

El análisis sintáctico superficial o la simplificación de una tarea compleja

Ferran Pla

Departament de Sistemes Informàtics i Computació

Universitat Politècnica de València

fpla@dsic.upv.es

<http://www.dsic.upv.es/~fpla>

Índice

- Introducción al Análisis Sintáctico
- Definición de Análisis Superficial y Aplicaciones
- Aproximaciones
- Evaluación
- Conclusiones

Análisis Sintáctico

El **Análisis Sintáctico** de textos consiste en recuperar la estructura sintáctica (árbol sintáctico) asociada a una oración.

Aproximación ‘clásica’

- **Representación** de la información sintáctica.
 - redes de transición, gramáticas, etc.
- **Algoritmos** de análisis.
 - estrategia, dirección del análisis, orden de aplicación de las reglas, etc.

Análisis Sintáctico: ejemplo

$G = \{N, T, S, P\}$

$N = \{S, NP, SP\}$

$T = \{nprop, n, v, det, adj\}$
(categorías morfológicas)

(1) $S \rightarrow NP VP$

(2) $NP \rightarrow nprop$

(3) $NP \rightarrow det n$

(4) $NP \rightarrow det n adj$

(5) $VP \rightarrow v$

(6) $VP \rightarrow VP NP$

Gramática CFG

La	det	
niña	n	POS
baja	v	TAGGING
la	det	
persiana	n	

(3) $NP v det n$

(5) $NP VP det n$

(3) $NP VP NP$

(6) $NP VP$

(1) S

Análisis ascendente

Más información en: [Allen, 1995], [Moreno et al., 1999], ...

Análisis Sintáctico: problemas

- **Gramáticas de amplia cobertura**
 - problemas en la definición
 - sobregeneración,
 - subgeneración,
 - robustez, etc.
- **Ambigüedad estructural**
 - distintos análisis para una oración.
- **Eficiencia**
 - alta complejidad algorítmica.

Alternativas: aproximaciones inductivas

Aprender las gramáticas a partir de una colección de ejemplos analizados

- Necesidad de Treebanks. Ej: *Penn Treebank* para el inglés. [Marcus et al., 1993]

Inferencia Gramatical

- [González and Thomason, 1978], [Fu and Booth, 1975]

Aprendizaje estadístico de analizadores completos

- [Charniak, 1997]
- [Collins, 1996]
- [Ratnaparkhi, 1997]

... un tema de investigación abierto

Simplificación: Análisis Parcial

‘Partial Parsing’

‘Chunking’

‘Shallow Parsing’

[Abney, 1991]

Obtener información sintáctica de la oración en **unidades sintácticas de interés** de forma **eficiente** y fiable **sacrificando la completitud** y profundidad del análisis global.

Ejemplos de Análisis Parcial

- detección de unidades sintácticas básicas no solapadas
“Chunking”
- detección de unidades sintácticas más complejas
- ligamiento preposicional
- detección de cláusulas
- detección de roles
- reconocimiento de entidades ...

Análisis Superficial o Parcial

Características

- Análisis robusto para textos no restringidos.
- Algoritmos más eficientes computacionalmente.
- Combinación de diferentes técnicas.
- Posibilidad de conseguir distintos grados de profundidad del análisis.

Aplicaciones

Análisis Sintáctico Total

- Utilización de los chunks como entrada al analizador

Traducción Automática

- Traducir unidades sintácticas en vez de palabras

Sistemas Diálogo

- Análisis del lenguaje coloquial

Extracción de Información

- Identificar unidades sintácticas correspondientes a la información buscada

Recuperación de Información

- Índices sintácticos, semánticos ...

Sistemas de Pregunta/Respuesta

Aplicaciones de RAH

Análisis Superficial: aproximaciones

Basadas en conocimiento (lingüísticas)

- Gramática parcial + Algoritmo análisis [Ferrández et al., 1998]
- Análisis incremental utilizando máquinas de estados finitos [Abney, 1991, Abney, 1996, Abney, 1997]
[Aït-Mokhtar and Chanod, 1997] [Ejerhed, 1988] ...

Basadas en corpus (inductivas o de aprendizaje)

- <http://lcg-www.uia.ac.be/~erikt/research/>
<http://lcg-www.uia.ac.be/conll2000/chunking/>
<http://lcg-www.uia.ac.be/conll2001/clauses/>
<http://lcg-www.uia.ac.be/conll2002/ner>
- **Journal of Machine Learning Research**
<http://www.ai.mit.edu/projects/jmlr/papers/>

Chunking: máquinas de estados finitos en cascada

[Abney, 1991, Abney, 1996, Abney, 1997]

- Secuencia de **niveles**. Cada nivel reconoce un conjunto de estructuras sintácticas no solapados (**chunks**) que se describen mediante **expresiones regulares o patrones**.

Ejemplo: [Molina et al., 1999]

Nivel 1 // núcleos nominales y verbales

$\text{NSN} \rightarrow (\text{NC} \mid \text{NP})^+$

$\text{NSV} \rightarrow (\text{VM} \mid \text{VA VMP})$

Nivel 2 // sintagmas nominales

$\text{SN} \rightarrow \text{TD? AQ}^* \text{ NSN AQ}^*$

Nivel 3 // sintagmas preposicionales

$\text{SPR} \rightarrow \text{SP SN}$

Nivel 4 // sintagmas verbales

$\text{SV} \rightarrow \text{NSV} (\text{SN} \mid \text{SPR})^*$

- Las expresiones regulares de un nivel se traducen a un **autómata de estados finitos determinista**.
- Gran eficiencia computacional debido a la arquitectura.

Inicios de las aproximaciones inductivas

NP-Chunking

- Se siguen aproximaciones similares a las de “POS tagging” [Church, 1988]
- [Ramshaw and Marcus, 1995].
 - El chunking se puede plantear como un **problema de etiquetado**, definiendo un conjunto de etiquetas para marcar los chunks.
 - Además, proponen un conjunto de aprendizaje y de prueba estándar, extraído del *Penn Treebank Corpus*, para contrastar diferentes aproximaciones inductivas (Aprendizaje: secciones 15–18, Prueba:sección 20).

Obtención de datos de aprendizaje

¿Cómo obtener/definir los chunks?

¿Cómo definir las etiquetas?

¿Se pueden definir distintos conjuntos de etiquetas equivalentes?

¿Influye el conjunto elegido en el resultado?

