

# Técnicas Aplicadas al Reconocimiento de Implicación Textual

Jesús Herrera, Anselmo Peñas, Felisa Verdejo

Departamento de Lenguajes y Sistemas Informáticos  
Universidad Nacional de Educación a Distancia  
Madrid, España  
{jesus.herrera, anselmo, felisa}@lsi.uned.es

**Resumen** Tras establecer qué se entiende por implicación textual, se expone la situación actual y el futuro deseable de los sistemas dirigidos a reconocerla. Se realiza una identificación de las técnicas que implementan actualmente los principales sistemas de Reconocimiento de Implicación Textual.

## 1. Introducción

El término “implicación textual” se utiliza para indicar la situación en la que la semántica de un texto en lenguaje natural se puede inferir de la semántica de otro texto en lenguaje natural. Más específicamente, si la verdad de un enunciado implica la verdad de otro enunciado, llamado también hipótesis. Sean los dos textos siguientes:

1. El encuentro del G8, de tres días, tendrá lugar en Escocia.
2. La reunión del Grupo de los Ocho durará tres días.

es evidente que la semántica del segundo se puede inferir de la semántica del primero; se dice, pues, que existe implicación textual entre ambos textos (el primero implica al segundo). También puede observarse que el reconocimiento de implicación textual requiere un procesamiento tanto a nivel léxico (por ejemplo, sinonimia entre *encuentro* y *reunión* o entre *G8* y *Grupo de los Ocho*) como sintáctico.

El reconocimiento de implicación textual (RTE, Recognising Textual Entailment) se ha empezado a tener en cuenta internacionalmente como tarea de interés y desarrollo independiente en el *Recognising Textual Entailment Challenge*, organizado por primera vez entre junio de 2004 y abril de 2005 por la Red de Excelencia PASCAL<sup>1</sup>. Aún así, de un modo u otro el reconocimiento de la implicación entre textos en lenguaje natural ha sido objeto de estudio en los últimos años, bien como parte de sistemas más complejos bien como aplicación independiente. Lin y Pantel [17] ofrecen una clasificación de los campos en los que había sido de utilidad el reconocimiento de equivalencia semántica (variantes y paráfrasis), identificando los siguientes:

<sup>1</sup> Pattern Analysis, Statistical Modeling and Computational Learning.  
<http://www.pascal-network.org/>

- Generación de lenguaje: donde se han focalizado los esfuerzos básicamente en las transformaciones de texto basadas en reglas, para satisfacer restricciones externas como la longitud y la legibilidad.
- Resumen automático: en el que detectar el parafraseo es importante para evitar redundancias en los resúmenes.
- Recuperación de información: donde es común generar o buscar variantes de los términos de la consulta en los textos (por ejemplo, para expansión de la consulta).
- Minería de textos: en la que se intenta encontrar reglas de asociación semántica entre términos.

A las anteriores se pueden añadir otras aplicaciones como validación de respuestas en pregunta-respuesta o comparación de traducciones en traducción automática.

A pesar de que diversas aplicaciones actualmente existentes en los ámbitos del Procesamiento del Lenguaje Natural y el Acceso a Información necesitan determinar la existencia de relaciones de equivalencia e implicación entre fragmentos de texto en lenguaje natural, no existe una sistematización de técnicas y herramientas para realizar esta labor de manera específica. El interés a largo plazo de los sistemas de RTE es que funcionen de manera genérica para dar servicio a toda la gama de aplicaciones que necesitan determinar implicaciones entre fragmentos de texto en lenguaje natural.

La experiencia dada por la participación en el PASCAL RTE *Challenge*, así como los trabajos que previamente fueron desarrollados en el ámbito del RTE, sirven de punto de partida para analizar una serie de técnicas, recursos y herramientas que han sido aplicados para el desarrollo de los mismos.

