

12th Conference of the European Chapter of the Association for Computational Linguistics

Franco M. Luque

Grupo de Procesamiento de Lenguaje Natural
Fa.M.A.F., Universidad Nacional de Córdoba

CONICET

6 de mayo de 2009

Resumen de la Charla

- 1 Introducción
- 2 Workshop CLAGI
- 3 Tutorial CCGs
- 4 Conferencia Principal
- 5 Conclusiones

Introducción

Atenas

- 3,5 millones de habitantes, 3,5 miles de años de historia.
- Ruinas, iglesias ortodoxas, anarquistas, turistas y africanos.
- Gente piola, pronunciación parecida, y buenos precios.
- Buena infraestructura. Mucho transporte público.

EACL 2009

- Conferencia trianual. Esta fue la doceava.
- La EACL más grande en términos de cantidad de papers presentados.

Computational Linguistic Aspects of Grammatical Inference

Menno van Zaanen (Tilburg, Holanda). Alignment-Based Learning.
Colin de la Higuera (Saint-Etienne, Francia). Mucha IG, poco PLN.

- Objetivo: Juntar a la gente de IG con la de PLN.
- Endogámico: 12 expositores aprox. y no muchos más participantes.
- Un poco tirado de los pelos. Temas muy distintos entre sí.
- Las charlas: Entre aburridas e inentendibles (para mí).
- Mi charla: Igual. Algunos entendieron.
- Panel final: En realidad era para proponer una *shared task* sobre GI que fuera atractiva para la comunidad de CL. No les dieron/dimos mucha bola.

Combinatory Categorical Grammars for Robust Natural Language Processing

Mark Steedman (Edimburgo, Escocia). Inventor de las CCGs.

El problema

- Ciertas construcciones poco frecuentes contribuyen mucho a determinar la aceptabilidad de un sistema.
- Ejemplos: *Argument cluster coordinations*, (“It could cost taxpayers \$15 million and BPC residents \$1 million.”) dependencias cruzadas y dependencias de largo rango.
- Las técnicas estándar de Machine Learning son malas para obtener sistemas que se comporten bien ante eventos raros.
- Las CFGs no son lo suficientemente expresivas para capturar estos fenómenos.
- Las métricas como eval-b no sirven para evaluar esto.

Combinatory Categorical Grammars

Por qué usarlas

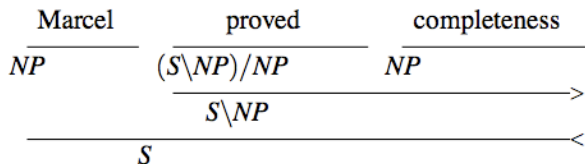
- Son Nearly Context Free, la generalización menos expresiva de las CFGs que se conoce dentro de la jerarquía de Chomsky.
- Permite capturar muchas construcciones importantes que las CFGs no.
- Se puede parsear en tiempo polinomial con programación dinámica.

Qué son

- Se reemplazan las reglas sintácticas por categorías léxicas y un conjunto fijo de reglas combinatorias.
- El orden en que vamos combinando las palabras usando reglas, la derivación, determina la estructura sintáctica.
- Las reglas permiten una gran cobertura y mucha ambigüedad.

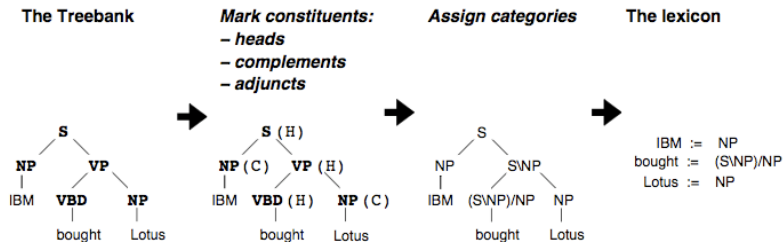
Ejemplo

- $\text{proved} := (S \setminus NP) / NP$
 (“proved” toma un NP a derecha y un NP a izquierda y devuelve una S.)
- $\text{Marcel, completeness} := NP$
- $\text{Marcel proved completeness} := NP (S \setminus NP) / NP NP \rightarrow S$



Experimentos

- Extraer un lexicón CCG del Penn Treebank (ver figura).
- Crear un CCG-bank (con derivaciones CCG en vez de árboles).
- Desarrollar modelos estadísticos generativos y discriminativos.
- Desarrollar un *supertagger* para asignar categorías CCG a las palabras. *Multitagging*: asigna varias categorías por palabra (las más probables). Integrarlo a los parsers.