Ejemplo de conversión de árboles a chunks

PENN TREEBANK

```
( (S
  (NP-SBJ-1 (PRP You) )
  (VP (MD will)
    (VP (VB start)
      (S
        (NP-SBJ (-NONE- *-1) )
        (VP (TO to)
          (VP (VB see)
            (NP
              (NP (NNS shows) )
              (SBAR
                (WHADVP-2 (WRB where) )
                (S
                  (NP-SBJ (NNS viewers) )
                  (VP (VBP program)
                    (NP (DT the) (NN program) )
                    (ADVP-LOC (-NONE- *T*-2 )))))))))))))
  (. .) ))
```

CHUNKS

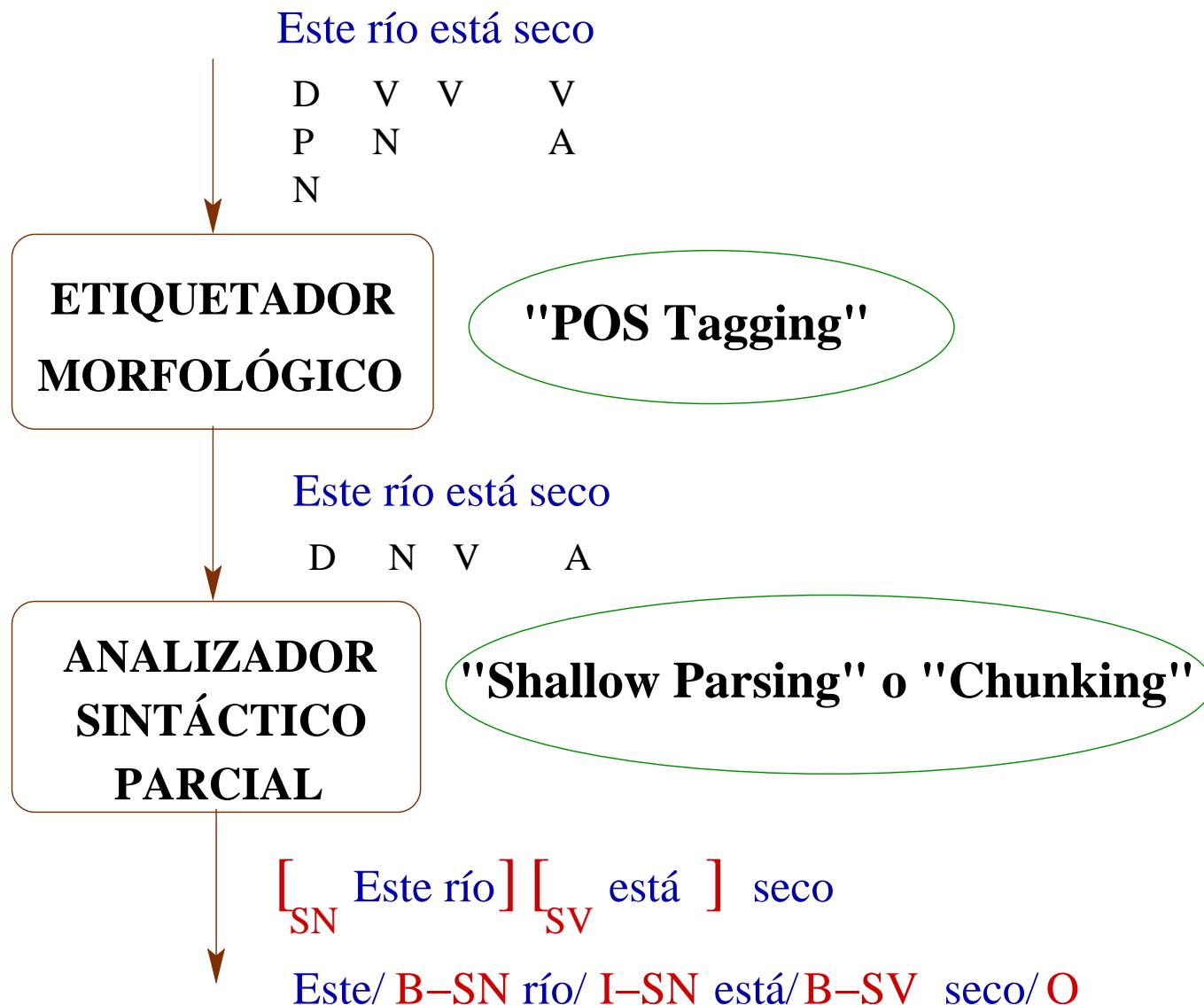
```
[ NP You ]
[ VP will
start
to
see ]
[ NP shows ]
[ ADVP where ]
[ NP viewers ]
[ VP program ]
[ NP the
program ]
.
```

Representación de los chunks

Objetivo: buscar un conjunto de etiquetas equivalente a la segmentación en chunks. [Ramshaw and Marcus, 1995]
[Tjong Kim Sang and Veenstra, 1999]

CHUNKS	PALABRAS	IOB	IOE
[NP You]	You	B-NP	E-NP
[VP will	will	B-VP	I-VP
start	start	I-VP	I-VP
to	to	I-VP	I-VP
see]	see	I-VP	E-VP
[NP shows]	shows	B-NP	E-VP
[ADVP where]	where	B-ADVP	E-ADVP
[NP viewers]	viewers	B-NP	E-NP
[VP program]	program	B-VP	E-VP
[NP the	the	B-NP	I-NP
program]	program	I-NP	E-NP
.	.	O	O

Esquema de un Analizador



Técnicas similares para problemas distintos

Formalismo	POS Tagging	Chunking
HMM	[Church, 1988] [Cutting et al., 1992], [Merialdo, 1994], [Brants, 2000]	[Church, 1988] [Brants, 1999] [Molina and Pla, 2002]
ME	[Ratnaparkhi, 1996]	[Skut and Brants, 1998] [Osborne, 2000],[Koeling, 2000]
TBL	[Brill, 1992, Brill, 1995]	[Ramshaw and Marcus, 1995]
MBL	[Daelemans et al., 1996]	[Veenstra, 1998], [Argamon et al., 1998], [Daelemans et al., 1999]
Comb.	[Van Halteren et al., 1998], [Brill and Wu, 1998]	[Tjong Kim Sang, 2000a]
Winnow		[Li and Roth, 2001] [Zhang et al., 2001]
SVMs		[Kudo and Matsumoto, 2000]

Evaluación

$$\text{Accuracy}(A) = \frac{\# \text{ etiquetas correctas en el análisis propuesto}}{\# \text{ etiquetas en el análisis de referencia}}$$