## 2. Trabajos Previos en RTE

El concepto de implicación entre textos es un clásico en los sistemas de Lenguaje Natural. De hecho, toda la teoría de dependencia conceptual de Schwank se dirigía a formalizar este tipo de inferencias. El advenimiento del acceso masivo a información textual en formato digital ha supuesto un éxito de los métodos empíricos, como el análisis estadístico y el aprendizaje automático. Estos métodos suelen aplicarse a un nivel de representación del conocimiento bastante superficial.

Los trabajos realizados sobre implicación textual en los últimos años tienen orígenes y motivaciones diversos. Dekang Lin y Patrick Pantel [17] propusieron un método no supervisado de extracción de reglas de inferencia a partir de texto (algoritmo DIRT), del tipo “*X es autor de Y*” = “*X escribió Y*”, “*X resolvió Y*” = “*X encontró una solución a Y*” o “*X causó Y*” = “*Y lo provocó X*”. Su algoritmo está basado en una versión extendida de la Hipótesis de Distribución de Harris [12], que expone que las palabras que ocurren en los mismos contextos tienden a ser similares; en lugar de utilizar esta hipótesis aplicada a palabras, la aplicaron a caminos en árboles de dependencias extraídos de un corpus.

El trabajo de Lin y Pantel estaba dirigido a simplificar la tarea de creación de bases de conocimiento de ese tipo de reglas, que habitualmente se realiza de manera manual y es muy laboriosa.

Christof Monz y Maarten de Rijke [18], defendiendo un tratamiento superficial de los textos en lugar de utilizar representaciones complejas a la hora de determinar implicaciones entre textos, construyeron un modelo basado en bolsas de palabras para tratar el problema de implicación.

Hristo Tanev, Milen Kouylekov y Bernardo Magnini [23] desarrollaron un sistema de reconocimiento de implicación textual para utilizarlo como subsistema de otro de pregunta-respuesta, que presentaron a la tarea de pregunta-respuesta del TREC<sup>2</sup> 2004; el objetivo a largo plazo era disponer de un sistema de pregunta-respuesta cuyo núcleo fuese un motor de implicaciones, capaz de realizar inferencias a partir de una gran base de datos de reglas de implicación. En esta línea de investigación estaban en colaboración con la Universidad israelí Bar Ilan, lo que dio como resultado conjunto un módulo para la extracción automática de reglas de implicación a partir de la *web* [22].

### 3. ¿Cómo Son Actualmente los Sistemas de RTE?

El PASCAL RTE *Challenge* tenía como meta proporcionar una primera oportunidad para presentar y comparar diferentes aproximaciones para modelizar y reconocer la implicación textual. La tarea a resolver por los sistemas en este *Challenge* era la detección automática de implicación semántica entre parejas de textos en lenguaje natural (monolingüe inglés). Para ello, los organizadores proporcionaron a los participantes sendos corpora de entrenamiento y de prueba, compuestos por pares de textos cortos en lenguaje natural pertenecientes al dominio de las noticias de prensa. Los componentes de cada par de fragmentos textuales se denominaron “texto” e “hipótesis”, respectivamente. Los sistemas debían detectar si el significado de la hipótesis se podía inferir del significado del texto; es decir, el sentido de la implicación estaba predeterminado.

Los pares <texto, hipótesis> que conformaban los corpora proporcionados a los participantes del PASCAL RTE *Challenge* habían sido elegidos de modo que cubriesen características propias de diferentes aplicaciones de procesamiento de texto, obteniéndose la siguiente clasificación: Recuperación de Información, Documentos Comparables, Lectura Comprensiva, Pregunta-Respuesta, Extracción de Información, Traducción Automática y Adquisición de Paráfrasis. A partir de esta primera clasificación, resulta una tarea de interés estudiar el comportamiento de los sistemas actuales para determinar su validez y proponer futuros cambios.

Al PASCAL RTE *Challenge* se presentaron 16 equipos de todo el mundo, proponiendo modelos para abordar el problema de RTE. A continuación se revisan las diversas técnicas, recursos y herramientas de las que se sirven los sistemas de RTE.

