Trabajo Fin de Máster: Word Embeddings para la anotación automática de gestos aplicada a Nao

Trabajo Fin de Máster

Mario Almagro Cádiz

Trabajo de investigación para el

Máster en I.A. avanzada: Fundamentos, métodos y aplicaciones

Universidad Nacional de Educación a Distancia

Dirigido por el

Prof. Dr. D. Víctor Fresno Fernández

y el

Prof. Dr. D. Félix de la Paz López

Febrero 2017

Resumen

En los últimos años, la integración progresiva de la robótica de servicios en los hogares ha despertado el interés de la comunidad científica por las interacciones de tipo robot-humano. Estudios recientes han arrojado nuevas líneas de investigación en torno a la expresión emocional y la gesticulación de los robots humanoides.

Una herramienta que facilite la integración de la comunicación gestual y la comunicación verbal supondría un gran avance para el diseño de comportamientos naturales. Con afán de mejorar la interfaz social en el ámbito robótico, este trabajo plantea una metodología para la anotación gestual ante textos abiertos, englobando tanto la tokenización de las palabras, como el etiquetado gramatical o *POS tagging* y el uso de *word embeddings* con el fin de aplicar criterios semánticos a la asignación gestual. Esta metodología ha sido trasladada a un algoritmo encapsulado en una aplicación web, accesible desde el dominio *http://trasgu.lsi.uned.es:8080/MotionTags*.

Para concluir este trabajo, se ha empleado dicho algoritmo en un cuento, disponible en el enlace *https://www.youtube.com/embed/lzViIUvNyJM?rel=0&hl=es&cc_lang_pref=es&cc_load_policy=1*, con el objetivo de realizar una breve demostración mediante un robot tipo *Nao*.

Abstract

In recent years, the progressive integration of service robotics in households has attracted the scientific community's interest in robot-human interactions. Recent studies have opened up new research lines into emotional expression and gesture for humanoid robots.

A tool that facilitates the integration of gestual and verbal communication would be a major step forward in design for natural behaviors. In an effort to improve the social interface in robotics, this work presents a methodology for gestural annotation in open texts, grouping together word segmentation, POS tagging and the use of *word embeddings* in order to apply semantic criteria to gestural assignment. This methodology has been developed into an algorithm encapsulated in a web application, accessible from the *http://trasgu.lsi.uned.es:8080/MotionTags* domain.

To conclude this work, this algorithm has been applied to a children's story, available in the link *https://www.youtube.com/embed/lzViIUuNyJM?rel=0&hl=es&cc_lang_pref=es&cc_load_policy=1*, with the aim of carrying out a brief demo with a *Nao* robot.

Índice general

Índice de Figuras

Índice de Tablas

Capítulo 1

Introducción

1.1. Motivación

Con su aparición, los nuevos avances tecnológicos siempre se materializan inicialmente como productos o servicios prescindibles en la sociedad. Conforme evolucionan y se propagan, su impacto en la mejora de la calidad de vida se hace cada vez más notable. Avances como la electricidad, el teléfono e Internet se han convertido en factores fundamentales que sustentan la sociedad y sus hábitos. La robótica está ahora mismo en plena expansión, extendiéndose por todos los ámbitos sociales, y en el futuro se espera que forme parte de aquellos recursos indispensables para todo ser humano.

El objetivo de la robótica es la generación de herramientas que resuelvan las necesidades sociales. Los primeros robots surgieron en el ámbito industrial para realizar tareas que suponían un riesgo para los trabajadores. Su impacto en el aumento de la productividad provocó su extensión a tareas cada vez más complejas y flexibles, dando lugar a sistemas adaptativos e inteligentes. Mientras que la industria integra diseños robóticos desde hace años, los estrictos requerimientos de seguridad han frenado su avance en el sector servicios.

Las sociedades occidentales han padecido cómo el sector industrial, agente fundamental de la actividad económica durante el siglo XX, ha ido perdiendo peso desde los inicios del siglo XXI en favor de una economía más centrada en las actividades del sector terciario. Así, el desarrollo de la robótica que durante el siglo anterior se encontraba enfocado en la industria, actualmente dirige sus esfuerzos al avance de la robótica de servicios. Ámbi-

tos como la limpieza, el repostaje, el arado, la construcción o la enseñanza constituyen la nueva área de investigación, siendo el ámbito médico y de rehabilitación los más demandados.

La robótica de servicios modifica el paradigma establecido hasta el momento a través de diseños que deben coexistir con seres humanos y ser capaces de enfrentarse a entornos más impredecibles (Schraft, 1994). Con el desarrollo de tareas en convivencia con las personas, la adquisición de nuevos interfaces para la interacción social cobra mayor importancia; los robots deben poder comunicarse con las personas con el fin de transmitir información de la tarea a realizar y adquirir nuevas directrices. Con ello, la interacción humana es vista como un interfaz universal que permitiría el control de los robots por parte de cualquier usuario, surgiendo así el campo de la interacción humano-robot o *HRI*.

Para mejorar el campo de la *HRI*, la robótica tiene que ser capaz de provocar emociones que generen una mayor afinidad. Para este propósito el diseño es tan decisivo como el comportamiento (Minato et al., 2004), ya que aparentemente cuanto más se parece un robot a un ser humano mayor empatía despierta en las personas (Riek et al., 2009a; Riek et al., 2009b); de ahí la importancia de la robótica humanoide. Masahiro Mori postula que afinidad y semejanza no constituyen una relación puramente incremental, sino que alberga un mínimo local: un diseño demasiado semejante pero visiblemente diferente provocará cierto rechazo atribuido a un instinto de repulsa hacia otras especies (Mori, MacDorman, y Kageki, 2012). Pese a todo, los diseños humanoides actuales aún distan profundamente de la apariencia humana y necesitan seguir perfeccionándose.

Un robot que cubra las demandas de las personas debe captar las emociones humanas mediante distintos medios, como la vista o el sonido, además de ser capaz de expresar sus propias emociones para transmitir información y facilitar su entendimiento. Aunque tradicionalmente los proyectos robóticos se han centrado en dar un soporte físico, las nuevas corrientes de investigación se están enfocando hacia la robótica social asistida o *SAR*, en la que predomina el intercambio emocional. Bajo este contexto, una emoción es entendida como el resultado de la evaluación entre la información recibida del entorno y la información interna del dispositivo. Con ello, la gesticulación adquiere una gran importancia en la robótica como complemento lingüístico, necesaria tanto para incrementar la empatía como para expresar

efusividad o intencionalidad. Los gestos tienen la función de enfatizar o ampliar la información verbal, así como añadir nueva información, responder a un interrogante o animar al interlocutor en el discurso (Meena, Jokinen, y Wilcock, 2012). Pueden ser clasificados en base a su forma o función, como los gestos discursivos y deícticos. El estudio de Nehaniv (Nehaniv et al., 2005) refleja una taxonomía basada en 5 clases diferenciadas por contexto e intencionalidad:

- **Gestos de manipulación.** Movimientos que suelen estar asociados a la modificación del entorno, careciendo de intencionalidad comunicativa. Ejemplos de ello son el movimiento de los brazos al caminar, rascarse o ajustarse las gafas.
- **Gestos de interacción.** Gestos de manipulación sobre un entorno animado que requieren cooperación; por ejemplo, extender un objeto a alguien o dar la mano.
- **Gestos de señalización o deícticos.** Indicaciones de objetos o localizaciones de interés; como por ejemplo, señalar una estrella.
- **Gestos expresivos del discurso.** Gestos empleados para acompañar al discurso. Generalmente se usan para indicar efusividad, como el movimiento de manos.
- **Gestos simbólicos.** Gestos para comunicar el contenido o significado semántico de un mensaje. Algunos ejemplos consisten en saludar, indicar cómo es una explosión o imitar una pelota.

Aunque existen distintos estudios que definen tipologías gestuales (Kendon, 1986; Nespoulous y Lecours, 1986; McNeill y Levy, 1980; Quek, 1994), es posible sintetizar unos rasgos comunes. Las tres primeras clases de gestos pueden ser agrupadas bajo el tipo de *gestos reactivos*, apartados del discurso y muy extendidos en la robótica, así como las otras dos clases pertenecen a los *gestos co-verbales*.

Generalmente, la líneas de investigación en el campo de la robótica han tenido una mayor predilección por la detección e integración de gestos de interacción, manipulación y deícticos, centrándose especialmente en las posiciones y trayectorias de las manos. Existe también algún proyecto que ha desarrollado expresiones emotivas, como el robot *WE-4RII* (Itoh et al., 2004).

Aunque el acompañamiento gestual del discurso ha sido un campo bastante inexplorado en la robótica (Salem et al., 2009), se han conseguido algunos avances en el entorno digital de los avatares. La sincronización de *gestos co-verbales* ha sido abordada por trabajos pertenecientes a dos vertientes: aproximaciones basadas en reglas y modelos estadísticos. Con ello, han surgido diferentes sistemas que se basan en el análisis de fragmentos de vídeos discursivos para desarrollar métricas con las que establecer reglas. También es frecuente el uso de sistemas que requieren textos anotados con gestos para generar modelos *HMMs* y establecer distintos patrones de comportamiento; como por ejemplo, el trabajo de Chung-Cheng Chiu et al. (Chiu, Morency, y Marsella, 2015).