Medidas usuales: Precisión y Cobertura

$$\text{Precisión}(P) = \frac{\# \text{ constituyentes correctos en el análisis propuesto}}{\# \text{ constituyentes en el análisis propuesto}}$$

$$\text{Cobertura}(C) = \frac{\# \text{ constituyentes correctos en el análisis propuesto}}{\# \text{ constituyentes en el análisis de referencia}}$$

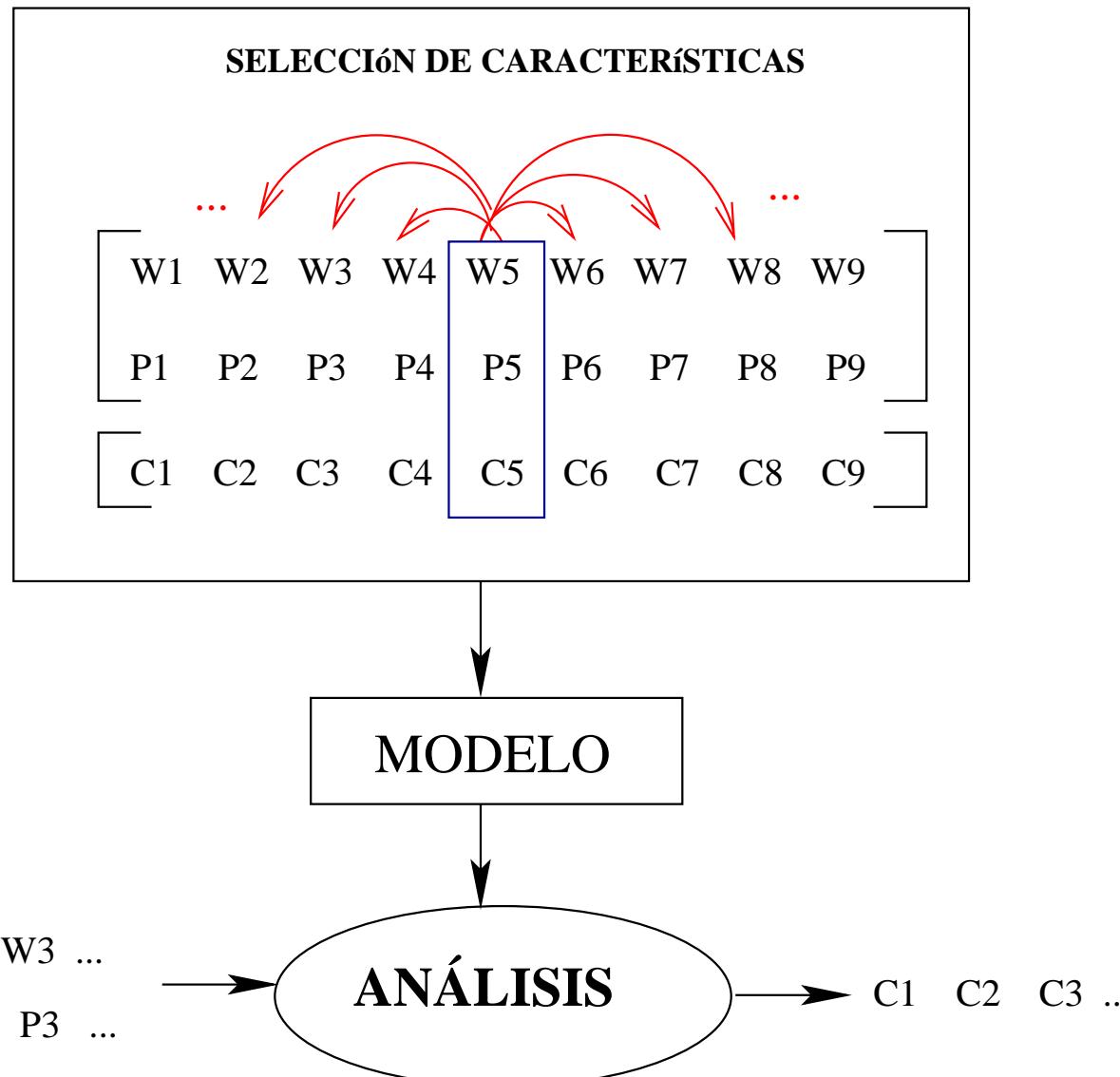
$$F_\beta = \frac{(\beta^2 + 1) \times P \times C}{\beta^2 \times P + C}$$

	Frase	W1	W2	W3	W4	W5	W6	W7
Ejemplo:	Referencia	B-NP	I-NP	B-NP	I-NP	B-VP	B-NP	I-NP
	Salida	B-NP	I-NP	I-NP	I-NP	B-VP	B-NP	I-NP

Evaluación: A=5/7 P=2/3 C=2/4

Aproximaciones basadas en corpus

APRENDIZAJE



Memory-Based Learning (MBL)

- Almacenar en memoria los ejemplos del cjto. de entrenamiento: vectores de características y la etiqueta asociada.
- Etiquetado: problema de clasificación. Una muestra se clasifica en términos de su similitud con los ejemplos en memoria.

IB1-IG: “Overlap Metric” + “Gain Ratio”

$$D(X, Y) = \sum_{i=1}^n \omega_i \sigma(x_i, y_i)$$

$$\sigma(x_i, y_i) = \begin{cases} 0 & \text{si } x_i = y_i \\ 1 & \text{otro caso} \end{cases}$$

Máxima Entropía (ME)

$$P(t|h) = \frac{1}{Z(h)} \prod_{j=1}^k \alpha_j^{f_j(h,t)}$$

- Función de probabilidad que combina un conjunto de **características** $f_j(h, t) \in \{0, 1\}$

$$f_j(h_i, t_i) = \begin{cases} 1 & \text{si } w_i = \text{"the"} \text{ y } t_i = \text{"B-NP"} \\ 0 & \text{otro caso} \end{cases}$$

- Los **parámetros** $\{\alpha_1, \dots, \alpha_k\}$ se eligen para maximizar la probabilidad del conjunto de entrenamiento (**Generalized Iteractive Scaling**).
Equivalente a **maximizar la entropía** de la distribución.

