo expandir números: teléfonos, códigos postales, etc.
- Muchas veces (aunque no siempre) alcanza con saber la **clase de palabra** para desambiguar, aunque no siempre:

lead:VB/*lead*:NN

live:VB/*live*:ADJ

desert:VB/*desert*:NN

bass:NN/*bass*:NN :-(

como:VB/*como*:CONJ

Ortografías especiales

“Yer’ great puddin’ of a son don’ need fattenin’ anymore, Dursley, don’ worry.” (Hagrid)



¿Normalización? ¿Pronunciación?

Sistema text-to-speech (TTS)

- **Front end**

- **Input:** Texto en algún formato: HTML, email, etc.

- Procesamiento del texto de entrada: tokenización, segmentación, normalización.
- **Análisis lingüístico: clase de palabra, parsing.**
- Análisis fonético: grafemas → fonemas.
- Análisis prosódico: asignación de prosodia (F0, dur, int).

- **Output:** Secuencia de fonemas + prosodia deseada.

- **Back end**

- **Input:** Secuencia de fonemas + prosodia deseada.

- Síntesis del habla propiamente dicha:
Articulatoria, de formantes, concatenativa, etc.

- **Output:** Archivo de audio.

Clases de palabras

(Part-of-speech, POS)

- Hay palabras que exhiben **comportamiento parecido**:
 - Aparecen en contextos similares.
 - Desempeñan funciones similares en la oración.
 - Sufren transformaciones similares.
- Clases de palabras:
 - sustantivo *silla, parlante, bajo*
 - verbo *estar, estoy, comprendería, bajo*
 - adjetivo *púrpura, altas, ridículo, parlante, bajo*
 - adverbio *desafortunadamente, despacio, bajo*
 - preposición *de, por, a, bajo*
 - pronombre *yo, mi, mío*
 - determinante *el, un, ese, estas*
 - conjunción *y, o, ni, sino*
 - ...

POS tagset

- **Conjunto de rótulos/etiquetas.**
- Podríamos elegir un conjunto reducido:
 - N, V, Adj, Adv, Otros.
- **Brown Corpus:** 1M palabras, 87 tags
 - Más informativo, pero más difícil de etiquetar.
- **Penn Treebank:** corpus del Wall Street Journal, 1M palabras, 45-46 tags.
 - El más usado en la actualidad para el inglés.
- Ejemplo para el español: **EAGLE tagset**
 - <http://nlp.lsi.upc.edu/freeling-old/doc/tagsets/tagset-es.html>

Penn Treebank Tagset

Tag	Description	Example	Tag	Description	Example
CC	Coordin. Conjunction	<i>and, but, or</i>	SYM	Symbol	<i>+, %, &</i>
CD	Cardinal number	<i>one, two, three</i>	TO	“to”	<i>to</i>
DT	Determiner	<i>a, the</i>	UH	Interjection	<i>ah, oops</i>
EX	Existential ‘there’	<i>there</i>	VB	Verb, base form	<i>eat</i>
FW	Foreign word	<i>mea culpa</i>	VBD	Verb, past tense	<i>ate</i>
IN	Preposition/sub-conj	<i>of, in, by</i>	VBG	Verb, gerund	<i>eating</i>
JJ	Adjective	<i>yellow</i>	VCN	Verb, past participle	<i>eaten</i>
JJR	Adj., comparative	<i>bigger</i>	VBP	Verb, non-3sg pres	<i>eat</i>
JJS	Adj., superlative	<i>wildest</i>	VBZ	Verb, 3sg pres	<i>eats</i>
LS	List item marker	<i>1, 2, One</i>	WDT	Wh-determiner	<i>which, that</i>
MD	Modal	<i>can, should</i>	WP	Wh-pronoun	<i>what, who</i>
NN	Noun, sing. or mass	<i>llama</i>	WP\$	Possessive wh-	<i>whose</i>
NNS	Noun, plural	<i>llamas</i>	WRB	Wh-adverb	<i>how, where</i>
NNP	Proper noun, singular	<i>IBM</i>	\$	Dollar sign	<i>\$</i>
NNPS	Proper noun, plural	<i>Carolinas</i>	#	Pound sign	<i>#</i>
PDT	Predeterminer	<i>all, both</i>	“	Left quote	<i>(‘ or “)</i>
POS	Possessive ending	<i>’s</i>	”	Right quote	<i>(’ or ”)</i>
PRP	Personal pronoun	<i>I, you, he</i>	(Left parenthesis	<i>([, (, { , <)</i>
PRP\$	Possessive pronoun	<i>your, one’s</i>)	Right parenthesis	<i>([, (, { , <)</i>
RB	Adverb	<i>quickly, never</i>	,	Comma	<i>,</i>
RBR	Adverb, comparative	<i>faster</i>	.	Sentence-final punc	<i>(. ! ?)</i>
RBS	Adverb, superlative	<i>fastest</i>	:	Mid-sentence punc	<i>(: ; ... – -)</i>
RP	Particle	<i>up, off</i>			