---

<sup>2</sup> Text REtrieval Conference. <http://trec.nist.gov/>

### 3.1. Técnicas Lingüísticas

De una manera u otra, todas las técnicas de procesamiento lingüístico son susceptibles de ser incorporadas en un sistema de RTE. A continuación se muestran las utilizadas por los participantes en el PASCAL RTE *Challenge*.

**Preprocesamiento.** Aparte de la necesaria identificación de *tokens*, existen sistemas que realizan un preprocesamiento de los textos antes de aplicarles un análisis morfosintáctico, que se sitúa en el primer nivel dentro del procesamiento lingüístico. Este preprocesamiento se corresponde, en la mayoría de los casos, con la segmentación de oraciones y sintagmas; ésta ha sido utilizada bien como preparación para el análisis morfológico bien para la creación de estructuras de representación de los textos.

Un ejemplo es el del MITRE<sup>3</sup> [3], que previamente al análisis morfológico efectuado por su sistema, aplica a los textos y las hipótesis un proceso de segmentación de frases.

El sistema de la Universidad de Concordia [2] no llega a realizar un análisis morfológico, utilizando la segmentación de sintagmas nominales como apoyo para crear estructuras de predicados con argumentos. Se crea una estructura por cada texto e hipótesis y se establece la similitud entre las estructuras de cada par de fragmentos textuales para determinar si existe implicación entre ambos.

**Análisis morfológico-léxico.** Dentro de éste se diferencian: la extracción de lemas o *stems*, el etiquetado de categorías gramaticales, el uso de analizadores morfológicos y la extracción de relaciones impuestas por la morfología derivacional.

El análisis morfológico se ha utilizado como un primer procesamiento de los textos para obtener información con la que alimentar etapas posteriores que permitan evaluar la implicación entre textos.

La **extracción de lemas** es una técnica que se ha utilizado con cierta profusión y que, en algunos casos, supone gran parte del procesamiento total que realiza el sistema de RTE. Además de ser necesaria para acceder a recursos léxicos como diccionarios, lexicones o *wordnets*, la lematización ha sido utilizada con tres fines diferentes: para evaluar la coincidencia de lemas en medidas de similitud tratando los textos como bolsas de palabras, como atributos de representaciones en forma de grafo de los textos y para ajustar parámetros en algoritmos de evaluación de similitud. Así, por ejemplo, en el sistema de las universidades de Edimburgh y Leeds [4] la lematización es el procesamiento del lenguaje más elaborado que se realiza y, tras él, sólo se aplica una medida de solapamiento entre lemas de la hipótesis y el texto para determinar la existencia de implicación entre ambos. El sistema de la Universidad de Illinois at Urbana-Champaign [7] utiliza los lemas como parte de los atributos asociados a los nodos de árboles conceptuales con los que representan tanto los textos como las hipótesis. La Universidad “Tor Vergata” de Roma, asociada con la Universidad de Milano-Bicocca

<sup>3</sup> The MITRE Corporation, Estados Unidos.

[20], ha desarrollado un sistema en el que se aplica el análisis morfológico para la extracción de lemas que, junto con los *tokens* y otros elementos, se utilizan para ajustar – mediante un algoritmo de aprendizaje SVM – los parámetros de la medida global de similitud entre los dos grafos con los que representan el texto y la hipótesis.

La **extracción de *stems*** ha sido una técnica utilizada básicamente para alimentar otros módulos del sistema. El uso de *stems* en el caso monolingüe inglés está justificado por el buen comportamiento que han demostrado, dada la simplicidad de la morfología del inglés; en el futuro, cuando se desarrollen sistemas de RTE para otras lenguas será necesario evaluar la posibilidad de trabajar sólo con *stems* o, por el contrario, se habrá de hacer uso de los lemas. Como ejemplo, el sistema de las universidades “Tor Vergata” y Milano-Bicocca [20] utiliza los *stems* comparándolos adecuadamente para obtener una medida de subsunción de nodos de los grafos con los que representan la información textual. Con esta medida, junto con otra de subsunción de vértices, determinan la subsunción global entre los grafos que representan el texto y la hipótesis; la medida de subsunción global sirve para establecer la implicación entre el texto y la hipótesis.