Transformation-Based Learning (TBL)

Método

- Etiquetado inicial
- Definición de un conjunto de reglas o patrones formado por un conjunto limitado de rasgos
- Se compara el etiquetado inicial con la referencia y se instancian los patrones definidos
- Se ordenan y eligen las reglas que producen mejoras en el etiquetado
- Se itera hasta que no se observan mejoras

Ejemplo [Ramshaw and Marcus, 1995]

Si $T_0=I$ y $T_1=O, P_0=JJ$, cambiar $T_0=O$

Si $T_0=-$ y $T_{-2}=I, T_{-1}=I, P_0=DT$ cambiar $T_0=B$

Hidden Markov Models (HMM)

$$\hat{O} = \arg \max_O P(O|I) = \arg \max_O \left(\frac{P(O) \cdot P(I|O)}{P(I)} \right); O \in \mathcal{O}^T$$

$$I = i_1, i_2, \dots, i_T; O = o_1, o_2, \dots, o_T$$

Asumciones de Markov (simplificación): bigramas

$$\hat{O} \approx \arg \max_{o_1 \dots o_T} \left(\prod_{j:1 \dots T} P(o_j|o_{j-1}) \cdot P(i_j|o_j) \right)$$

Probabilidad de transición: $P(o_j|o_{j-1})$

Probabilidad de emisión: $P(i_j|o_j)$

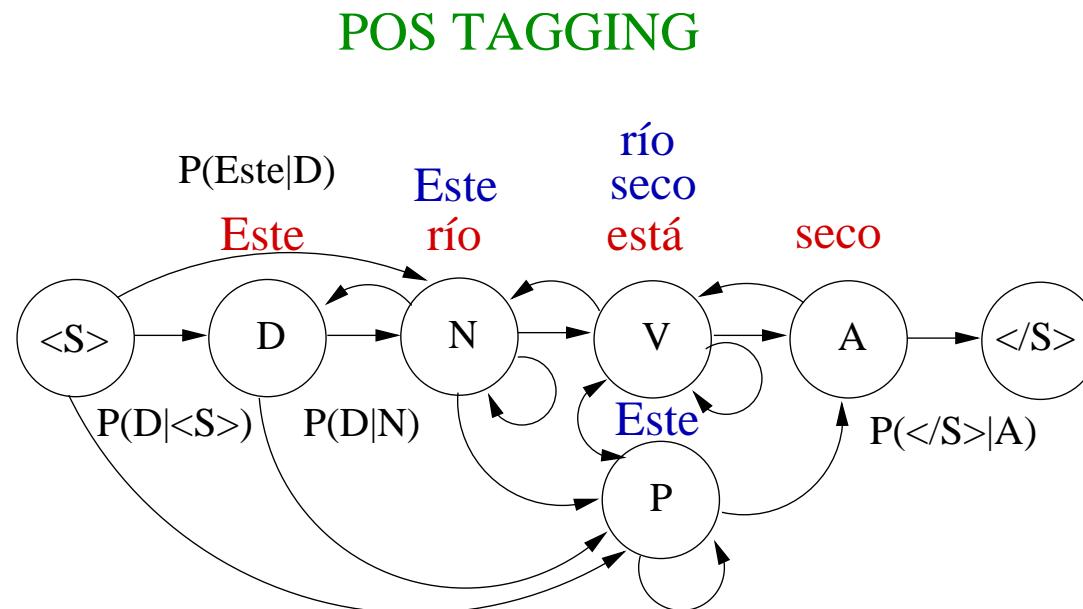
Solución: algoritmo de Viterbi

HMM: representación

ENTRADA

Este río está seco

D	V	V	V
P	N		A
N			



SALIDA

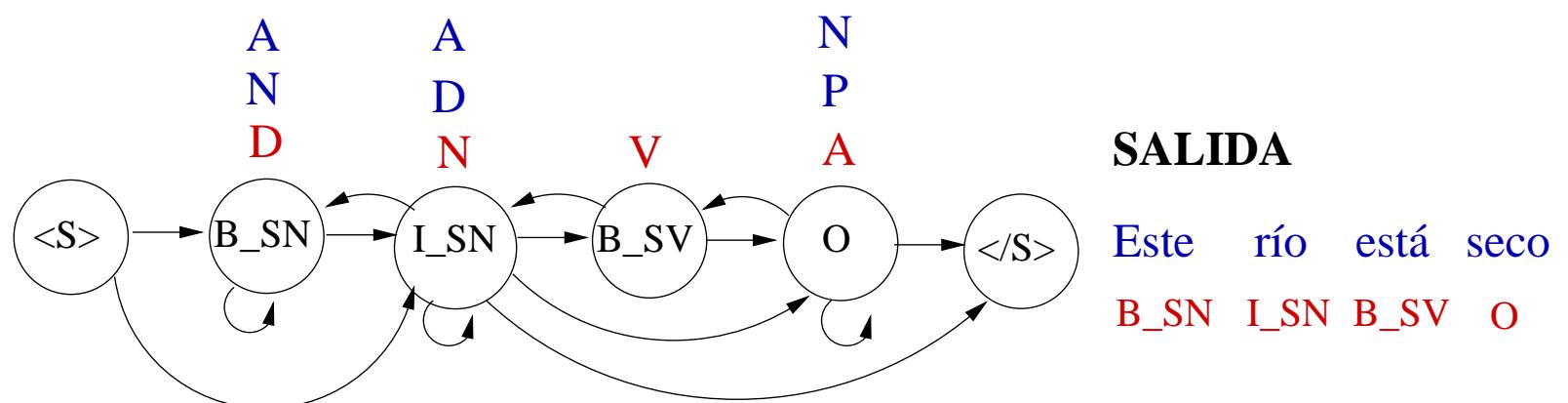
Este río está seco

D	N	V	A
---	---	---	---

ENTRADA

Este río está seco

D	N	V	A
---	---	---	---



SALIDA

Este río está seco

B_SN	I_SN	B_SV	O
------	------	------	---

HMM Especializados [Pla, 2000]

Objetivo: conseguir una mejor modelización (considerando más características) con el fin de reducir la sobregeneralización inherente a los modelos de n-gramas.

Método:

- **Selección:** determinación de la información relevante para el problema.
- **Especialización:** refinamiento del conjunto de etiquetas para obtener modelos más precisos.

Características: no se realizó ningún cambio en el proceso de entrenamiento y etiquetado de una aproximación clásica basada de HMM.