Escasas son las investigaciones al respecto que emplean un procesamiento del lenguaje natural, y las pocas existentes únicamente lo han utilizado con el fin de determinar qué tipo de gesto es más conveniente activar, utilizando para ello etiquetas y modelos gramaticales. Aunque la asociación de gestos a palabras es común, las técnicas que se han venido usando se basan fundamentalmente en la aplicación de reglas simples estilo *si encuentro* $word_1$ *ó* $word_2$ *ó* $word_3$*ó* $\ldots \rightarrow$ *se activa G1.*

Con el afán de suplir una técnica de asociación más flexible, este trabajo plantea una aproximación basada en similitudes semánticas; mientras que tradicionalmente se ha limitado el lanzamiento de estos gestos a distintas palabras clave, con la nueva aproximación un gesto es activado con cualquier palabra cuyo significado sea relativamente cercano semánticamente. La pretensión de este trabajo de investigación es, por tanto, mejorar la naturalidad de las interfaces robóticas dentro del contexto descrito con la integración de gestos simbólicos; a diferencia de otros estudios, en éste se emplearán métricas de similitud semántica entre los gestos y las palabras para determinar qué expresión gestual es la más acorde al significado verbal.

1.2. Hipótesis y objetivos

Las investigaciones en torno a la interacción robot-humano dentro del campo *HRI* se han centrado generalmente en los gestos no verbales, del tipo *deícticos*. Es por ello por lo que este trabajo se centra en los *gestos co-verbales* de tipo *simbólicos*, no cubiertos en profundidad por la literatura. La integración de los *gestos co-verbales* requiere un análisis del lenguaje

natural, ya que deben acompañar al discurso. Las primeras aproximaciones del desarrollo de los gestos simbólicos se basan en la detección exacta de palabras clave; sin embargo, esta técnica supone una metodología rígida, ineficaz y limitada, incapaz de reaccionar ante variaciones lingüísticas. Un análisis semántico sería capaz de paliar esa rigidez.

Tradicionalmente, el procesamiento semántico se ha realizado a través de bases de datos léxicas como *WordNet*, conformando un árbol de relaciones de conceptos. Sin embargo, en los últimos años se ha comenzado a impulsar el uso de modelos estadísticos semánticos generados por el aprendizaje automático de redes neuronales, denominados *word embeddings*. Éstos configuran el contexto de significados de las palabras en vectores numéricos de n dimensiones. Partiendo de la naturaleza de los *word embeddings* puede ser interesante mantener una separación gramatical en los términos para evitar comparaciones entre contextos diferentes.

Por último, uno de los robots que más se está extendiendo en el ámbito educativo es el modelo *Nao* (Mubin et al., 2013). Su adquisición se ha popularizado por su diseño avanzado y atractivo en contraposición con su bajo coste. Por ello, están surgiendo numerosas investigaciones en el ámbito robótico que emplean estos modelos durante la experimentación; entre ellas, una gran parte en el contexto de interacciones robot-humano. Este trabajo empleará el mismo modelo para visualizar los resultados. Consecuencia de esa popularidad es la disponibilidad de una amplia librería de gestos pre-programados, proporcionada de manera oficial en el nuevo *framework Naoqi 2.x*. Ésta se usará para conformar el listado de gestos empleados en la demostración visual.

A partir de todas las premisas anteriores se pueden formular las siguientes hipótesis:

- Los gestos suelen estar asociados a determinadas palabras; por ello, una aproximación adecuada puede ser la activación de un gesto por una palabra y otras relacionadas semánticamente.
- Se asume que los nombres, verbos, adjetivos y adverbios son los elementos más relevantes en la semántica de una oración, por lo que es esperable que tengan un papel notable a la hora de activar un gesto desde un texto escrito en lenguaje natural.
- La consideración de diferentes categorías gramaticales en los datos

puede mejorar la tasa de aciertos durante la experimentación.

- Puede hacerse necesario el uso de un umbral de distancia semántica mínimo para asociar los gestos realmente relacionados a una palabra.
- La taxonomía de *WordNet* y los modelos de *word embeddings* suponen aproximaciones apropiadas para el cálculo de similitudes semánticas.
- El robot *Nao* supone uno de los mejores medios para implementar interfaces sociales en el ámbito universitario, tanto por su coste de adquisición como por su *framework*.
- La librería de animaciones *Animations* proporcionada por *Naoqi* tiene la cobertura apropiada para llevar a cabo una demostración visual.

Dada la ausencia de estudios sobre la integración de los *gestos co-verbales* asociados al significado, el **objetivo** de este trabajo es la mejora de la naturalidad de las interfaces sociales en el ámbito robótico a través del diseño de una metodología para la anotación gestual automática de textos libres.

1.3. Propuesta

En base al creciente interés en el campo de la interacción humano-ordenador o *HCI*, así como en mejorar la naturalidad en la comunicación digital, esta investigación se ha planteado con el fin de integrar los gestos de tipo simbólico bajo el contexto de la interacción social en la robótica, y más concretamente, en el ámbito educacional. Actualmente las herramientas más consolidadas en la integración de los gestos simbólicos se asientan en la idea de asignar palabras relacionadas a cada uno de ellos, de forma que son activados con la detección de cualquiera de esas palabras en las oraciones. Con la intención de mejorar esta aproximación básica, la propuesta del proyecto consiste en el diseño de una nueva metodología de anotación gestual mediante el estudio de la aplicación de métricas de similitud semántica a texto libre. Este estudio se desarrolla con un análisis comparativo de funciones de similitud semántica basadas en *WordNet* y *word embeddings* sobre un conjunto de gestos. Finalmente, se aplicará dicha metodología a un algoritmo cuyos resultados se verificarán sobre un modelo robótico *Nao*

El diseño de la metodología arranca con la asignación de palabras significativas a los gestos, conformando un espacio de representación semántica.

Estos espacios de representación se emplearán para activar los gestos con la detección de palabras con significados relacionados en el texto correspondiente. El estudio de los efectos de las métricas de similitud semántica en el contexto gestual se llevará a cabo desde el punto de vista de las taxonomías léxicas con *WordNet*, y los modelos de aprendizaje automático basados en *word embeddings*. Inicialmente, se considerará la generación de un conjunto de datos de prueba basado en los términos más frecuentes en inglés, asociándolos a etiquetas en representación de los gestos. Se emplearán páginas web que ofrecen términos relacionados para conformar las bolsas de palabras de cada gesto o término. Posteriormente se propone desarrollar dos algoritmos en *C++*. El primero implementando 10 métricas de similitud basadas en *WordNet*, englobando los principios de distancia en la taxonomía, cantidad de información o *IC* y comparación entre glosas. Respecto al otro algoritmo, usará los *word embeddings* para calcular la distancia euclídea y la similitud coseno entre términos y palabras.

Teniendo en cuenta la naturaleza de *WordNet* y los *word embeddings*, se analizarán los resultados tras aplicar una restricción para la comparación semántica entre categorías gramaticales iguales. Por otro lado, el estudio constará de dos sistemas alternativos: un primer sistema en el que se establecerá un umbral mínimo de similitud con el que cada palabra puede ser asociada a cualquier gesto cuyo término lo supere, y un segundo sistema en el que no se establece ningún umbral y cada palabra es simplemente asociada a aquel gestos cuyo término es más cercano. Por último, se plantea la posibilidad de modificar la distribución de los datos en las distintas categorías gramaticales para analizar su efecto.

Finalmente, se configurará un algoritmo con los resultados más destacables a niveles de precisión y cobertura, así como respecto al coste computacional. Si es necesaria la separación gramatical, se propone el uso de las librerías de *FreeLing* para el análisis *POS tagging*. El algoritmo resultante se encapsulará en un servicio web público accesible en el dominio `http://trasgu.lsi.uned.es:8080/MotionTags`. Para comprobar la viabilidad del mismo, se aplicará a un cuento en español para su integración en *Nao* con los gestos de la librería *Animations*.

1.4. Estructura del documento

Capítulo 1. Introducción. Este capítulo detalla los principales motivos que han llevado a la realización del trabajo, así como las bases sobre las que se asienta. Por último, se define la propuesta a implementar.

Capítulo 2. Estado del arte. Este capítulo describe en mayor detalle la robótica social, presentando su origen y su historia hasta el presente. Se muestran las técnicas actuales más utilizadas para resolver la interacción robot-humano, y en especial la robótica gestual, así como sus debilidades.

Capítulo 3. Diseño experimental. Este capítulo presenta los detalles de la experimentación llevada a cabo, y plantea las técnicas empleadas para el diseño de la metodología definitiva.