Resultados y Trabajo Futuro

C&C parser (Clark and Curran, 2004)

- Modelo discriminativo (condicional log-linear con features).
- Combinado con un supertagger.
- El parsing es muy rápido (gracias al supertagging).
- Logra estado del arte en parsing de dependencias en general.
- Captura muchas dependencias de largo rango.

Trabajo Futuro

- Combinar el modelo generativo con el supertagger.
- Construir interpretaciones semánticas utilizando lógicas de primer orden como FOPL o DRTs.

Conferencia Principal

Algunas Charlas (a las que asistí)

- Unsupervised Methods for Head Assignments (Federico Sangati and Willem Zuidema).
- Incremental Parsing with Parallel Multiple Context-Free Grammars (Krasimir Angelov).
- Improving Grammaticality in Statistical Sentence Generation: Introducing a Dependency Spanning Tree Algorithm with an Argument Satisfaction Model (Stephen Wan et. al.).
- EM Works for Pronoun Anaphora Resolution (Eugene Charniak and Micha Elsner).

Unsupervised Methods for Head Assignments

Federico Sangati and Willem Zuidema

- Se basa en que un treebank anotado con heads define una única Lexicalized Tree Substitution Grammar (LTSG).
- Luego define algoritmos para inducir LTSGs.
- Dos algoritmos: minimización de entropía y una técnica greedy llamada “maximización de familiaridad”.
- Baseline: reglas de Magerman y Collins, leftmost y rightmost.
- Mejores resultados (el de familiaridad):
 - Contra dependency banks: supera todo menos M&C.
 - Con un parser LTSG (entrena con la LTSG inducida): el mejor.
 - Con el parser de Bikel: tan bien como M&C.

Incremental Parsing with Parallel Multiple Context-Free Grammars

Krasimir Angelov

- PMCFGs: extensión de las CFGs que en las rhs's lleva una tupla de strings. Permite modelar constituyentes discontinuos.
- Da un algoritmo incremental y top-down para parsearlas.
- Las PMCFGs se pueden ver como CFGs con infinitas reglas. El algoritmo explota esto.
- Hay pruebas formales de soundness y de completitud.
- Los Tree Adjoining Grammars y varios modelos más (GF, LCFRS, FCTT) se pueden convertir a PMCFGs o sea que el algoritmo sirve para ellos.

Improving Grammaticality in Statistical Sentence Generation: Introducing a Dependency Spanning Tree Algorithm with an Argument Satisfaction Model

Stephen Wan et. al.

- Supone que tiene elegido un cjto. de palabras (trabajo previo).
- Usa modelos estadísticos de dependencias entrenados con el PTB. Modelan cantidad y posición de los argumentos.
- Busca el ordenamiento de las palabras que da el árbol de dependencias más probable.
- El problema se traduce a *weighted bipartite graph matching*.
- Evaluación: *string regeneration task*: tomar oraciones desordenadas y encontrar el ordenamiento original. Usa métrica BLEU.
- Resultados: muy buenos respecto de baselines. No tiene otros trabajos con los que compararse.

EM Works for Pronoun Anaphora Resolution

Eugene Charniak and Micha Elsner

- Esta charla me gustó por lo simple y clara.
- Personal Pronoun Anaphora Resolution: Dado un PP decidir a qué entidad ya vista en el texto se refiere.
- Define un modelo generativo de PP en términos de las entidades vistas. Modela generación de persona, género y número de los pronouns.
- Utiliza EM sobre un subset del PTB para aprender los parámetros del modelo.
- Evaluación: contra otros sistemas de resolución de anáforas.
- Resultados: les pasa el trapo a todos.

Conclusiones

- Muy buena experiencia.
- Nivel muy variable de los trabajos. Sólo unos pocos realmente interesantes.
- No vi nada parecido a lo que hago yo.
- Faltaron actividades sociales (sólo un banquete de 50 euros al que no fui).
- Es muy difícil presentar en 4 slides 4 papers que no leí.
- Se usa poner un slide al final que diga “Gracias”.

*Ευχαριστω*¹
¡Gracias!