Se realiza una transformación del conjunto de entrenamiento

HMM Especializados: ejemplo

C. ENTRENAMIENTO

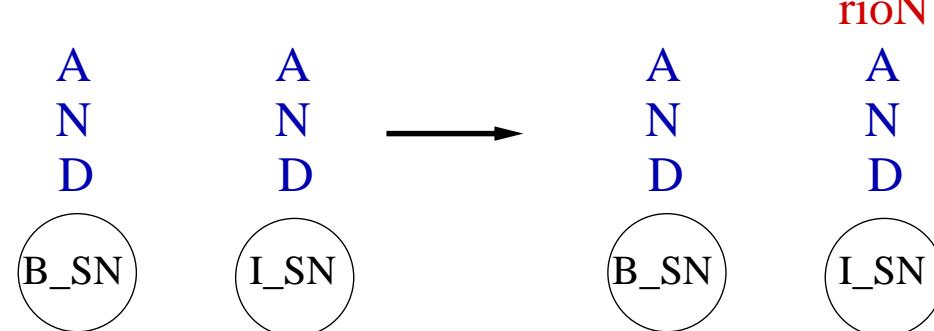
Este río está seco

D	<u>N</u>	V	A	
B-NP		I-NP	B-VP	O

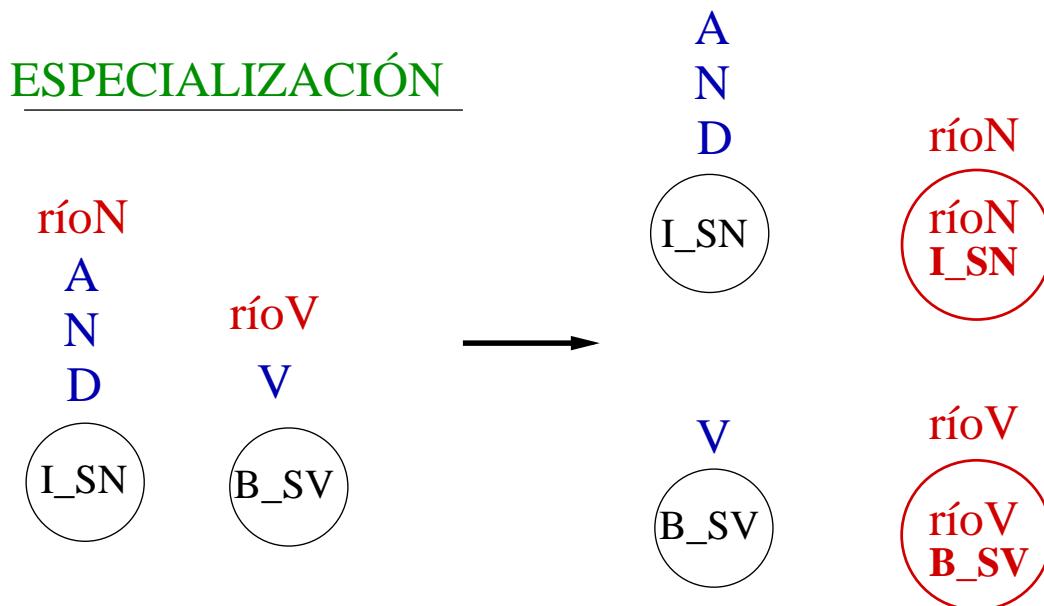
Me río ...

P	<u>V</u>
B-NP B-VP ...	

SELECCIÓN



ESPECIALIZACIÓN



Transformación de un conjunto de entrenamiento

\mathcal{T}			$\tilde{\mathcal{T}}$		
I		O	\tilde{I}		\tilde{O}
Este	D	B-NP	D		D·B-SN
río	N	I-NP	río·N		río·N·I-NP
está	V	B-VP	V		V·B-VP
seco	A	O	A		A·O
.	.	O	.		·O

Determinación experimental de la especialización

[Molina and Pla, 2002]

Elección de rasgos con mayor F_β sobre un conjunto de desarrollo.

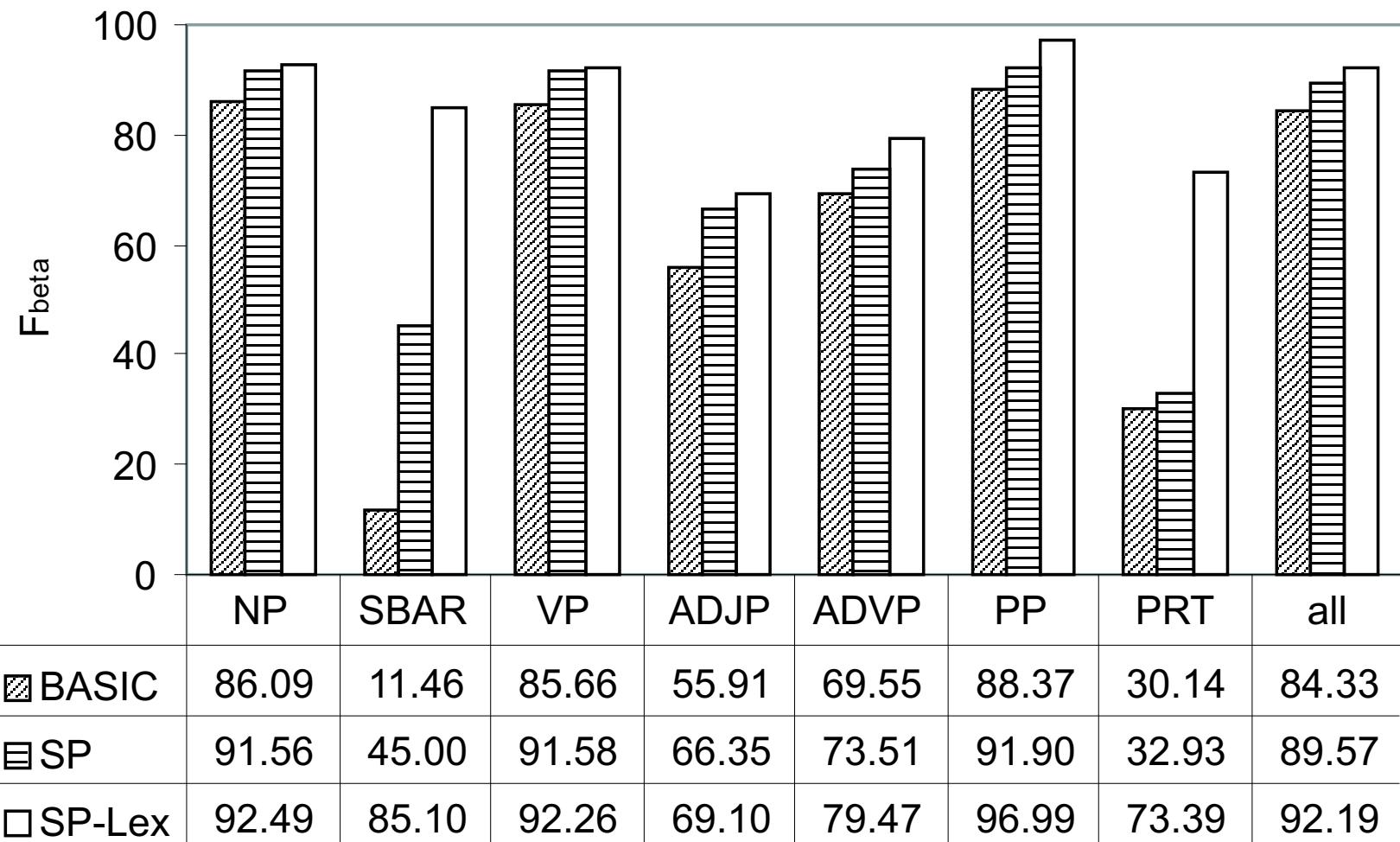
$$f_s(\langle w_i \cdot p_i, ch_i \rangle) = \langle p_i, p_i \cdot ch_i \rangle$$