Clases de palabras

(Part-of-speech, POS)

- POS Tagging
 - Proceso de asignar una etiqueta de clase de palabra (POS) a cada palabra de un texto.

El informe dice que no hay evidencia

Det Sust Verbo Conj Adv Verbo Sust

- ¡No es un mapeo directo!
 - *La compra* se concretó. *la* = artículo *compra* = sust
 - *La compra* con crédito. *la* = pronombre *compra* = verbo
 - Ejemplo en inglés: *Time flies like an arrow.*
Time [VB,NN] *flies* [VB,NN] *like* [VB,PR] *an* DT *arrow* NN

Rule-based POS tagging

- Variantes del siguiente algoritmo:
 - 1) Asignar a cada palabra todos sus tags posibles.
 - 2) Usar un conjunto de reglas predefinidas para eliminar tags selectivamente.
 - 3) Iterar hasta que cada palabra tenga exactamente un tag.

Rule-based POS tagging

			NN		
			RB		
	VBN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	promised	to	back	the	bill

Rule-based POS tagging

Regla: *Eliminar VBN si VBD es una opción cuando VBN|VBD sucede a “<start> PRP”.*

VBN = verb, past participle

VBD = verb, past tense

			NN		
			RB		
	VBN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	promised	to	back	the	bill

Rule-based POS tagging

Regla: *Eliminar VBN si VBD es una opción cuando VBN|VBD sucede a “<start> PRP”.*

VBN = verb, past participle

VBD = verb, past tense

			NN		
			RB		
	VBN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	promised	to	back	the	bill

Rule-based POS tagging

Regla: *Después de DT no puede venir VB (si hay otra opción disponible).*

DT = determiner

VB = verb, base form

			NN		
			RB		
	VBN		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	promised	to	back	the	bill

Rule-based POS tagging

Regla: *Después de DT no puede venir VB (si hay otra opción disponible).*

DT = determiner

VB = verb, base form

			NN		
			RB		
	VB N		JJ		VB
PRP	VBD	TO	VB	DT	NN
She	promised	to	back	the	bill

Algoritmos de POS tagging

- Rule-based tagging (ej: [EnCG ENGTWOL tagger](#))
- Transformation-based tagging (ej: [Brill tagger](#))
- Machine learning:
 - Support Vector Machines (SVM), Hidden Markov Models (HMM), Deep Neural Networks (DNN).
- Performance cercana a la humana: **95-99%**
- ¿Problema aparentemente “resuelto”?
 - ¡No!
 - Muy atado al dominio de los textos de entrenamiento.
 - Habla espontánea: problema abierto.
 - *el este el libro tiene un tiene una foto en la t- no en la contratapa*

Análisis Sintáctico

Gramática

S → NP VP
NP → DT NP
NP → NN
VP → VB PP
VP → RN VP
VP → VB NP
...

→ Reglas para la construcción de oraciones.


DT	NN	VB	CS	RN	VB	NN
El	informe	dice	que	no	hay	evidencia

Análisis Sintáctico

Gramática

S → NP VP
NP → DT NP
NP → NN
VP → VB PP
VP → RN VP
VP → VB NP
...