El **etiquetado de categorías gramaticales** ha sido utilizado de dos maneras: el sistema del MITRE [3] y el de la Universidad Ca’ Foscari y el ITC-irst<sup>4</sup> [8] lo incluyen como uno de los módulos de análisis lingüístico en cascada; pero la Universidad de Illinois at Urbana-Champaign [7] utiliza las categorías gramaticales como parte de los atributos asociados a los nodos de árboles conceptuales con los que representan tanto los textos como las hipótesis.

El **uso de analizadores morfológicos** como tales sólo se ha dado en el sistema del MITRE [3], que aplica un analizador morfológico (Minnon et al., 2001) cuya acción se suma a la del etiquetado de categorías gramaticales, para alimentar con sus resultados a las etapas siguientes (un analizador sintáctico de constituyentes, un analizador de dependencias y un generador de proposiciones lógicas).

La **extracción de relaciones impuestas por la morfología derivacional** es otra técnica poco utilizada, que encuentra un ejemplo en el sistema de la Language Computer Corporation [9], que extrae relaciones entre palabras a partir de la morfología derivacional contenida en *WordNet*.

**Reconocimiento de multipalabras.** No es una técnica muy extendida, dándose sólo dos casos: el sistema de la UNED<sup>5</sup> [13] la utiliza para detectar implicación entre unidades léxicas; para ello realiza una búsqueda aproximada de multipalabras de *WordNet* en los textos, aplicando la distancia de Levenshtein. Esto permite establecer relaciones semánticas (sinonimia, hiponimia, etcétera) no sólo entre palabras sino también con multipalabras. El otro caso es el de uno de los módulos en cascada del sistema de la Universidad Ca’ Foscari y el ITC-irst [8].

---

<sup>4</sup> ITC-irst, Centro per la Ricerca Scientifica e Tecnologica, Italia

<sup>5</sup> Universidad Nacional de Educación a Distancia, España.

**Reconocimiento de entidades, expresiones numéricas y expresiones temporales.** Son técnicas todavía poco utilizadas. En el caso del reconocimiento de entidades se encuentran sólo dos ejemplos: la Universidad de Stanford [21] y la de Illinois at Urbana-Champaign [7]. El sistema de Stanford detecta entidades nombradas y resuelve correferencias, con el fin de encontrar dependencias entre nodos de los grafos con los que representa los textos. En el caso de la Universidad de Illinois, las entidades nombradas se utilizan como atributos de los nodos de los grafos que representan los textos. En cuanto a la detección de expresiones numéricas y temporales se tienen otros dos ejemplos: La Universidad de Stanford [21] realiza un tratamiento de expresiones numéricas, capaz de determinar inferencias del tipo “*2113 es más que 2000*”. La Universidad Ca’ Foscari y el ITC-irst [8] efectúan una detección de expresiones temporales, para realizar comprobaciones de coherencia.

**Análisis sintáctico.** El **análisis de dependencias** es una de las técnicas más utilizadas, situación propiciada probablemente en el caso del inglés por la disponibilidad pública de analizadores de dependencias de gran eficiencia temporal y alta cobertura, como el desarrollado por Dekang Lin [17]. Generalmente se obtiene el árbol asociado como representación del texto analizado, aunque también como auxiliar para llegar a una representación en forma lógica. Como ejemplos del primer tipo de uso, se tienen el sistema de la UNED [13] y el del grupo de la Universidad de Trento y el ITC-irst [16]. El primero evalúa la existencia de implicación entre texto e hipótesis mediante el solapamiento entre los árboles de dependencias de ambos fragmentos de texto. El segundo evalúa la existencia de implicación entre texto e hipótesis mediante la distancia de edición entre los árboles de dependencias de ambos fragmentos de texto. Como ejemplo del otro tipo de uso, el MITRE [3] implementa en cascada subsistemas de análisis lingüístico, incluyendo una etapa de análisis de dependencias; antes del analizador de dependencias hay un analizador sintáctico de constituyentes y, después, un generador de predicados lógicos.