$$f_{s-Lex}(\langle w_i \cdot p_i, ch_i \rangle) = \begin{cases} \langle w_i \cdot p_i, w_i \cdot p_i \cdot ch_i \rangle & \text{if } w_i \in \mathcal{W}_s \\ \langle p_i, p_i \cdot ch_i \rangle & \text{if } w_i \notin \mathcal{W}_s \end{cases}$$

specialization criteria	precision	recall	$F_{\beta=1}$	$ \tilde{\mathcal{O}} $	$ \mathcal{W}_s $
BASIC	84.16 %	84.52 %	84.34	22	0
Selección	85.45 %	86.93 %	86.18	22	0
SP	90.44 %	89.96 %	90.20	317	0
SP + Lex-WCC	91.99 %	91.56 %	91.77	830	154
SP + Lex-WHF	91.87 %	92.14 %	92.00	1,086	144
SP + Lex-WTE	92.22 %	92.00 %	92.11	592	38
SP + Lex-WCH	92.03 %	92.25 %	92.14	1,305	217
SP + Lex-COM	92.10 %	92.35 %	92.23	1,341	225

Resultados sobre un test no visto

SHMM (trigramas)– F_β



Combinación de clasificadores

- **Winnow**
[Littlestone, 1988] [Zhang et al., 2001]
- **SNoW**
[Roth, 1998]
- **SVMs**
[Vapnik, 1995] [Kudo and Matsumoto, 2000]
- **AdaBoost**
[Freund and Schapire, 1995] [Carreras and Márquez, 2001]
- **Voting:** Majority voting , TotPrec, TagPair, ...
[Tjong Kim Sang, 2000a]

Comparación Experimental

CoNLL2000

TRAIN

- Secciones 15–18 del WSJ
- 211,727 tokens, 106,978 chunks
- NP: 51 %, VP: 20 %, PP: 20 %, ADVP: 4 %, SBAR: 2 %, ADJP: 2 %, PRT: 1 %, ...

TEST

- Sección 20 del WSJ
- 47,337 tokens, 23,852 chunks
- NP: 52 %, PP: 20 %, VP: 20 %, ADVP: 4 %, SBAR: 2 %, ...

Resultados de la Comparación

System	Method	w_L	w_R	p_L	p_R	c_L	F_β
[Kudo and Matsumoto, 2000]	SVM _(C)	2	2	2	2	2	93.91
[Zhang et al., 2001]	Winnow	2	2	2	2	2	93.51
[Van Halteren, 2000]	WPDV _(C)	1-5	1-5	3-5	3	2	93.32
[Li and Roth, 2001]	Winnow						93.02
[Tjong Kim Sang, 2000b]	MBL _(C)	4	4	4	4		92.50
[Molina and Pla, 2002]	SHMM	2		2		2	92.19
[Zhou et al., 2000]	HMM-MBL	1		1		1	92.12
[Déjean, 2000]	Rule			1	1		92.09
[Koeling, 2000]	ME	1	1	3	2	3	91.97
[Osborne, 2000]	ME	2	2	2	2	2	91.94
[Veenstra and v.d. Bosch, 2000]	MBL	5	3	5	3		91.54
[Pla et al., 2000]	HMM	1		1		1	90.14
[Johansson, 2000]	Rule			0-3	0-3		87.23
[Vilain and Day, 2000]	Rule						85.76

- **Eficiencia S-HMM (Pentium III 1 GHz. 256M)**
TRAIN: 22,000w/s TEST: 60,000w/s

Conclusiones sobre las diferentes aproximaciones

- Gran diversidad de aproximaciones.
- Selección de características (rasgos).
- Importancia de la representación.[Tjong Kim Sang, 2002]
- Métodos robustos: tolerancia al ruido.
- Mejores resultados en la detección de chunks que los analizados totales. [Li and Roth, 2001]
- Métodos combinados obtienen buenas prestaciones. Aunque en ciertas condiciones no es así. [Osborne, 2002]

Hacia un análisis más completo

¿Se puede representar estructuras recursivas con un conjunto de etiquetas?

¿Se puede aplicar las técnicas de etiquetado a este problema?

¿Cómo garantizar la consistencia del análisis?

Ejemplo:

Detección de cláusulas

Shared task CoNLL2002

Detección de cláusulas

	\mathcal{T}		$\xrightarrow{f_s}$	$\tilde{\mathcal{T}}$	
	I	O		\tilde{I}	\tilde{O}
You	PRP	B-NP	(S*)	PRP·B-NP	PRP·(S1*)
will	MD	B-VP	*	MD·B-VP	MD·*1
start	VB	I-VP	*	VB·I-VP	VB·*1
to	TO	I-VP	*	TO·I-VP	TO·*1
see	VB	I-VP	*	VB·I-VP	VB·*1
shows	NNS	B-NP	*	NNS·B-NP	NNS·*1
where	WRB	B-ADVP	(S*)	WRB·B-ADVP	WRB·(S2*)
viewers	NNS	B-NP	(S*)	NNS·B-NP	NNS·(S3*)
program	VBP	B-VP	*	VBP·B-VP	VBP·*3
the	DT	B-NP	*	DT·B-NP	DT·*3
program	NN	I-NP	*S)S)	NN·NN	NN·*S3)S2)
.	.	O	*S)	.·O	O·*1S)

Detección de cláusulas: resultados

CoNLL2001

system	precision	recall	$F_{\beta=1}$
[Carreras and Màrquez, 2001]	84.82 %	73.28 %	78.63
[Molina and Pla, 2001]	70.89 %	65.57 %	68.12
[Tjong Kim Sang, 2001]	76.91 %	60.61 %	67.79
[Patrick and Goyal, 2001]	73.75 %	60.00 %	66.17
[Déjean, 2001]	72.56 %	54.55 %	62.77
[Hammerton, 2001]	55.81 %	45.99 %	50.42

Los resultados muestran que todavía queda mucho por mejorar

Conclusiones finales

- Resultados “aceptables” de las aproximaciones de aprendizaje automático.
- Tendencias: cooperación de diferentes paradigmas.
- ¿Cuál es el papel de la Lingüística?