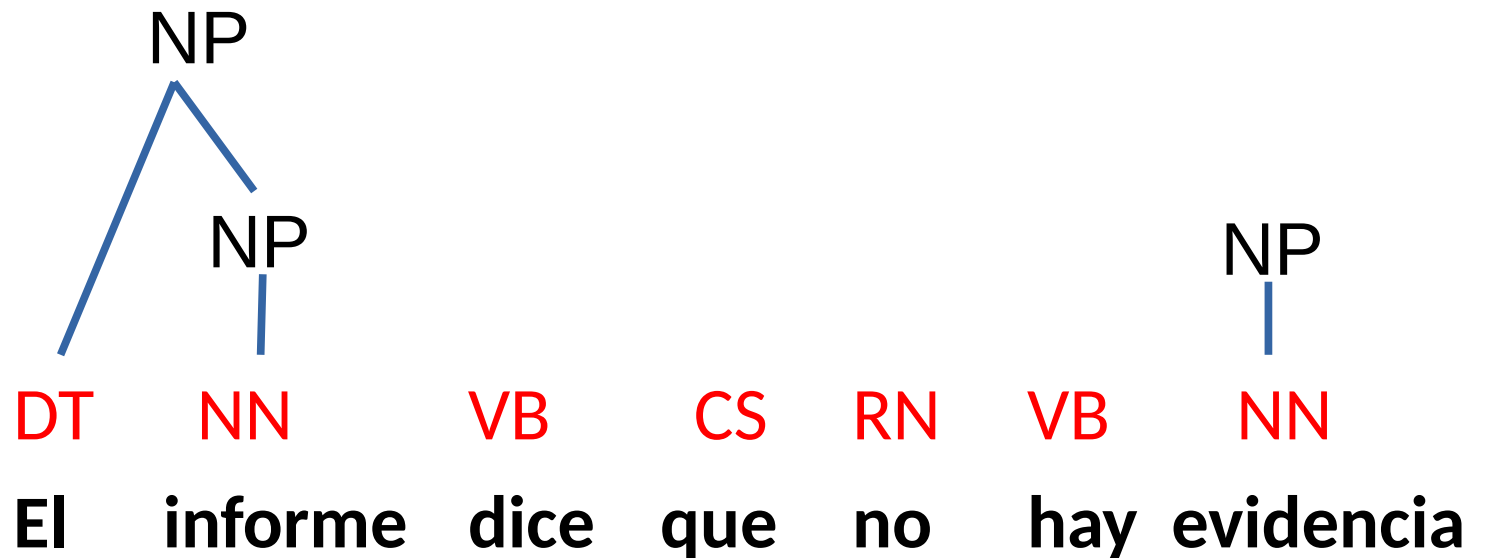
NP
|
DT NN VB CS RN VB NN
El informe dice que no hay evidencia