Capítulo 4. Experimentación y análisis de resultados. Este capítulo analiza y evalúa los resultados de los métodos propuestos para la tarea con el fin de terminar presentando la configuración que mejor se adecue al problema de la anotación gestual.

Capítulo 5. Aplicación práctica de la metodología. Este capítulo presenta el diseño de una demostración práctica para la visualización de los resultados arrojados por la metodología diseñada. Para ello, se aplicará el algoritmo desarrollado a un cuento de *Teo*, que posteriormente se integrará en un robot *Nao*.

Capítulo 6. Conclusiones y trabajo futuro. Este capítulo recopila las diferentes conclusiones extraídas del trabajo realizado, y propone algunas líneas de trabajo futuro.

Capítulo 2

Estado del arte

Este capítulo desarrolla una revisión de la robótica de servicios, concretamente de la reciente expansión de los robots humanoides. Por consiguiente, se aborda la creciente demanda de interfaces sociales para la interacción robot-humano, así como la generación de gestos para la mejora comunicativa. Por último, se hace una breve introducción a la similitud semántica, a través de las bases de datos léxicas y los modelos de datos formados por aprendizaje automático.

2.1. Robótica

2.1.1. El origen de la robótica

Durante los años cincuenta aparecieron los primeros prototipos de manipuladores industriales bajo la demanda de herramientas para tareas de alto riesgo. El primer manipulador programable fue patentado en 1948 por George Devol (Devol, 1948); ese mismo año, R.C. Goertz implementó el primer tele-manipulador con el fin de gestionar productos radiactivos (and others, 1953). En 1958 apareció el primer robot bajo el nombre de *Unimate*, de la mano de Devol y Joseph F. Engelberger; éste consistía en una computadora en combinación con un manipulador (Ballard et al., 2012). Sin embargo, no fue hasta 1962 cuando se instaló el primer robot (de modelo *Unimate*) en una cadena industrial, propiedad de *General Motors*.

A lo largo de la década de los sesenta, la industria submarina y aeroespacial comenzaron a interesarse por la robótica, incrementando la financiación y dándole un gran impulso a la investigación. Con ello, surgieron diferentes

robots como la serie soviética *Lunokhod* o el modelo estadounidense *Lunar Roving Vehicle* de exploración espacial (Morea, 1992). En los años siguientes, las líneas de investigación se centraron en el ámbito industrial, dirigiendo sus esfuerzos a mejorar la precisión a través de la calibración cinemática, optimizar los métodos de planificación del movimiento para ejecutar trayectorias o desarrollar técnicas de control más eficaces. No fue hasta los años noventa cuando los diseños comenzaron a propagarse hacia ámbitos cotidianos, dando lugar a los robots personales (Dario, Guglielmelli, y Laschi, 2001).

Finalmente, con una demanda cada vez más creciente en el sector servicios, las investigaciones pusieron el foco en los robots de servicio. De acuerdo a los informes de *Naciones Unidas*, hoy en día cerca del 24 % de la población de los países europeos tiene más de 60 años, y las estimaciones dictan que ese sector se incrementará hasta el 34 % para el año 2050 (evolución en la Figura 2.1). Las condiciones de vida actuales a partir de esas edades son alarmantemente insatisfactorias, con una proyección más negativa. La robótica de servicios pretende ser la solución definitiva a este problema.

En los últimos años, los robots de servicio se han integrado con éxito en hospitales (Ozkil et al., 2009), museos (Germak et al., 2015) y supermercados (como el robot *CompaRob* (Sales et al., 2016) o *TOOMAS* (Gross et al., 2009)), con el afán de ejecutar tareas de limpieza, distribución, entretenimiento o educativas. También se han desarrollado proyectos de ayuda a invidentes (Graf y Staab, 2009). A pesar de todo, la mayoría de los estudios se han centrado en proporcionar una ayuda mecánica, dejando de lado la parte cognitiva. Hay un vacío en la investigación de robots de interacción con los pacientes, atención al público o tutela de tareas, que ahora se está empezando a paliar.

2.1.2. Áreas de investigación en la robótica

Tradicionalmente las líneas de investigación en la robótica se han dividido en tres áreas principales, en función de las tareas desempeñadas: los robots de manipulación, los robots móviles y los robots biológicamente inspirados.

Los primeros diseños robóticos surgieron como brazos robóticos destinados a tareas industriales, como la pintura o soldadura de piezas en cadenas de montaje. En la década de los años noventa, los robots de manipulación se extendieron a la industria alimentaria y farmacéutica, en las que se re-

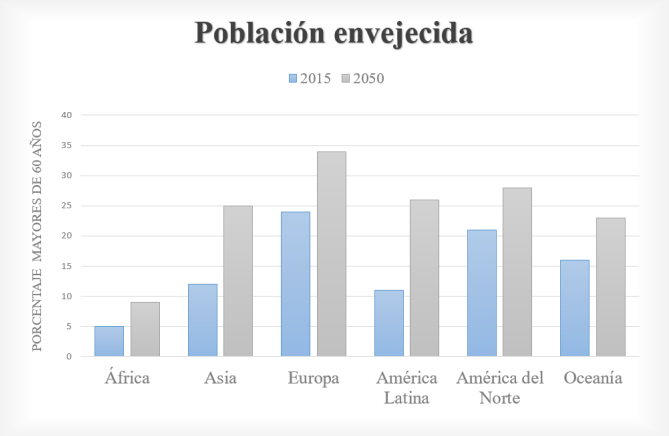


Figura 2.1: Proyección del porcentaje de la población mayor de 60 años, por región principal, 1994, 2014 y 2050. Informe de Naciones Unidas: World Population Prospects The 2015 Revision.

quería el manejo de distintos productos en tamaño, forma y consistencia. Con este pretexto, las investigaciones científicas enfocaron sus estudios en el control remoto y el aprendizaje automático tutelado con el fin de mejorar la flexibilidad en la manipulación de objetos. Recientemente los manipuladores han aterrizado en la robótica de servicios, introduciéndose con una gran demanda en el ámbito médico con la cirugía (Lanfranco et al., 2004) o la rehabilitación (Díaz, Gil, y Sánchez, 2011). Otras áreas como el repostaje también han tenido cabida (Scott et al., 2015).

A diferencia de los mecanismos manipuladores, los robots móviles se caracterizan por su sistema locomotor. Aparecieron por primera vez a finales de los años sesenta como herramientas de transporte. Aunque inicialmente se basaban en técnicas con trayectorias fijas en un entorno estático, la navegación por detección de líneas adquirió una mayor importancia a nivel industrial. En el ámbito científico, los proyectos abordaron problemas de navegación por la acumulación de errores en los sensores, la generación de mapas y la auto-localización. Actualmente se ha incrementado la complejidad en entornos dinámicos, como es el caso de los extendidos robots domésticos de limpieza. Desde su masiva introducción en los hogares, ponen de manifiesto la decepción generada por la ausencia de interfaces con inteligencia aparente (Forlizzi y DiSalvo, 2006). Los ámbitos actuales más notables son la aeronáutica con el desarrollo de drones (Springer, 2013), así como los robots submarinos (Chu et al., 2012).

En cuanto a los robots de carácter biológico, sus diseños se basan en los mecanismos de locomoción adaptativos que emplean diferentes especies, incluyendo el ser humano. Aunque es el área menos explotada dada su complejidad, se ha convertido en una de las líneas de investigación que genera mayor interés, con importantes avances en el diseño de las articulaciones robóticas. A pesar de todo, desde los años noventa han surgido escasas investigaciones sobre la movilidad de las articulaciones debido a los problemas de estabilidad dinámica; la versatilidad del terreno impide el control de la movilidad en entornos no controlados. Sin duda, los robots humanoides son los más destacables en base a su éxito en la interacción social. Aun así, los diseños biológicos se están extendiendo a otros campos como la agricultura o la silvicultura (Billingsley, Visala, y Dunn, 2008).

2.1.3. Robots humanoides

Los primeros bípedos autónomos aparecieron en 1967, con el estudio de Vukobratovic sobre los exoesqueletos (Vukobratovic, Frank, y Juricic, 1970). En 1972 se desarrolló el prototipo *WL-5* en la Universidad de Waseda, constituyendo el primer robot bípedo. La simplicidad de los primeros humanoides se vio menguada con las investigaciones sobre la generación de pasos, la estabilidad y el diseño (Garcia et al., 2007). Se han propuesto métodos no dinámicos limitados a entornos estáticos, así como métodos basados en adaptaciones periódicas para la generación de pasos (Roussel, Canudas-de Wit, y Goswami, 1998; Kajita et al., 2003). Además, se han realizado importantes avances en la estabilidad mediante los cálculos del punto *ZMP* (aquel en el que se aplica la resultante de las fuerzas de reacción del suelo) o el *pseudo-ZMP*, siendo la proyección del centro de masas. Aun así, otros problemas persisten: la estabilidad dinámica del movimiento de los bípedos y la distribución de cada pie en una superficie distinta son complejidades no resueltas que continúan investigándose (Takubo, Inoue, y Arai, 2005; Komura et al., 2005). En cuanto a los avances en los diseños, se han producido mejoras como la introducción de nuevos sensores y actuadores, o el desarrollo de pieles (Cannata et al., 2008) y músculos artificiales (Shahinpoor et al., 1998; Bar-Cohen, 2004; Yip y Niemeyer, 2015), que pretenden mejorar tanto la estabilidad como el movimiento.