Referencias

- [Abney, 1991] Abney, S. (1991). *Parsing by Chunks*. R. Berwick, S. Abney and C. Tenny (eds.) Principle-based Parsing . Kluwer Academic Publishers, Dordrecht.
- [Abney, 1996] Abney, S. (1996). *Statistical Methods and Linguistics*. Judith Klavans and Philip Resnik (eds.) The Balancing Act. MIT Press, Cambridge, MA.
- [Abney, 1997] Abney, S. (1997). *Part-of-Speech Tagging and Partial Parsing*. S. Young and G. Blothoof (eds.) Corpus-Based Methods in Language and Speech Processing. An ELSNET book. Kluwer Academic Publishers, Dordrecht.
- [Aït-Mokhtar and Chanod, 1997] Aït-Mokhtar, S. and Chanod, J. (1997). Incremental Finite-State Parsing. In *Proceedings of the 5th Conference on Applied Natural Language Processing*, Washington D.C., USA.
- [Allen, 1995] Allen, J. F. (1995). *Natural Language Understanding*. Computer Science. 2nd. ed. Benjamin Cummings.
- [Argamon et al., 1998] Argamon, S., Dagan, I., and Krymolowski, Y. (1998). A Memory-based Approach to Learning Shallow Natural Language Patterns. In *Proceedings of the joint 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics, COLING-ACL*, pages 67–73, Montréal, Canada.
- [Brants, 1999] Brants, T. (1999). Cascaded Markov Models. In *Proceedings of the EACL99*, Bergen, Norway.
- [Brants, 2000] Brants, T. (2000). TnT – a statistical part-of-speech tagger. In *Proceedings of the Sixth Applied Natural Language Processing (ANLP-2000)*, Seattle, WA.
- [Brill, 1992] Brill, E. (1992). A Simple Rule-Based Part-of-speech Tagger. In *Proceedings of the 3rd Conference on Applied Natural Language Processing, ANLP*, pages 152–155. ACL.
- [Brill, 1995] Brill, E. (1995). Transformation-based Error-driven Learning and Natural Language Processing: A Case Study in Part-of-speech Tagging. *Computational Linguistics*, 21(4):543–565.
- [Brill and Wu, 1998] Brill, E. and Wu, J. (1998). Classifier Combination for Improved Lexical Disambiguation. In *Proceedings of the joint 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics, COLING-ACL*, pages 191–195, Montréal, Canada.
- [Carreras and Màrquez, 2001] Carreras, X. and Màrquez, L. (2001). Boosting trees for clause splitting. In Daelemans, W. and Zajac, R., editors, *Proceedings of CoNLL-2001*, pages 73–75. Toulouse, France.
- [Charniak, 1997] Charniak, E. (1997). Statistical Techniques for Natural Language Parsing. *AI Magazine*. <http://www.cs.brown.edu/people/ec>.
- [Church, 1988] Church, K. W. (1988). A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text. In *Proceedings of the 1st Conference on Applied Natural Language Processing, ANLP*, pages 136–143. ACL.

- [Collins, 1996] Collins, M. (1996). A new Statistical Parser based on Bigram Lexical Dependencies. In *Proceedings of the 34th Annual Meeting of the Association for Computational Linguistics*, Santa Cruz, CA. ACL.
- [Cutting et al., 1992] Cutting, D., Kupiec, J., Pederson, J., and Sibun, P. (1992). A Practical Part-of-speech Tagger. In *Proceedings of the 3rd Conference on Applied Natural Language Processing, ANLP*, pages 133–140. ACL.
- [Daelemans et al., 1999] Daelemans, W., Buchholz, S., and Veenstra, J. (1999). Memory-Based Shallow Parsing. In *Proceedings of EMNLP/VLC-99*, pages 239–246, University of Maryland, USA.
- [Daelemans et al., 1996] Daelemans, W., Zavrel, J., Berck, P., and Gillis, S. (1996). MBT: A Memory-Based Part-of-speech Tagger Generator. In *Proceedings of the 4th Workshop on Very Large Corpora*, pages 14–27, Copenhagen, Denmark.
- [Déjean, 2000] Déjean, H. (2000). Learning Syntactic Structures with XML. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Déjean, 2001] Déjean, H. (2001). Using allis for clauising. In Daelemans, W. and Zajac, R., editors, *Proceedings of CoNLL-2001*, pages 64–66. Toulouse, France.
- [Ejherhed, 1988] Ejherhed, E. (1988). Finding Clauses in Unrestricted Text by Finitary and Stochastic Methods . In *Proceedings of Second Conference on Applied Natural Language Processing*, pages 219–227. ACL.
- [Ferrández et al., 1998] Ferrández, A., Palomar, M., and Moreno, L. (1998). Anaphor resolution in unrestricted texts with partial parsing. In *Proceedings of COLING-ACL'98*, pages "385–391", Montréal, Canada.
- [Freund and Schapire, 1995] Freund, Y. and Schapire, R. E. (1995). A Decision-Theoretic Generalization of On-line Learning and an Application to Boosting. In *Proceedings of the 2nd European Conference on Computational Learning Theory, EuroCOLT'95*, Barcelona, Spain. Also in: *Journal of Computer and System Sciences* 55:1, 1997, 119–139.
- [Fu and Booth, 1975] Fu, K. and Booth, T. (1975). Grammatical Inference: introduction and survey. Parts I and II. *IEEE Transactions on Systems, Man and Cybernetics*, 5:303–309,409–423.
- [González and Thomason, 1978] González, R. and Thomason, M. (1978). *Syntactic Pattern Recognition*. Addison-Wesley Publishing Company.
- [Hammerton, 2001] Hammerton, J. (2001). Clause identification with long short-term memory. In Daelemans, W. and Zajac, R., editors, *Proceedings of CoNLL-2001*, pages 61–63. Toulouse, France.
- [Johansson, 2000] Johansson, C. (2000). A Context Sensitive Maximum Likelihood Approach to Chunking. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Koeling, 2000] Koeling, R. (2000). Chunking with Maximum Entropy Models. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Kudo and Matsumoto, 2000] Kudo, T. and Matsumoto, Y. (2000). Use of Support Vector Learning for Chunk Identification. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Li and Roth, 2001] Li, X. and Roth, D. (2001). Exploring Evidence for Shallow Parsing. In *Proceedings of the 5th Conference on Computational Natural Language Learning (CoNLL-2001)*, Toulouse, France.