Análisis Sintáctico

Gramática

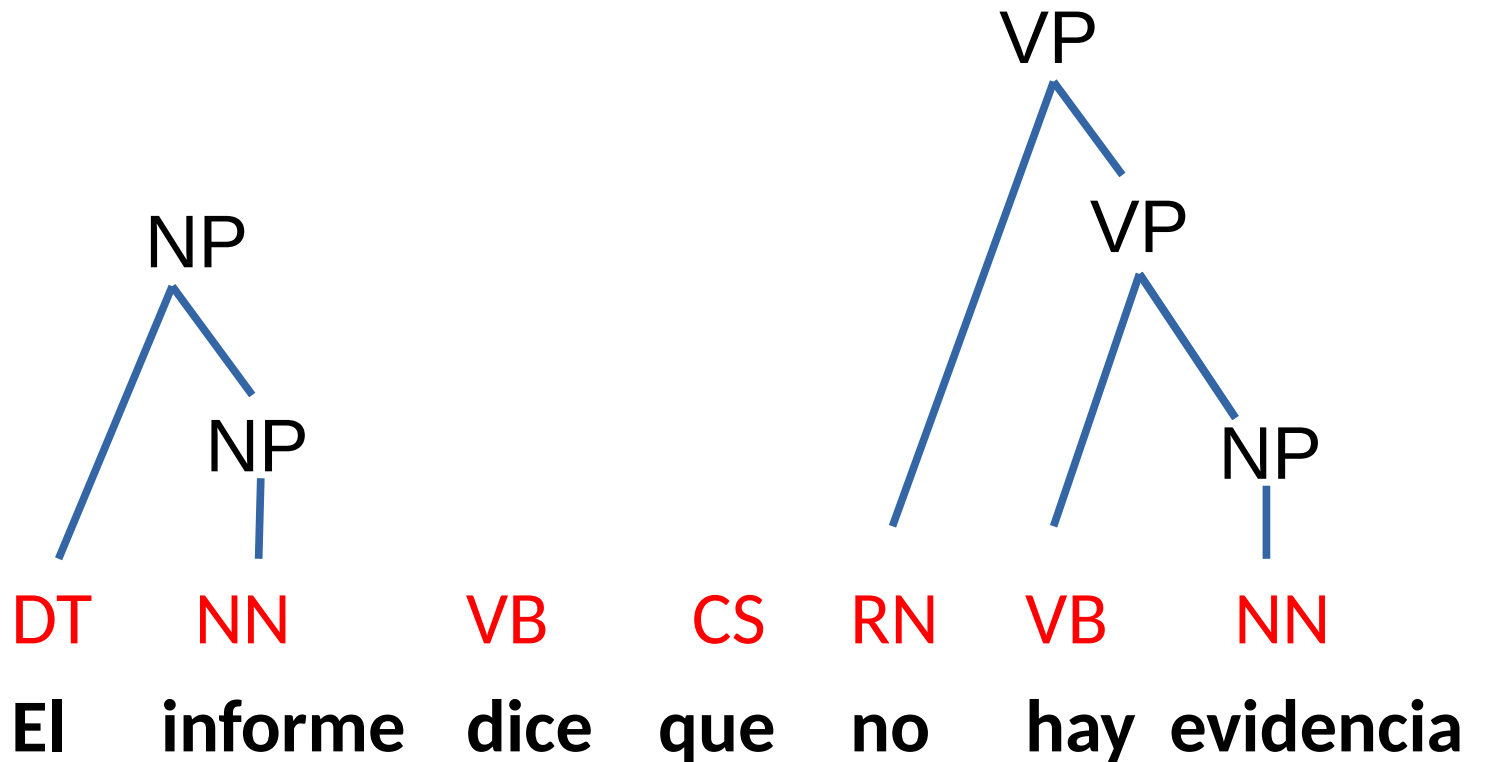
S → NP VP
NP → DT NP
NP → NN
VP → VB PP
VP → RN VP
VP → VB NP
...