El **análisis de constituyentes**, por contra, es una técnica poco empleada. La Universidad “Tor Vergata”, asociada a la de Milano-Bicocca, [20] utiliza los constituyentes para extender grafos de dependencias. La Universidad Ca’ Foscari, junto al ITC-irst [8] realizan análisis de constituyentes como parte de un análisis sintáctico híbrido.

**Análisis semántico.** El etiquetado de **roles semánticos** lo usaron las universidades de Illinois at Urbana-Champaign [7], la de Stanford [21] y la Ca’ Foscari en asociación con el ITC-irst [8]. El sistema de Illinois at Urbana-Champaign busca coincidencias entre los conjuntos de atributos y la estructura de los argumentos tanto en el nivel de los roles semánticos como en el nivel del análisis sintáctico. En el caso de Stanford, este etiquetado permite añadir relaciones entre palabras no identificadas mediante el análisis sintáctico y, además, para clasificar las frases como temporales, locativas, etcétera. En todos los casos las etiquetas se aplicaban a los nodos de los grafos con los que representaban los

fragmentos de texto. La Ca' Foscari y el ITC-irst utilizan los roles semánticos en una medida de similitud entre el texto y la hipótesis, mediante el conteo de etiquetas similares entre las del texto y las de la hipótesis.

### 3.2. Otras Técnicas

Además de las técnicas anteriormente reseñadas, la mayor parte de los sistemas implementan una o varias de las que a continuación se citan:

**Uso de tesauros, grandes corpora y *WordNet*.** Una parte significativa de los sistemas obtiene conocimiento a partir de tesauros y *WordNet*. Las consultas a *WordNet* han sido realizadas buscando bien la obtención de relaciones entre unidades léxicas a partir de relaciones de *WordNet* – como es el caso del sistema de la UNED, que busca relaciones de sinonimia, hiperonimia e implicación de *WordNet* para establecer implicaciones entre unidades léxicas del texto y de la hipótesis – bien la obtención de relaciones a partir de cadenas léxicas, como el sistema de la Universidad de Concordia [2]. Los tesauros han sido utilizados para extraer conocimiento de algún campo concreto, como el conocimiento geográfico que obtienen las universidades de Edimburgh y Leeds [4] a partir del “CIA factbook”. Grandes corpora como la *web* o el *Gigaword newswire corpus* han sido utilizados para extraer propiedades léxicas [3] o estadísticas de coocurrencia [10].

**Detección de paráfrasis.** El uso de paráfrasis se centra en la obtención de reglas de reescritura con las que poder mejorar el comportamiento al intentar determinar si dos expresiones son equivalentes. Tal es el caso del grupo de la Universidad de Illinois at Urbana-Champaign, que a partir de un corpus de reglas de paráfraseo desarrollado por Lin y Pantel (2001) obtuvo un conjunto de reglas de reescritura para que su sistema generase variantes de los textos [7].

**Demostradores automáticos.** Los sistemas que, tras realizar un análisis lingüístico, representan en forma lógica los textos entre los que se tiene que resolver la posible inferencia, utilizan demostradores automáticos para llevarlo a cabo, como es el caso del de la Universidad Macquaire [1].

**Aprendizaje automático.** Algunos sistemas han hecho uso de este tipo de algoritmos como, por ejemplo, el de las universidades “Tor Vergata” y Milano-Bicocca [20], que aplicaban un SVM para evaluar los parámetros de una medida de evaluación.