Han surgido también nuevas investigaciones sobre exoesqueletos con el objetivo de incrementar la velocidad, fuerza y fortaleza de las personas.

El primer exoesqueleto comercializado fue el *HAL-5* (Chu, Kazerooni, y Zoss, 2005), empleado en la asistencia de personas con movilidad reducida. Aunque la mayoría de las aplicaciones de los exoesqueletos están destinadas a entornos militares o de rescate, se están empezando a emplear extremidades robóticas en el ámbito médico (Bogue, 2009; Kiguchi y Fukuda, 2004).

A pesar de que las primeras investigaciones sobre robots humanoides se centraban en los problemas locomotores, el interés de las investigaciones actuales ha sido depositado en el campo de la interacción humano-robot o *HRI* (Breazeal, 2004), siendo el objetivo la introducción de los robots en el entorno doméstico. Desde los primeros sistema básicos (Broadbent, Stafford, y MacDonald, 2009; Broekens, Heerink, y Rosendal, 2009) para la asistencia a personas de movilidad reducida como *MOVAID* (Dario y Susani, 1996) o *IMA* (Kawamura et al., 1996), se han desarrollado otros más complejos (Jayawardena et al., 2010; Wu, Fassert, y Rigaud, 2012), llegando incluso a comercializar proyectos como *Hopis* (Foulk, 2007) o *Yorisoi Ifbot* (Belew, 2007). También se han extendido a otros ámbitos, surgiendo así los proyectos *Gauguin* (Yamazaki et al., 2009), *CiceRobot* (Chella y Macaluso, 2009) o *RI-MAN* (Mukai et al., 2008) para la guía en museos. Otros estudios se han centrado simplemente en la interacción social (Upadek, 2010), emergiendo proyectos como el robot *Asimo* (Sakagami et al., 2002) para la atención del personal, el proyecto *HRP-2* (Kaneko et al., 2008) o el robot *Qrio* (Geppert, 2004), capaz de dirigir una orquesta. Se han abierto además numerosas líneas de investigación en torno a las interfaces robot-humano; con estudios sobre el reconocimiento de voz (Valin, Michaud, y Rouat, 2007; Valin et al., 2007; Yamamoto et al., 2006), electromiogramas para monitorizar el sistema nervioso (Kang et al., 2011; Wang, Yang, y Xie, 2012; Nam et al., 2014) o la percepción visual con la detección de personas y el entorno (Tresadern y Reid, 2004; Courty y Marchand, 2003), se pretende facilitar la comunicación a cualquier usuario y permitir un control directo sin experiencia previa. Otros estudios se han centrado en la transmisión de emociones, como el robot expresivo *Kismet* (Breazeal y others, 1998; Breazeal y Aryananda, 2002).

El aprendizaje social ha cobrado también cierta importancia, dando lugar a la adquisición de nuevas habilidades y la tutela en nuevas tareas (Calinon, Guenter, y Billard, 2007; Asfour et al., 2008; Nakaoka et al., 2007). Distintas investigaciones señalan la importancia de impulsar el aprendizaje, de forma

que las personas puedan enseñar de forma natural las diferentes tareas más o menos complejas a los robots (Schaal, 1999). Existen diferentes formas de aprendizaje social; unos estudios hacen uso del aprendizaje por demostración (Atkeson y Schaal, 1997a; Schaal y others, 1997), y otros en cambio emplean la imitación (Billard, 2002; Derimis y Hayes, 2002; Mataric, 2000). Aunque el aprendizaje actualmente es visto como una herramienta indispensable en el futuro ámbito robótico, parece necesaria la implantación de conocimientos previos.

2.2. Interfaz social

2.2.1. Aspectos generales

Se han estudiado repetidamente las interacciones entre los seres humanos y las nuevas tecnologías, infiriendo la posibilidad de formar los mismos vínculos que aparecen entre las personas (Reeves y Nass, 1996; Benyon y Mival, 2008). Al ser la interacción social una habilidad adquirida en toda sociedad, una interfaz social supone un modo de acceso universal en pos de una mayor facilidad para el control tecnológico, fomentando el campo de la interacción humano-ordenador o *HCI*.

En la actualidad, han surgido sistemas con interfaces para la interacción muy diversos, tanto etéreos como con soporte visual. Un ejemplo de sistemas no visuales con interfaces sociales son los *chatbots* (Jenkins et al., 2007). Estos tienen el propósito de facilitar información, por lo que intentan averiguar las expectativas de los usuarios en el sistema correspondiente. Los sistemas visuales comprenden los robots y avatares que, debido a sus similitudes morfológicas, posibilitan una comunicación natural a través de la gesticulación, la manipulación de objetos, los movimientos y otras características. Teniendo en cuenta diferentes estudios, no se reciben los mismos estímulos al observar un movimiento virtual y otro llevado a cabo en el plano real (Perani et al., 2001; Han et al., 2005). Bajo esta premisa, la robótica humanoide y el *HRI* representan una herramienta más eficaz para la interacción.

Los avatares suelen combinarse con un lenguaje natural y un motor gesticular en sistemas gráficos con el objetivo de mantener una conversación (Cassell y others, 2000). Aunque estas interfaces suelen utilizarse para realizar una tarea específica, como es el caso del agente inmobiliario *Rea* (Cassell et al.,), también pueden emplearse para el aprendizaje, como el manejo de

las herramientas navales impartido por el sistema pedagógico *Steve* (Rickel y Johnson, 2000). La introducción de caras y expresiones faciales, combinada con el lenguaje, mejora considerablemente la predisposición de los usuarios en cualquier aplicación (Takeuchi y Nagao, 1993). Desde hace años hay distintos proyectos centrados en el desarrollo de expresiones faciales con apariencia animada (Bruce, Nourbakhsh, y Simmons, 2002), dibujos (Takanobu et al., 1998; Scheeff y others, 2000) y apariencia real (Hara y Kobayashi, 1996; Hara et al., 1998). Los rasgos faciales robóticos más realistas han sido diseñados históricamente en la Universidad científica de Tokio. La tendencia actual dicta incorporar caras en los diseños de robots móviles en los entornos educativos o de entretenimiento, ya que incrementa la interacción con las personas. Por ejemplo, en el ámbito educativo se ha desarrollado en *LEGO* la cara robótica *Feelix* (Cañamero y Fredslund, 2001), que introduce una interacción táctil. Conforme aparecen más robots de servicio con facciones robóticas, crece el interés de los investigadores en analizar la reacción e interacción del ser humano; concretamente, las nuevas tendencias hacia robots terapéuticos en el campo de la robótica social asistida o *SAR* hacen indispensables los análisis sobre los efectos ejercidos en los seres humanos. Consecuencia de ello son los proyectos enfocados al tratamiento del trastorno autista (Scassellati, Admoni, y Matarić, 2012). Kozima et al. estudian las percepciones de los niños en torno a los movimientos de un brazo robótico (Kozima, Michalowski, y Nakagawa, 2009). Kim et al. exploran cómo afecta a la percepción de la personalidad de un robot la variación en la expresividad de sus gestos (Kim, Kwak, y Kim, 2008). Kiesler y Goetz muestran un estudio sobre las técnicas de caracterización de modelos mentales que las personas generan sobre los robots, variando su aspecto y expresiones (Kiesler y Goetz, 2002). Bruce et al. estudian la predisposición a interacciones cortas con robots en función de la existencia o no de expresiones faciales e indicadores de atención (Bruce, Nourbakhsh, y Simmons, 2002).