- [Littlestone, 1988] Littlestone, N. (1988). Learning Quickly when Irrelevant Attributes Abound. *Machine Learning*, 2:285–318.
- [Marcus et al., 1993] Marcus, M. P., Marcinkiewicz, M. A., and Santorini, B. (1993). Building a Large Annotated Corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2). <http://morph.ldc.upenn.edu/Catalog/docs/treebank2/cl93.html>.
- [Merialdo, 1994] Merialdo, B. (1994). Tagging English Text with a Probabilistic Model. *Computational Linguistics*, 20(2):155–171.
- [Molina and Pla, 2001] Molina, A. and Pla, F. (2001). Clause detection using HMM. In *Proceedings of the 5th Conference on Computational Natural Language Learning (CoNLL-2001)*, Toulouse, France.
- [Molina and Pla, 2002] Molina, A. and Pla, F. (2002). Shallow Parsing using Specialized HMMs. *Journal of Machine Learning Research*, 2:595–613.
- [Molina et al., 1999] Molina, A., Pla, F., Moreno, L., and Prieto, N. (1999). APOLN: A Partial Parser of Unrestricted Text. In *Proceedings of 5th Conference on Computational Lexicography and Text Research COMPLEX-99*, pages 101–108, Pecs, Hungary.
- [Moreno et al., 1999] Moreno, L., Palomar, M., Molina, A., and Ferrández, A. (1999). *Introducción al Procesamiento del Lenguaje Natural*. Servicio de Publicaciones Universidad de Alicante.
- [Osborne, 2000] Osborne, M. (2000). Shallow Parsing as Part-of-Speech Tagging. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Osborne, 2002] Osborne, M. (2002). Shallow Parsing using Noisy and Non-Stationary Training Material. *Journal of Machine Learning Research*, 2:695–718.
- [Patrick and Goyal, 2001] Patrick, J. D. and Goyal, I. (2001). Boosted decision graphs for nlp learning tasks. In Daelemans, W. and Zajac, R., editors, *Proceedings of CoNLL-2001*, pages 58–60. Toulouse, France.
- [Pla, 2000] Pla, F. (2000). *Etiquetado Léxico y Análisis Sintáctico Superficial basado en Modelos Estadísticos*. Phd. Thesis, Dep. de Sistemes Informàtics i Computació. Universitat Politècnica de València.
- [Pla et al., 2000] Pla, F., Molina, A., and Prieto, N. (2000). Improving Chunking by means of Lexical-Contextual Information in Statistical Language Models. In *Proceedings of ConNL-2000*, Lisbon, Portugal.
- [Ramshaw and Marcus, 1995] Ramshaw, L. and Marcus, M. (1995). Text Chunking Using Transformation-Based Learning. In *Proceedings of third Workshop on Very Large Corpora*, pages 82–94. <ftp://ftp.cis.upenn.edu/pub/chunker/wvlcbook.ps.gz>.
- [Ratnaparkhi, 1996] Ratnaparkhi, A. (1996). A Maximum Entropy Part-of-speech Tagger. In *Proceedings of the 1st Conference on Empirical Methods in Natural Language Processing, EMNLP*.
- [Ratnaparkhi, 1997] Ratnaparkhi, A. (1997). A Linear Observed Time Statistical Parser Based on Maximum Entropy Models. In *Proceedings of the 2nd Conference on Empirical Methods in Natural Language Processing, EMNLP*, Brown University, Providence, RI.
- [Roth, 1998] Roth, D. (1998). Learning to Resolve Natural Language Ambiguities: A Unified Approach. In *Proceedings of the National Conference on Artificial Intelligence, AAAI '98*. <http://l2r.cs.uiuc.edu/~danr>.

- [Skut and Brants, 1998] Skut, W. and Brants, T. (1998). A Maximum–Entropy Partial Parser for Unrestricted Text. In *Proceedings of the 6th Workshop on Very Large Corpora*, Montréal, Canada. cmp-lg/9807006.
- [Tjong Kim Sang, 2000a] Tjong Kim Sang, E. F. (2000a). Noun Phrase Representation by System Combination. In *Proceedings of ANLP-NAACL 2000*, Washington, USA. Morgan Kaufman Publishers. <http://lcg-www.uia.ac.be/~erikt/papers/naacl2000.ps>.
- [Tjong Kim Sang, 2000b] Tjong Kim Sang, E. F. (2000b). Text Chunking by System Combination. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Tjong Kim Sang, 2001] Tjong Kim Sang, E. F. (2001). Memory-based clause identification. In Daelemans, W. and Zajac, R., editors, *Proceedings of CoNLL-2001*, pages 67–69. Toulouse, France.
- [Tjong Kim Sang, 2002] Tjong Kim Sang, E. F. (2002). Memory-based shallow parsing. *Journal of Machine Learning Research*, 2:559–594.
- [Tjong Kim Sang and Veenstra, 1999] Tjong Kim Sang, E. F. and Veenstra, J. (1999). Representing Text Chunks. In *Proceedings of EACL'99*, Association for Computational Linguistics. <http://xxx.lanl.gov/abs/cs.CL/9907006>.
- [Van Halteren, 2000] Van Halteren, H. (2000). Chunking with WPDV Models. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Van Halteren et al., 1998] Van Halteren, H., Zavrel, J., and Daelemans, W. (1998). Improving Data Driven Wordclass Tagging by System Combination. In *Proceedings of the joint 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Computational Linguistics*, COLING-ACL, pages 491–497, Montréal, Canada.
- [Veenstra, 1998] Veenstra, J. (1998). Fast NP Chunking Using Memory-based Learning Techniques. In *Proceedings of BENELEARN-98: Eighth Belgian-Dutch Conference on Machine Learning*, Wageningen, the Netherlands. <ftp://ilk.kub.nl/pub/papers/ilk.9807.ps.gz>.
- [Veenstra and v.d. Bosch, 2000] Veenstra, J. and v.d. Bosch, A. (2000). Single-Classifier Memory-Based Phrase Chunking. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Vilain and Day, 2000] Vilain, M. and Day, D. (2000). Phrase Parsing with Rule Sequence Processors: an Application to the Shared CoNLL Task. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.
- [Zhang et al., 2001] Zhang, T., Damerau, F., and Johnson, D. (2001). Text chunking using regularized Winnow. In *proceedings of the Joint EACL-ACL Meeting (ACL2001)*, Toulouse, France.
- [Zhou et al., 2000] Zhou, G., Su, J., and Tey, T. (2000). Hybrid Text Chunking. In *Proceedings of CoNLL-2000 and LLL-2000*, Lisbon, Portugal.