**Definición de un marco probabilístico.** Como único ejemplo se tiene el propuesto por la Universidad Bar Ilan, que define un marco probabilístico para modelizar la noción de implicación textual [10]; así mismo, recurre a una representación de bolsa de palabras para describir un modelo de implicación léxica

a partir de estadísticas de coocurrencia en la *web*. Se dice que un texto implica probabilísticamente a una hipótesis si el texto incrementa la probabilidad de que el valor de verdad asignado a la hipótesis sea *cierto*. Para tratar la implicación léxica, se establece un modelo probabilístico según el cual se espera que cada palabra de la hipótesis sea implicada por alguna palabra del texto; ésto se puede ver, alternativamente, como la inducción de una alineación entre términos de la hipótesis y el texto, de manera similar a como se realiza en traducción automática estadística [5]. Así pues, la implicación probabilística entre texto e hipótesis la calculan en función de la implicación léxica referida. Las probabilidades de implicación léxica las estiman de manera empírica mediante un proceso no supervisado fundamentado en coocurrencias en la *web*.

**Traducción automática.** El MITRE ha desarrollado un sistema inspirado en los modelos de traducción automática estadística [3], siguiendo el siguiente proceso:

1. Entrenamiento del sistema de traducción automática mediante el *Gigaword newswire corpus* [11], buscando implicaciones entre titulares y subtítulos.
2. Estimación manual de la fiabilidad del entrenamiento anterior.
3. Refinamiento del corpus obtenido anteriormente mediante el entrenamiento del clasificador de documentos SVMlight [15].
4. Inducción de modelos de alineamiento en el subconjunto seleccionado del *Gigaword newswire corpus*, mediante las herramientas GIZA++ [19].
5. Utilización de un clasificador de los vecinos más próximos a distancia  $k$  con cada uno de los pares <texto, hipótesis> del corpus de prueba, para elegir el valor dominante de verdad de entre los vecinos más próximos a distancia 5 en el corpus de desarrollo.

## 4. Conclusiones

A grandes rasgos, se pueden diferenciar algunas tendencias en el desarrollo sistemas de RTE:

- Los que tratan los textos como bolsas de palabras y la extracción de lemas es el análisis lingüístico más profundo que realizan.
- Los que se fundamentan en una representación sintáctica de los textos, incluyendo algunos procesamientos morfológico-léxicos de manera accesoria para incrementar la capacidad del sistema; en este caso existe una preponderancia en la determinación de la implicación a partir del solapamiento de árboles de dependencias.
- Los que realizan un tratamiento lingüístico profundo, lo que implica hacer un análisis clásico por etapas, recorriendo un amplio rango de niveles de análisis: morfológico-léxico, sintáctico y semántico.

Son escasos los ejemplos de tratamientos exclusivamente estadísticos o de sistemas que realicen un análisis lingüístico profundo.



Los resultados obtenidos en el PASCAL RTE *Challenge* no son indicativos sobre la idoneidad de las técnicas empleadas, ya que todos los participantes – en un marco común – obtuvieron valores de las medidas de evaluación muy similares, oscilando entre el 49,5% y el 58,6% de precisión [6] en el reconocimiento de implicación entre los pares de textos e hipótesis.

Así pues, el estado actual no permite decidir si es más efectivo el tratamiento lingüístico superficial o profundo, o bien un tratamiento exclusivamente estadístico. La experiencia del *Challenge* sugiere una revisión profunda del funcionamiento de los sistemas con respecto a los corpora utilizados, para detectar qué tipos de inferencias no se han llevado a cabo y así proponer nuevas vías de mejora.

## Agradecimientos

Este trabajo ha sido parcialmente subvencionado por el Ministerio de Ciencia y Tecnología, con cargo a los presupuestos del Proyecto TIC-2003-07158-C04-02: R2D2-SyEMBRA.