2.2.2. Emociones

Los organismos complejos gestionan las emociones como una forma de motivación para cubrir ciertas necesidades. Éstas determinan las reacciones frente a las condiciones externas y del individuo (Plutchik, 1991; Izard, 2001). Según Frijda, las emociones positivas son provocadas por eventos que desembocan en algún objetivo personal, mientras que las emociones

negativas promueven acciones para prevenir o solucionar situaciones adversas (Frijda y others, 1994). De esta forma, las emociones favorecen ciertos estímulos y evitan otros. En el estudio de Darwin sobre las expresiones, se concluye que las señales emocionales se adquirieron durante la evolución debido a su efectividad en la comunicación (Darwin, Ekman, y Prodger, 1998). Ciertamente se pueden considerar como una herramienta de modificación conductiva (Levenson, 1994). Como dice Scherer, el comportamiento que adopta un individuo en relación a un evento está determinado en parte por la reacción de otros individuos ante éste (Scherer, 1994). Distintos estudios multiculturales deducen que las emociones primarias (Ekman, 1992; Izard, 1993) facilitan respuestas adaptativas a los eventos diarios, como por ejemplo el enfado, el miedo, la alegría, la pena o la sorpresa. Cada emoción tiene un propósito biológico o social, motivando cierta respuesta adaptativa. Las emociones también son modificadas durante las interacciones, así como otras nuevas aprendidas (Ekman y Oster, 1982). Por último, Narahara et al. señalan en su estudio la relación entre las propiedades físicas de los gestos en la robótica y la percepción emocional que despierta (Narahara y Maeno, 2007)

2.2.3. Gestos

Cada día los robots tienen una mayor presencia en nuevos ámbitos y tareas; por ello, ante un entorno variable e impredecible, se hace necesaria una comunicación no verbal para referirse a distintos eventos (Sauppé y Mutlu, 2014). Tradicionalmente, la comunidad científica ha dirigido sus esfuerzos al reconocimiento de gestos en lugar de a su síntesis (Salem et al., 2010). Además, entre las pocas aproximaciones a la síntesis se ha empleado el término *gesto* para referirse a la manipulación de objetos, más que a la comunicación no verbal. Entre los sistema de detección gestual, se han desarrollado numerosos trabajos en el ámbito del aprendizaje gestual a través de la imitación (Calinon y Billard, 2006; Calinon y Billard, 2007; Atkeson y Schaal, 1997b; Breazeal y Scassellati, 2002).

Debido al éxito de las investigaciones sobre el reconocimiento de gestos, hay infinidad de estudios en torno a los robots capaces de reaccionar ante gesticulaciones (Breazeal, Hoffman, y Lockerd, 2004; Loper et al., 2009), surgiendo así proyectos como el robot *ALBERT* (Rogalla et al., 2002) o el robot *BIRON* (Haasch et al., 2004). Aunque la mayoría de los trabajos se

han centrado en el reconocimiento posiciones estáticas de la mano mediante modelos *HMMs* (Nam, Wohn, y others, 1996; Francke, Ruiz-del Solar, y Verschae, 2007; Dardas y Georganas, 2011), también han surgido estudios incluyendo información sobre la cabeza y el torso (Nickel y Stiefelhagen, 2007; Stiefelhagen et al., 2004) o sobre los movimientos dinámicos (Lee y Kim, 1999; Nickel y Stiefelhagen, 2003; Ramamoorthy et al., 2003). Murthy y Jadon desarrollan una revisión de los sistemas de reconocimiento gestual basados en el movimiento de las manos (Murthy y Jadon, 2009).

Aunque recientemente se ha puesto el foco en la síntesis e integración de gestos, la mayoría de los trabajos de síntesis emplean gestos predefinidos (Salem et al., 2010), y los trabajos de integración no avanzan más allá de emplear gestos no verbales o modelos de lenguaje n-gram para los *gestos co-verbales*. El robot *Fritz* es uno de los pocos proyectos que genera las expresiones gestuales a tiempo real (Bennewitz et al., 2007). En contraste con el ámbito robótico, la generación de *gestos co-verbales* en interfaces virtuales ha sido una línea de investigación más recurrente. El agente inmobiliario *REA* antes mencionado fue uno de los primeros interfaces en emplear este tipo de gestos. Aunque la mayoría se basa en accesos a lexicones de palabras, han surgido sistemas más complejos que también emplean información visual como *Greta* (Niewiadomski et al., 2009) o el agente *Max* (Kopp y Wachsmuth, 2004).

Existen ciertas discrepancias sobre qué puede considerarse un gesto. Por ejemplo, D. McNeill determina que un gesto es la consecuencia de movimientos espontáneos de piernas y brazos (McNeill, 2008), mientras que A. Kendon define los gestos como acciones comunicativas con una intencionalidad, distinto a cualquier otro movimiento espontáneo (Kendon, 2004). Kendon establece tres fases para establecer la ejecución de un gesto: una fase de preparación al comienzo del movimiento, el trazo del movimiento a la mitad y una fase de recuperación al final del movimiento.

Del mismo modo, la taxonomía gestual genera diferencia de opiniones y frecuentemente está orientada a tareas específicas. Kendon distingue entre *gestos autónomos*, independientes del discurso, y *gesticulación*, asociada al discurso (Kendon, 1986). Nespoulous y Lecours también realizan una separación en *gestos centrífugos* o con intenciones comunicativas y *gestos centrípetos*, interpretados como indicaciones (Nespoulous y Lecours, 1986). McNeill y Levy clasifican los gestos como *deíticos*, *icónicos*, *metafóricos* y

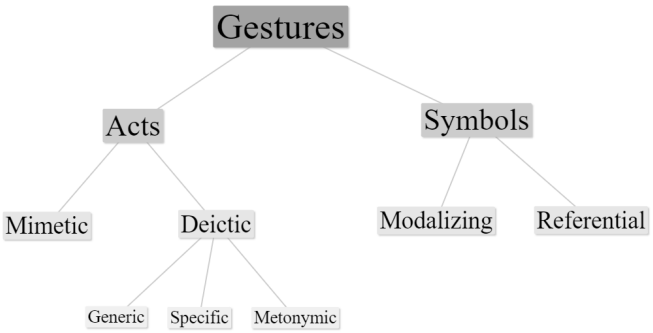


Figura 2.2: Taxonomía gestual diseñada por Francis KH Quek.

beats (esporádicos); siendo los primeros indicaciones y referencias a objetos y lugares, los segundos representaciones explícitas de objetos y acciones, los *metafóricos* descripciones de ideas abstractas, y por último, los *beats* movimientos con total carencia de significado que enfatizan el discurso (McNeill y Levy, 1980). Posteriormente, McNeill incluye un quinto tipo llamado *emblemas* en el que incluye aquellos gestos que necesitan del lenguaje para su comprensión, como los saludos (McNeill, 2008). Quek define una taxonomía más detallada y enfocada a las interfaces sociales a partir de los trabajos anteriores, disponible en la Figura 2.2 (Quek, 1994; Quek, 1995). Clasifica los gestos en dos categorías principales en función de su rol lingüístico, *símbolos* y *acciones*. Los *símbolos* pertenecen al tipo de *gestos co-verbales*, que pueden ser a su vez gestos referentes a objetos o conceptos y relativos a la estructura del discurso. En cuanto a las *acciones*, pueden ser de tipo *mímicas*, como por ejemplo simular cerrar la puerta para indicar que lo haga el receptor, o de tipo *deícticas*, relacionadas con la señalización. A su vez, estas últimas son agrupadas en *específicas* cuando el sujeto señala un objeto o lugar concreto, *genéricas* cuando se indica un objeto semejante al referido y *metonímicas* para referirse a una parte del objeto en cuestión. El trabajo de Nehaniv detallado en la introducción sigue una línea similar.

El contexto de los gestos está compuesto por muchos factores determinantes para su significado, como la cultura que lo integra, el ambiente social o el entorno laboral. Estos factores se pueden agrupar en la *información espacial* relativa al lugar donde es llevado a cabo, la *información afectiva* en relación al nivel de efusividad, la *información simbólica* conforme a los signos y la *información geométrica* respecto a la trayectoria (Mitra y Acharya, 2007).

A pesar de que los proyectos más frecuentes integran los *gestos deícticos* (Sugiyama et al., 2007; Hato et al., 2010) o de tipo colaborativos (Riek et al., 2010), se han logrado también avances con las otras tipologías. El robot *WE-4RII* se ha implementado para la expresión de emociones y gestos emotivos (Itoh et al., 2004). Los *gesto co-verbales* han sido abordados principalmente en avatares. Aunque éstos están sincronizados semánticamente, temporalmente y pragmaticamente con el discurso, la mayoría son generados de forma inconsciente según McNeil (McNeill, 1992). Los *gesto co-verbales* son influenciados por su intencionalidad comunicativa y la idea verbal que acompañan. K Bergmann et al. agrupa las principales complejidades de la sincronización de gestos en dos problemáticas, la distribución y el empaquetado de información (Bergmann, Kahl, y Kopp, 2013). Mientras que la distribución expresa cómo se complementa la información que reside en la ideas verbal y gestual, el empaquetado se refiere a la cantidad de información contenida en las mismas. Zheng y Meng afirman en su estudio que los gestos simbólicos pueden llegar a confundir si no son definidos meticulosamente (Zheng y Meng, 2012).

