In Part III of this thesis I present: 1) a methodology for the creation and annotation of corpora containing multiple textual meaning relations; 2) the first corpus annotated independently with Paraphrasing, Textual Entailment, Contradiction, Textual Specificity, and Semantic Similarity; 3) a statistical corpus-based analysis of the interactions, correlations, and overlap between the different meaning relations; 4) SHARel - a shared typology of textual meaning relations; 5) a corpus of paraphrasing, textual entailment, and contradiction annotated with SHARel. Part III of the thesis gives a new perspective on the research of textual meaning relations. I show that a joint study of multiple meaning relations is both possible and beneficial for processing and analyzing each individual relation. I provide the first empirical data on the interactions between paraphrasing, textual entailment, contradiction, and semantic similarity. The research in Part III resulted in the publication of two articles.

This thesis has advanced our understanding of important issues associated with the empirical analysis, corpus annotation, and computational treatment of textual meaning relations. I have addressed existing gaps in the research field, posed new research questions, and explored novel research directions. The findings and resources presented in this dissertation have been released to the community to facilitate further research and knowledge transfer.

Resumen

En esta tesis se exploran los aspectos lingüísticos y computacionales de las relaciones semánticas que puede haber entre dos o más expresiones lingüísticas complejas (sintagmas, cláusulas, oraciones, párrafos). En particular, se centra en la paráfrasis, la implicación, la contradicción y la similitud semántica. La tesis se compone de siete artículos y se estructura en tres partes.

En la *Parte I: "Similitud de palabras y sintagmas"*, realizo un estudio sobre la Hipótesis distribucional (HD). La HD es relevante en muchos de los trabajos actuales sobre el procesamiento del significado y de las relaciones de significado en el área de la Lingüística Computacional (LC) y el Procesamiento del Lenguaje Natural (PLN). En esta parte se exploran diferentes métodos para la cuantificación de la similitud semántica de palabras y de sintagmas. He calculado la importancia del tamaño del corpus y el papel que juega el preprocesado lingüístico. También muestro que la similitud semántica léxica puede interactuar con reglas de composición sintáctica lo que da como resultado patrones productivos al nivel de sintagma. La investigación de esta parte de mi tesis ha dado lugar a la publicación de dos artículos.

En la *Parte II: "Tipología de paráfrasis e identificación de paráfrasis"* me centro en la relación semántica de paráfrasis y en la tarea empírica de la identificación automática de paráfrasis (IP). La paráfrasis es una de las relaciones de significado más estudiadas, tanto a nivel teórico como aplicado. La IP es una de las tareas más populares en LC y en el PLN. En la Parte II de esta tesis presento: 1) EPT, una nueva tipología de fenómenos lingüísticos y de fenómenos basados en el razonamiento implicados en la paráfrasis: 2) WARP-Text, una nueva interfaz web para la anotación de diferentes tipos de paráfrasis; 3) ETPC: hasta el momento, el corpus de mayor tamaño anotado con tipos de paráfrasis; y 4) un entorno de evaluación cualitativa de sistemas automáticos de IP; Los resultados de esta segunda parte proporcionan un conocimiento más a fondo sobre la naturaleza de la relación de paráfrasis y mejoran la evaluación, interpretación y análisis de errores referentes a la tarea de IP. La investigación de esta segunda parte ha dado lugar a tres publicaciones.

En la *Parte III: "Paráfrasis, Implicación textual y Similitud semántica"*, pre-

sento una nueva línea en la investigación sobre las relaciones de significado. Llevo a cabo una investigación conjunta sobre paráfrasis, implicación textual, contradicción y similitud semántica. Tradicionalmente, estas relaciones se han estudiado separadamente y la transferencia de conocimiento entre ellas ha sido muy limitado. En esta tercera parte de la tesis presento: 1) una metodología para la creación y anotación de corpus que contienen diversas relaciones de significado; 2) el primer corpus anotado independientemente con Paráfrasis, Implicación textual, Contradicción, Especificidad y Similitud semántica; 3) un análisis estadístico de las interacciones, correlaciones y coincidencias entre las diferentes relaciones de significado; 4) SHARel, una tipología compartida para las relaciones semánticas textuales; 5) un corpus de paráfrasis , implicación textual y contradicción anotado con SHARel. Esta tercera parte de la tesis da una nueva perspectiva sobre la investigación en las relaciones de significado a nivel textual. Pongo de manifiesto que es posible el estudio conjunto de diversas relaciones de significado y también que repercute positivamente para cada una de las relaciones en particular. Proporciono por primera vez un conjunto de datos empíricos sobre la integración de paráfrasis, implicación textual, contradicción y similitud semántica. La investigación de esta tercera parte ha dado lugar a dos artículos.

Esta tesis ha permitido avanzar en la comprensión de aspectos importantes relacionados con el análisis empírico, la anotación de corpus, y el tratamiento computacional de las relaciones de significado a nivel textual. He tratado diversas áreas de conocimiento poco atendidas hasta ahora, he planteado nuevas preguntas para la investigación posterior y he explorado en nuevas directrices. Los resultados y recursos presentados en esta tesis son de libre disposición para el colectivo que investiga en LC y PLN con el fin de facilitar la investigación futura y la transferencia de conocimiento.