## Referencias

1. E. Akhmatova. Textual Entailment Resolution via Atomic Propositions. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 61–64, April 2005.
2. A. Andreevskaia, Z. Li, and S. Bergler. Can Shallow Predicate Argument Structures Determine Entailment? In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 45–48, April 2005.
3. S. Bayer, J. Burger, L. Ferro, J. Henderson, and A. Yeh. MITRE's Submissions to EU PASCAL RTE Challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 41–44, April 2005.
4. J. Bos and K. Markert. Combining Shallow and Deep NLP Methods for Recognizing Textual Entailment. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 65–68, April 2005.
5. P. F. Brown, S. A. Della Pietra, V. J. Della Pietra, and R. L. Mercer. The Mathematics of Statistical Machine Translation. In *Computational Linguistics 19(2)*, 1993.
6. I. Dagan, O. Glickman, and B. Magnini. The PASCAL Recognising Textual Entailment Challenge. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 1–8, April 2005.
7. R. de Salvo Braz, R. Girju, V. Punyakanok, D. Roth, and M. Sammons. Textual Entailment Recognition Based on Dependency Analysis and WordNet. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 29–32, April 2005.
8. R. Delmonte, S. Tonelli, M. A. Picollino Boniforti, A. Brsitot, and E. Pianta. VENSES – a Linguistically-Based System for Semantic Evaluation. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 49–52, April 2005.

9. A. Fowler, B. Hauser, D. Hodges, I. Niles, A. Novischi, and J. Stephan. Applying COGEX to Recognize Textual Entailment. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 69–72, April 2005.
10. O. Glickman, I. Dagan, and M. Koppel. Web Based Textual Entailment. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 33–36, April 2005.
11. D. Graff. English Gigaword. <http://www ldc.upenn.edu/Catalog/>, 2003.
12. Z. Harris. Distributional Structure. In J. J. Katz, editor, *The Philosophy of Linguistics*, pages 26–37, 1985.
13. J. Herrera, A. Peñas, and F. Verdejo. Textual Entailment Recognition Based on Dependency Analysis and WordNet. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 21–24, April 2005.
14. V. Jijkoun and M. de Rijke. Recognizing Textual Entailment Using Lexical Similarity. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 73–76, April 2005.
15. T. Joachims. *Learning to Classify Text Using Support Vector Machines*. Kluwer Academic Publishers, Massachusetts, 2002.
16. M. Kouylekov and B. Magnini. Recognizing Textual Entailment with Tree Edit Distance Algorithms. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 17–20, April 2005.
17. D. Lin and P. Pantel. DIRT - Discovery of Inference Rules from Text. In *Proceedings of ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 323–328, 2001.
18. C. Monz and M. de Rijke. Light-Weight Entailment Checking for Computational Semantics. In P. Blackburn and M. Kohlhase, editors, *Inference in Computational Semantics (ICoS-3)*, pages 59–72, 2001.
19. J. Och and H. Ney. A Systematic Comparison of Various Statistical Alignment Models. In *Computational Linguistics 29(1)*, 2003.
20. M. T. Paziienza, M. Pennacchiotti, and F. M. Zanzotto. Textual Entailment as Syntactic Graph Distance: a Rule Based and a SVM Based Approach. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 25–28, April 2005.
21. R. Raina, A. Haghighi, C. Cox, J. Finkel, J. Michels, K. Toutanova, B. MacCartney, M. C. de Marneffe, C. D. Manning, and A. Y. Ng. Robust Textual Inference using Diverse Knowledge Sources. In *Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment, Southampton, UK*, pages 57–60, April 2005.
22. I. Szpektor, H. Tanev, I. Dagan, and B. Coppola. Scaling Web-Based Acquisition of Entailment Relations. In *Proceedings of Empirical Methods in Natural Language Processing (EMNLP-04)*, 2004.
23. H. Tanev, M. Kouylekov, and B. Magnini. Combining Linguistic Processing and Web Mining for Question Answering. In *Proceedings of the 2004 Edition of the Text Retrieval Conference*, 2004.