Con ello, se han analizado fragmentos de discurso junto con los movimientos de la mano para generar gestos (Stone et al., 2004). De forma similar, Levine et al. extraen métricas de los movimientos gestuales en distintos videos para generar gestos en tiempo real (Levine, Theobalt, y Koltun, 2009). Sin embargo, las métricas son útiles para el énfasis y el nivel emocional, pero no indican nada de la semántica. Por ello, distintos sistemas asumen la entrada de texto con los tipos de gestos y sus parámetros anotados (Kopp y Wachsmuth, 2004; Hartmann, Mancini, y Pelachaud, 2005). Neff et al. emplean etiquetas semánticas manualmente anotadas para generar gestos a través de un modelo probabilístico entrenado (Neff et al., 2008). Del mismo modo, Endrass et al. utilizan técnicas dirigidas a corpus gestuales (Endrass et al., 2010). Otros sistemas emplean gestores de diálogo para planificar gestos puntuales a partir de los objetivos comunicativos (Tepper, Kopp, y Cassell, 2004). También existen sistemas gestuales tele-operados o basados en métodos *Wizard of Oz* (Dahlbäck, Jönsson, y Ahrenberg, 1993).

La arquitectura *REA* lexicaliza los gestos para tratarlos como palabras en un generador de lenguaje natural. El sistema *BEAT* propone una aproximación más avanzada para la sincronización gestual al discurso (Cassell, Vilhjálmsson, y Bickmore, 2004). Emplea un conjunto de reglas manual para

determinar qué tipo de gesto debe lanzar, gestionando por defecto los gestos de tipo *beats*. Por último, Victor Ng-Thow-Hing et al. proponen un sistema mejorado, con la integración de todos los tipos de gestos en un robot *Honda* (Ng-Thow-Hing, Luo, y Okita, 2010). Para ello, usan un etiquetado gramatical o POS tagging y 5 modelos gramaticales, correspondientes a cada tipo de gesto definido por McNeill, para determinar su pertenencia probabilística a una tipología. Finalmente, una vez definida la tipología del gesto, emplean un sistema de reglas basado en la identificación de palabras clave para determinar el gesto específico. Estos modelos gramaticales se basan en lexicones gestuales creados mediante la anotación de conferencias bajo las siguientes premisas: hay gestos más sencillos que otros para modelar, estos suelen estar asociados a determinadas palabras, los *beats* se acentúan en los cambios de tema y las palabras pueden estar asociadas a más de un gesto debido al contexto.

La importancia de los *gestos co-verbales* de tipo simbólico radica en su transmisión de la semántica en relación al mensaje oral. Un análisis semántico de la oración a la que refiere el gesto resulta indispensable para su integración. Se han descrito sistemas que asumen la asociación gestual a las palabras más relevantes, ya que estas conforman el significado de la oración. Este trabajo Fin de Máster considera esa técnica de integración y explora la vía de establecer comparaciones semánticas entre las palabras y los gestos. Por ello, a continuación se introduce un apartado de representación y similitud semántica aplicada a la ciencia computacional.

2.3. Similitud semántica

La similitud semántica ha sido un problema central en la inteligencia artificial, la psicología y la ciencia cognitiva durante años. Concretamente, en el entorno computacional se ha estado empleando en múltiples ámbitos, como el procesamiento del lenguaje natural (Li, Bandar, y McLean, 2003), la recuperación de información (Rada et al., 1989; Srihari, Zhang, y Rao, 2000), la desambigüación del sentido de las palabras (Patwardhan, Banerjee, y Pedersen, 2003), la segmentación (Kozima, 1994) o en los sistemas de recomendación (Blanco-Fernández et al., 2008). La comparación entre distintas palabras requiere de una representación semántica de los conceptos, y estas representaciones conforman el conocimiento que fundamenta la habilidad

para hacer inferencias acerca de relaciones, similitudes, contextos, etc.

Las representaciones semánticas pueden ser generadas a partir de dos tipos de datos: experienciales o distributivos. Los datos experienciales consisten en propiedades o atributos asociados a las referencias de las palabras, o entidades a las que representan, reflejando la información derivada de la percepción e interacción con el mundo físico. Desde esta perspectiva, las palabras reflejan el mundo y sus significados son vistos como patrones conformando las propiedades de los objetos físicos. En cambio, los datos distributivos son obtenidos a través de corpus lingüísticos y describen la distribución probabilística de las palabras a través del propio lenguaje. En este sentido, el significado de las palabras se basa en su rol y uso dentro del lenguaje. Mark Andrews et al. afirman que ambos representan distintos tipos de datos de origen extralingüísitco e intralingüísitco y, por tanto, fuentes distintas de información. Por ello, postulan que las representaciones semánticas del ser humano derivan de una combinación estadística óptima entre ambos tipos (Andrews, Vigliocco, y Vinson, 2009). Una misma palabra puede ser aprendida a través de sus propiedades o atributos reales, agrupándola junto con otras similares en distintas categorías, o mediante su uso en diferentes sentencias y el contexto que desprenden.

La visión clásica de la representación semántica considera las palabras como puntos de un espacio multi-dimensional (Landauer y Dumais, 1997) o nodos interconectados en una red semántica (Collins y Loftus, 1975), cada una con sus inconvenientes. Aunque las aproximaciones espaciales resaltan la importancia de reducir la dimensionalidad y emplean algoritmos simples, están limitadas por la geometría euclídea. En cambio, las redes semánticas no están tan limitadas, pero su estructura gráfica carece de interpretación clara, además de suponer un mayor coste computacional.

2.3.1. Redes semánticas

Tradicionalmente se ha adquirido la visión experiencial de representación semántica, desarrollando bases de datos léxicas para la organización de conceptos. En 1968, Collins y Quillian propusieron las redes semánticas como almacenes de conocimiento (Collins y Quillian, 1969), lo que dio pie a la generación de las ontologías lingüísticas. Así, han surgido distintas bases de conocimiento lingüístico como *SUMO* (Niles y Pease, 2001), *WordNet* (Kilgarriff y Fellbaum, 2000) o *Multilingual Central Repository* (Gonzalez-

Agirre, Laparra, y Rigau, 2012). Entre todas ellas, la ontología más completa es el proyecto *WordNet* de la universidad de Princeton, que define una extensa base de datos léxica del inglés (Kilgarriff y Fellbaum, 2000). A partir de *WordNet* se han ido desarrollando otros proyectos similares en diferentes idiomas (Fellbaum, 1998), destacando el proyecto de ámbito europeo *EuroWordNet* (Vossen, 1998).

El creador de *WordNet*, Fellbaum, describió el proyecto como un diccionario semántico que fue diseñado como una red. Las palabras de las distintas categorías (nombres, verbos, adjetivos y adverbios) son organizadas por una variedad de relaciones semánticas en conjuntos de sinónimos o *synsets* como representación de un concepto, así como cada palabra puede tener varios significados o *senses*. Las posibles relaciones semánticas entre palabras y conceptos son diversas; algunos ejemplos son la sinonimia, la hiponimia, la pertenencia, el dominio o la autonimia. Las relaciones forman una estructura jerárquica convirtiendo la red en una herramienta útil para el procesamiento lingüístico. Se suele argumentar que la semántica de un lenguaje suele ser reflejada en su mayoría por nombres o frases nominales, por lo que es frecuente el fomento de las similitudes entre éstos a través de la meronimia, la holonimia o la hiperonimia. Esta última engloba un 80 % de las relaciones. En la taxonomía que conforma *WordNet*, los conceptos más profundos representan a su vez mayor especifidad; al contrario que los conceptos superiores, pues suponen una mayor abstracción. La Figura 2.3 muestra una pequeño ejemplo de la distribución de los datos en *WordNet*.

Las principales métricas diseñadas para las bases de datos léxicas se basan en tres principios: la *distancia entre conceptos*, la *cantidad de información* o *IC* y la *similitud entre glosas*.

La *distancia entre conceptos* se mide como el mínimo número de nodos en el árbol que conecta ambos conceptos. La métrica más simple basada en esta medida es el *Path length*, que representa la inversa del mínimo número de nodos entre dos *synsets*. Leacock y Chodorow proponen la medida definida en la Ecuación 2.1, basada en la distancia entre dos *synsets* y la máxima profundidad de la taxonomía (Leacock y Chodorow, 1998). Para evitar errores con la presencia de un nodo único, se toma D como la máxima profundidad de la taxonomía en la que está el LCS o nodo superior común inmediato.