Análisis Sintáctico

Gramática

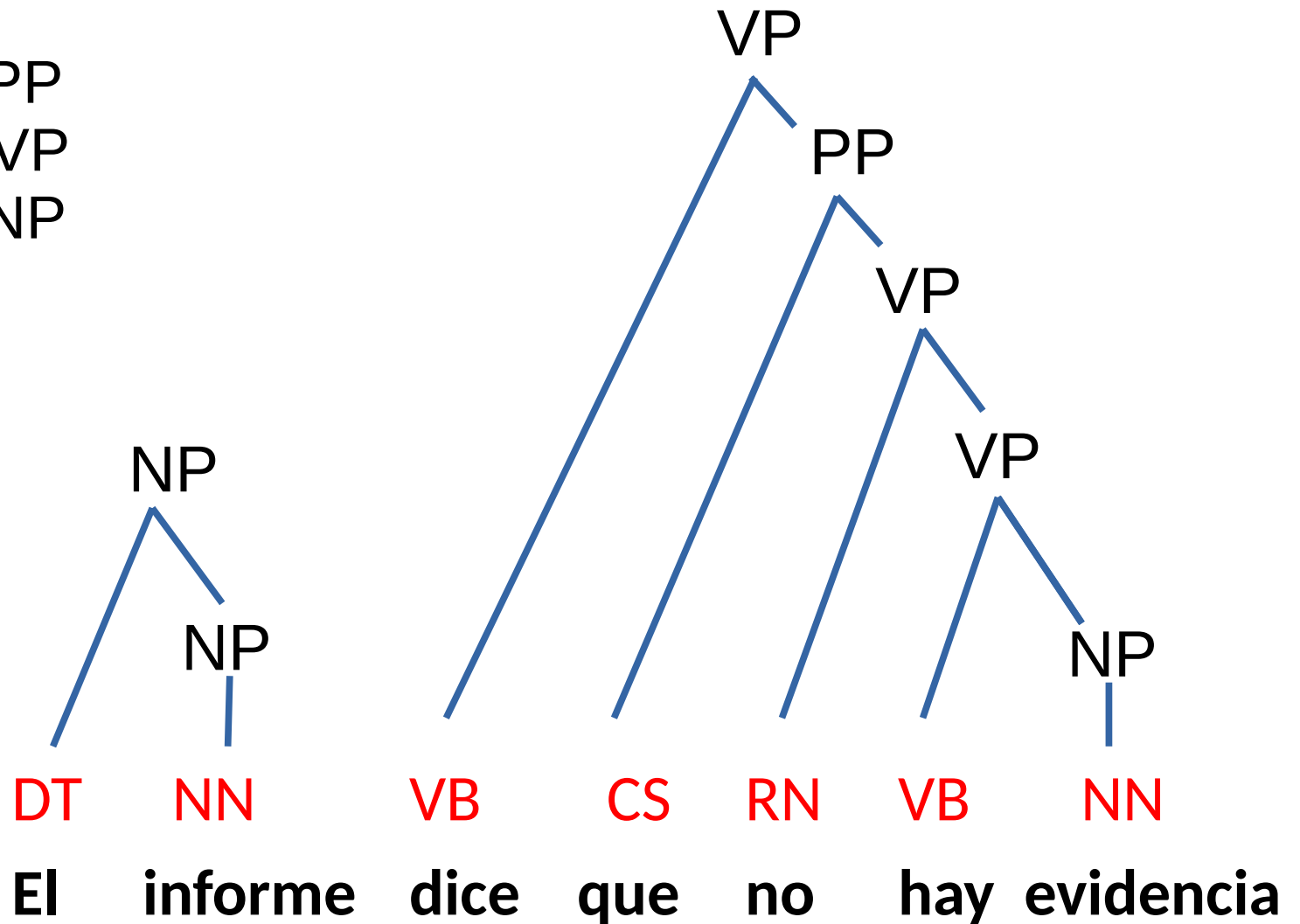
S → NP VP
NP → DT NP
NP → NN
VP → VB PP
VP → RN VP
VP → VB NP
...



Análisis Sintáctico

Gramática

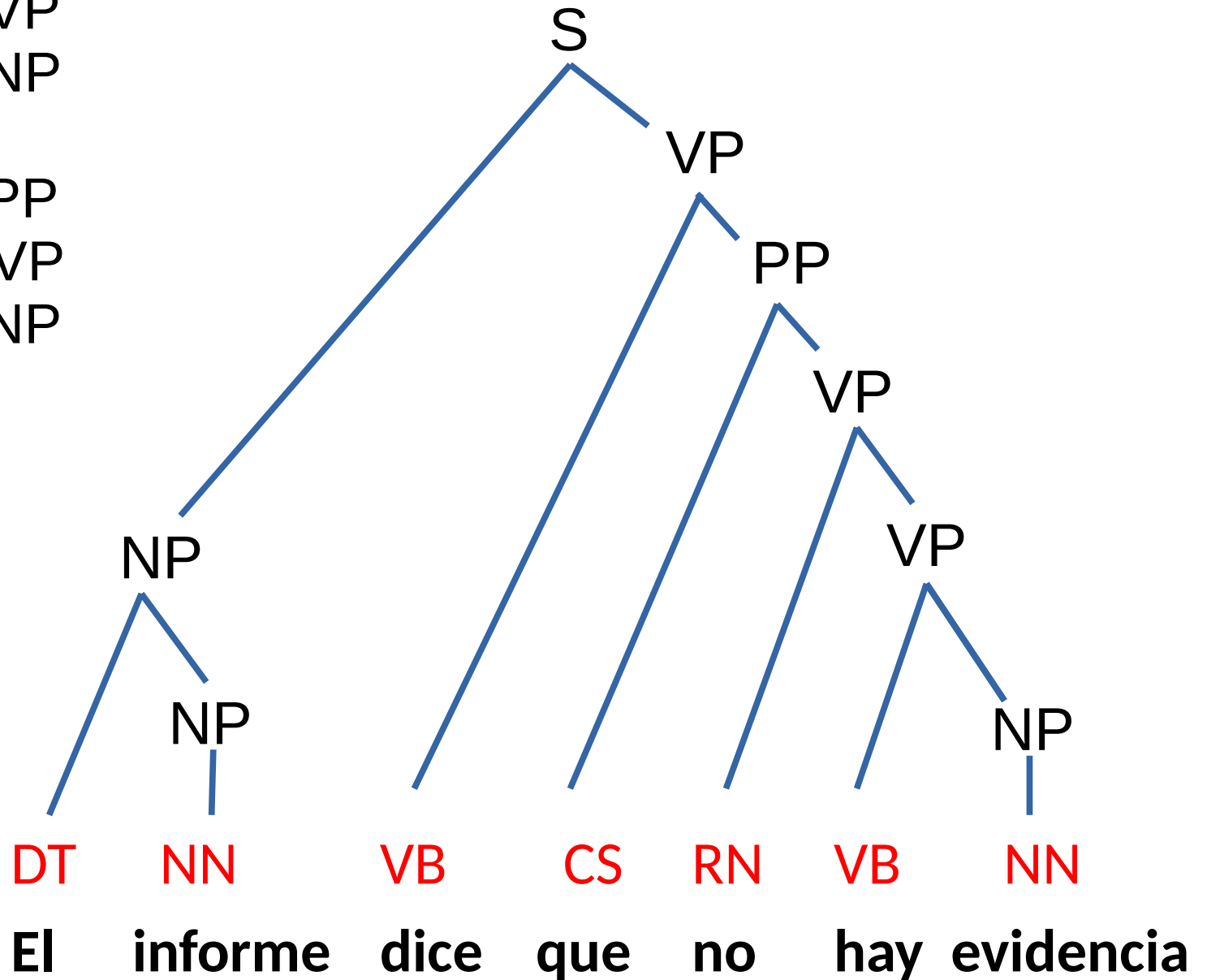
S → NP VP
NP → DT NP
NP → NN
VP → VB PP
VP → RN VP
VP → VB NP
...