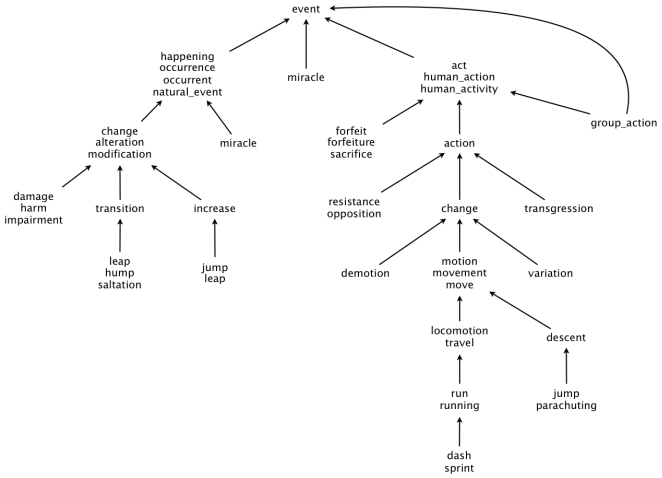


Figura 2.3: Ejemplo de la distribución de los datos en WordNet.

$$LCH = \log(length/(2 * D)) \tag{2.1}$$

Otra métrica basada en la *distancia entre conceptos* es la propuesta por Wu y Palmer en la Ecuación 2.2, que emplea la profundidad de los *synsets* y el LCS (Wu y Palmer, 1994).

$$WUP = 2 * depth(lcs)/(depth(s1) + depth(s2)) \tag{2.2}$$

Otro tipo de métricas surgen con medida de la *cantidad de información*, independientes de la conexión entre nodos. Por ejemplo, Resnik desarrolla una métrica en torno al IC del LCS, cuyo valor máximo no está limitado (Resnik, 1995). Tanto Jiang y Conrath como D. Lin definen otra métrica con la medida IC en la Ecuación 2.3 (Jiang y Conrath, 1997) y la Ecuación 2.4 (Lin y others, 1998) respectivamente.

$$JCR = 1/(IC(synset1) + IC(synset2) - 2 * IC(lcs)) \tag{2.3}$$

$$LIN = 2 * IC(lcs)/(IC(synset1) + IC(synset2)) \tag{2.4}$$

En cuanto a la *similitud entre glosas*, aparecen métricas como *Adapted Lesk*, basada en la suma de los cuadrados del número de términos superpuestos entre las glosas de dos *synsets* (Banerjee y Pedersen, 2003), o *Gloss*

Vector, como el coseno del ángulo que forman los vectores de segundo orden de co-ocurrencias de las glosas (Patwardhan, 2003). Esta última ha sido desarrollada por Patwardhan, y suele usarse con el conjunto de glosas de los *synsets* cercanos para incrementar el número de términos. Por último, la métrica *HSO* (Hirst y St-Onge) se basa en la búsqueda de conexiones léxicas entre los *senses* (Hirst, St-Onge, y others, 1998).

Conforme se ha incrementado el interés por la similitud semántica, han ido apareciendo distintos algoritmos con las implementaciones de alguna de las métricas descritas. *WS4J* es una librería *Java* que desarrolla los cálculos de todas las métricas anteriores exceptuando ambas de vectores basados en glosas (Shima, 2014). La libería *NLTK* es una herramienta *Python* enfocada al procesamiento del lenguaje natural que contiene distintas métricas *WordNet* de *distancia entre conceptos* e *IC* (Bird, 2006). En este trabajo se empleará la librería *WordNet::Similarity* diseñada por T. Pedersen et al., ya que contiene una implementación de todas las métricas descritas (Pedersen, Patwardhan, y Michelizzi, 2004).

2.3.2. Modelos *embeddings*

Siguiendo una perspectiva diferente al de las redes semánticas, Harris propuso en 1954 la hipótesis distributiva (Harris, 1954): el significado de una palabra puede ser obtenido de su entorno lingüístico (Baroni y Lenci, 2010; Baroni y Zamparelli, 2010; Blacoe y Lapata, 2012). En base a la idea de que las entidades semánticas complejas pueden construirse desde constituyentes más simples (Briscoe, 2011; Gennaro y McConnell-Ginet, 2000), surgen modelos sustentados en palabras individuales o frases cortas que generan vectores a partir del procesamiento de la distribución de información en distintos corpus por redes neuronales. Los esfuerzos para generar esos espacios vectoriales semánticos se centran en los modelos de variables latentes con la intención de reducir la dimensionalidad mediante la fusión de variables dependientes. Se han desarrollado múltiples técnicas para la reducción del espacio semántico: *Point-wise mutual information* (Bouma, 2009), *Latent Semantic Indexing* (Dumais et al., 1988), *pLSI* (Hofmann, 1999), *Latent Dirichlet Allocation* (Blei, Ng, y Jordan, 2003) y *embeddings*. De esta forma, los modelos son escalados mejorando la robustez frente a la heterogeneidad de los datos multi-relacionales. Aunque las representaciones de espacios vectoriales de palabras individuales son comprensibles, hay mucha incerti-

dumbre sobre cómo se crean esas representaciones de espacios vectoriales con frases.

Los modelos contienen representaciones de entidades usualmente en forma de vectores de dimensionalidad reducida o *embeddings*, cuyas relaciones surgen en forma de operadores. Los *word embeddings* son modelos que contienen representaciones matemáticas por cada palabra, frecuentemente vectores. Cada dimensión del vector se corresponde con una característica o propiedad semántica de la palabra. Pennington et al. considera la existencia de dos familias de modelos en el aprendizaje: métodos de factorización global de matrices como *LSI*, *LDA*, *pLSI* y *sLDA*, y métodos de ventana de contexto local como *skip-gram* o *CBOW*. En este último se basa el modelo *Word2Vec* de Mikolov (Mikolov et al., 2013). Pennington genera su propio modelo de regresión global log-bilineal llamado *GloVe* (Pennington, Socher, y Manning, 2014). Mnih y Kavukcuoglu también propusieron modelos log-bilineales, *vLBL* y *ivLBL* (Mnih y Kavukcuoglu, 2013). Otro modelo basado en la métrica *PPMI* fue propuesto por Levy et al. (Levy y Goldberg, 2014). Aunque todos estos sistemas se basan en métodos de aprendizaje no supervisado, también se han desarrollado aproximaciones semi-supervisadas *turian2010word*.

Diversos son los ámbitos y las aplicaciones que hacen uso de los *word embeddings*. Por ejemplo, un uso muy generalizado de los *word embeddings* es en técnicas de análisis de sentimientos en *Twitter*; determinando la polaridad de los *tweets* a partir de su semántica han surgido trabajos como el de D. Tang et al. (Tang et al., 2014) o X. Wang et al. (Wang et al., 2015). De forma similar son empleados en minería de opiniones, en trabajos como el de P. Liu et al. (Liu, Joty, y Meng, 2015). Otros estudios han aplicado estos modelos al análisis sintáctico (Chen y Manning, 2014) o a la recuperación de información (Vulić y Moens, 2015; Palangi et al., 2016). Otros trabajos han partido de los embeddings para implementar aplicaciones de más alto nivel, como herramientas de reconocimiento de voz (Bengio y Heigold, 2014) o traductores (Zou et al., 2013). Se han extendido hasta el ámbito de la percepción visual, apareciendo proyectos con el propósito de fusionar técnicas de ambos campos, como el trabajo de B Klein et al. sobre la asociación de imágenes a frases (Klein et al., 2015).

En el trabajo presente se pretende evaluar el uso de *word embeddings* en el problema de la anotación de gestos en texto libre, tratando de ampliar la

cobertura semántica de las unidades de texto que activan los gestos en un robot.

Capítulo 3

Diseño experimental

Este capítulo describe la experimentación llevada a cabo para determinar las bases de la metodología definitiva para la anotación gestual. Inicialmente se desarrolla una descripción de los datos empleados, prosiguiendo con las condiciones que han conformado la experimentación y finalizando con las técnicas destinadas a su realización.

3.1. Datos y condiciones

La experimentación se ha elaborado con el fin de analizar distintas técnicas de similitud semántica para la asociación de gestos propuesta. Con este análisis se pretende determinar qué técnicas y condiciones permiten una asociación óptima, teniendo en cuenta que es prioritaria una elevada precisión frente a una mejor cobertura. En este contexto, la ejecución de gestos en coherencia con el discurso adquiere mayor relevancia en comparación con el número de gestos coherentes; es decir, la activación de gestos incoherentes o erróneos es penalizada para la simulación de comportamientos naturales. Por ello, es preferible un número reducido de gesticulaciones correctas acompañadas de algún gesto inapropiado a un mayor número de expresiones coherentes pero con un incremento en la proporción de gestos erróneos.

3.1.1. Datos

Los datos para realizar la experimentación se han escogido en inglés bajo el criterio de ser el idioma más extendido globalmente (Crystal, 2012). El objetivo de la metodología buscada es la asociación de gestos a las pa-

labras relevantes de una oración entrante. Para ello, se ha asentado como base inicial la construcción de representaciones de significado de los gestos a través de la asignación de términos significativamente semánticos. En la experimentación no se emplearán gestos propiamente, sino conceptos, pues a efectos prácticos estos últimos son una abstracción de los primeros. Con ello, se han establecido un total de 60 etiquetas en representación de los conceptos. Cada concepto o etiqueta llevará asociado un único término que conformará su significado con el afán de simplificar el conjunto de datos. Estos términos, y por ende los conceptos, se establecerán en base a las palabras más frecuentes en el lenguaje anglosajón; se empleará el corpus *Word frequency data: Corpus of Contemporary American English* [1] como fuente de información para determinar la frecuencia de las palabras. Con ello, se asegura la cobertura y se garantizan unos resultados extrapolables a la mayoría de las conversaciones, ratificando la relevancia del experimento dentro del lenguaje.