Análisis Sintáctico

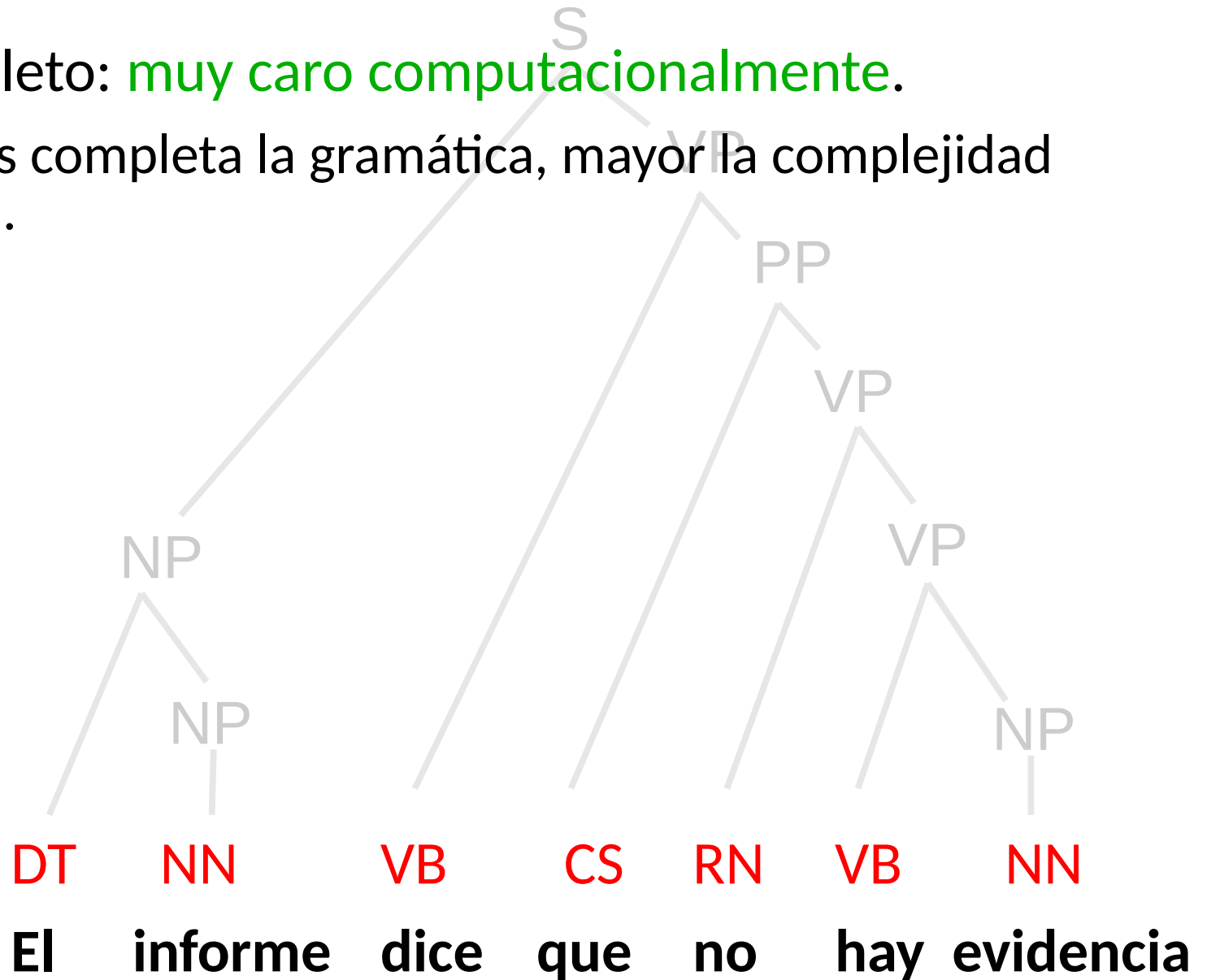
Gramática

S → NP VP
NP → DT NP
NP → NN
VP → VB PP
VP → RN VP
VP → VB NP
...



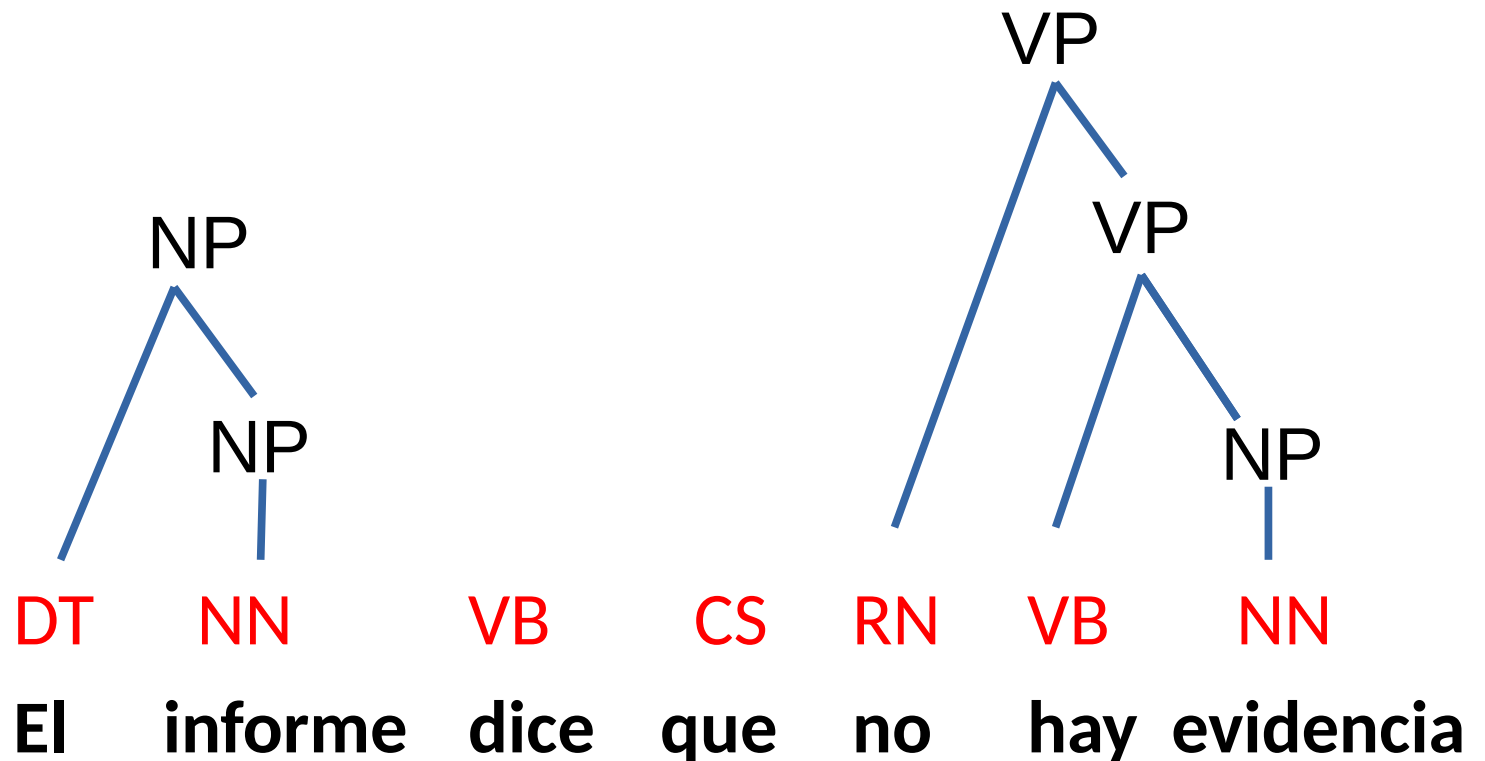
Análisis Sintáctico

- Parsing completo: **muy caro computacionalmente.**
 - Cuanto más completa la gramática, mayor la complejidad algorítmica.



Análisis Sintáctico

- Parsing completo: **muy caro computacionalmente**.
 - Cuanto más completa la gramática, mayor la complejidad algorítmica.
 - Alternativas: shallow parsing; chunking.



Análisis Sintáctico

- Parsing estadístico
 - Decidir qué regla usar en base a estadísticas tomadas de cuerpos de datos.
- Parsing de dependencias
 - Grafo: nodos=palabras, aristas=dependencias.
 - Construido en base a estadísticas tomadas de cuerpos de datos.
 - Ventaja: no hay que definir una gramática.

Análisis Lingüístico: Resumen

- Clase de palabras: POS tagging.
- Análisis sintáctico:
 - Full parsing, shallow parsing, chunking, parsing estadístico, parsing de dependencias.
- Herramientas de PLN en español:
 - FreeLing: <http://nlp.lsi.upc.edu/freeling>
 - Natural Language Toolkit (NLTK) <http://www.nltk.org/>

Sistema text-to-speech (TTS)

- **Front end**

- **Input:** Texto en algún formato: HTML, email, etc.

- Procesamiento del texto de entrada: tokenización, segmentación, normalización.
- Análisis lingüístico: clase de palabra, parsing.
- Análisis fonético: grafemas → fonemas.
- **Análisis prosódico: asignación de prosodia (F0, dur, int).**

- **Output:** Secuencia de fonemas + prosodia deseada.

- **Back end**

- **Input:** Secuencia de fonemas + prosodia deseada.

- Síntesis del habla propiamente dicha:
Articulatoria, de formantes, concatenativa, etc.

- **Output:** Archivo de audio.

Asignación de prosodia (o *entonación*)

- Texto de entrada:

En enero quería ir a Brasil, pero no encontré nada barato.

- Para que el habla sintetizada suene **natural**, debemos decidir:
 - dónde poner **pausas**, y
 - a qué palabras dar mayor **prominencia**.

- Posible asignación (una entre varias posibles):

En enero | quería ir a Brasil, | pero no encontré | nada barato. |

- Soluciones: Reglas manuales y/o machine learning. Tema abierto de inv.
- ¡Cuidado! Una asignación poco feliz puede cambiar la connotación:
 - Ejemplos: “**no cantes victoria**”, “**vi un perro con un telescopio**”.

Sistema text-to-speech (TTS)

- **Front end**

- **Input:** Texto en algún formato: HTML, email, etc.

- Procesamiento del texto de entrada: tokenización, segmentación, normalización.
- Análisis lingüístico: clase de palabra, parsing.
- Análisis fonético: grafemas → fonemas.
- Análisis prosódico: asignación de prosodia (F0, dur, int).

- **Output:** Secuencia de fonemas + prosodia deseada.

- **Back end**

- **Input:** Secuencia de fonemas + prosodia deseada.

- Síntesis del habla propiamente dicha:
Articulatoria, de formantes, concatenativa, etc.

- **Output:** Archivo de audio.