Además, con el objetivo de analizar la dependencia gramatical de las relaciones semánticas, se realizará la extracción de las palabras más frecuentes en cada categoría gramatical. De este modo, las 60 etiquetas y sus términos asociados quedarán divididos en cuatro grupos. Esta división se ha llevado a cabo manteniendo la misma proporción de 15 etiquetas por categoría; es decir, los datos están formados por 15 etiquetas con nombres asociados, otras 15 con verbos, 15 más con adjetivos y el resto con adverbios. Mediante diferentes páginas web de búsqueda de palabras relacionadas se ha generado a través de los términos un conjunto de palabras sobre el que aplicar las relaciones semánticas, anotando en cada una la referencia a aquellas etiquetas a las que se correspondería. Este tipo de páginas también se han empleado con restricción gramatical, arrojando palabras con la misma categoría gramatical que su término relacionado. Con todo, se ha establecido el mismo número de términos a cada etiqueta para equilibrar la experimentación.

En resumen, se han extraído 60 términos ingleses entre los más frecuentes del lenguaje, agrupándolos en las categorías gramaticales de nombres, verbos, adjetivos y adverbios. Junto a esos términos se han definido 60 etiquetas en representación de los gestos. Por cada uno, se ha realizado una búsqueda de 20 palabras relacionadas a través de páginas web que ofrecen términos relacionados a uno dado, con la misma categoría gramatical. Pue-

[1] http://www.wordfrequency.info/

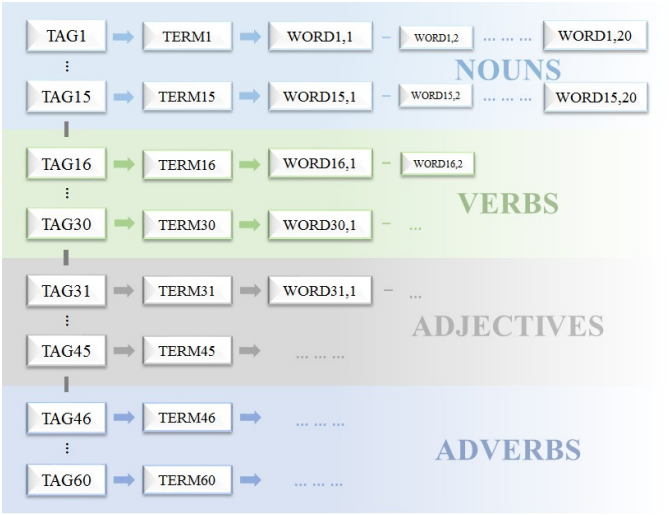


Figura 3.1: Diseño de los datos empleados en la experimentación.

de observarse un esquema de la composición de los datos en la Figura 3.1. Todas esas palabras relacionadas constituyen el conjunto de prueba sobre el que se van a aplicar las relaciones semántica con el propósito de verificar la asignación de *WORDs*, presentes en el texto, a *TERMs* relacionados con gestos.

3.1.2. Condiciones iniciales

Las condiciones de la experimentación se han determinado conforme a las distintas posibilidades planteadas para la configuración de la metodología: el tipo de métrica, el modelo de datos, el proceso empleado y la distribución de los datos.

Los experimentos consisten en el cálculo de los valores de 12 métricas de similitud semántica distintas entre cada término y palabra del conjunto de prueba, formando un total de N relaciones en base al producto del número de términos y el número de palabras. De estas métricas, 10 se basan en la estructura ontológica de *WordNet*, detalladas más adelante, y las otras dos en los modelos *word embeddings*. Éstos últimos consisten en modelos vectoriales pre-entrenados en los que cada palabra es representada mediante un vector de parámetros; por ello, las medidas de similitud más comunes entre palabras en este tipo de modelos son la distancia euclídea y la similitud coseno, ambas de tipo geométrico. Los modelos *word embeddings* propuestos

han sido generados por los proyectos *word2vec* y *GloVe*; éste último con los datos del repositorio *Wikipedia*[2]). Como los valores de ambas métricas variarán en función del modelo vectorial empleado, los resultados finalmente arrojarán dos valores distintos para cada una.

El objetivo de la experimentación es la determinación de la metodología que va a administrar los gestos, por lo que cada métrica es evaluada de forma independiente: la pertenencia de una palabra a un término se establece respecto a cada una de las métricas en paralelo. Por lo tanto, en los distintos experimentos se establecen catorce valores de similitud semántica diferentes entre cada término y palabra que se estudiarán de manera independiente.

En cuanto a la asignación de pertenencia, se han propuesto un proceso en el que una palabra activa todos aquellos términos considerados similares y otro proceso en el que una palabra únicamente activa el término más cercano. Ambas alternativas constituirán un experimento:

- La primera propuesta requiere un umbral para establecer el límite de pertenencia de una palabra a un término en relación a una métrica. Se estudiarán las relaciones *precision-recall* para intentar determinar cuál es el umbral de pertenencia óptimo de cada métrica.

- En la segunda propuesta únicamente se asignará a cada palabra el término con mayor valor de similitud en la métrica en cuestión. Para evitar la asociación de términos en ausencia de similitud, se incluirá una restricción de pertenencia mínima.

Respecto a la disposición de los datos, y como ya se ha comentado, se han definido tres escenarios distintos para la experimentación, dando lugar a tres experimentos diferentes:

- El primer escenario contempla la posibilidad de asociar palabras de distintas categorías gramaticales a un mismo término.

- El segundo escenario consiste en una separación gramatical, permitiendo únicamente las similitudes entre palabras y términos de la misma categoría.

- El último escenario es una extensión del primero, y se basa en imitar las proporciones gramaticales del lenguaje eliminando palabras de las

[2]https://es.wikipedia.org

categorías menos frecuentes; es decir, el nombre es la categoría más numerosa, mientras que el adverbio contiene el menor número de palabras.

3.2. Preparación del entorno de ejecución

Toda la experimentación se basa en el cálculo de distintas métricas de similitud semántica. Para ello, se ha utilizado la base de datos léxica *WordNet* y los modelos de lenguaje tipo *word embeddings* a través de dos algoritmos, tal y como se planteó en la propuesta.

Se ha implementado un algoritmo en C++ que acepta un término y una lista de palabras, posteriormente ejecuta los *scripts* correspondientes a la librería *WordNet::Similarity* (mencionada en la sección 2.3.2 del estado del arte), y finalmente devuelve una lista con los valores de las diferentes métricas de similitud entre cada una de las palabras y el término. Dicha librería proporciona la configuración e implementación en *Perl* de las métricas expuestas a continuación (se puede encontrar una descripción detallada en la sección 2.3.2 del estado del arte):

- ***Path length.*** Se basa en la mínima distancia medida en número de nodos entre dos conceptos.
- ***Leacock & Chodorow.*** Métrica que cuantifica el número de nodos entre conceptos en relación con la profundidad de la taxonomía.
- ***Wu & Palmer.*** Mide tanto la profundidad de los conceptos como la profundidad del concepto más abstracto inmediato, común a ambos.
- ***Resnik.*** Expresa la probabilidad de encontrar el concepto más abstracto común en un corpus determinado.
- ***Jiang & Conrath.*** Emplea la probabilidad de encontrar cada concepto, junto con su concepto común en un corpus.
- ***Lin.*** Relaciona también las probabilidades de encontrar en un corpus específico los conceptos y su confluencia.
- ***Adapted Lesk.*** Calcula la suma de los cuadrados del número de términos que se superponen en las glosas de los conceptos.

- ***Gloss Vector.*** Se basa en la formación de vectores con la co-ocurrencias de las glosas de los dos conceptos y sus posiciones angulares.

- ***Gloss Vector (pairwise).*** Establece de forma similar una relación angular entre los vectores formados a partir de las co-ocurrencias de las glosas de todos los conceptos cercanos a los dados.

- ***Hirst & St-Onge.*** Intenta determinar conexiones léxicas entre los distintos significados o *senses* de ambos conceptos.

Las representaciones *word embeddings* se han explotado a través de otro algoritmo que extrae de los ficheros ya pre-entrenados las representaciones vectoriales del término y las palabras. Aplicando álgebra básica, este algoritmo devuelve como resultado la distancia euclídea y la similitud coseno entre el término y cada palabra. Se han empleado exclusivamente dos ficheros vectoriales pre-entrenados, el modelo resultante del proyecto *word2vec*[3] de Mikolov y el modelo de Standford (proyecto *GloVe*[4]) basado en *Wikipedia*, al considerarse sus fuentes de entrenamiento las más adecuadas para un contexto general.

En esta experimentación, el criterio de acierto por cada métrica se corresponde con la asociación correcta entre una palabra y una etiqueta cuyo término fue extraído de las páginas web de relaciones semánticas utilizadas; del mismo modo, se considera un fallo a la asociación de una etiqueta cuyo término no se encontró entre los relacionados en dichas páginas web.

Una vez obtenidos los diferentes valores entre cada término y palabra según cada una de las métricas, se evaluará la propuesta a través de tres experimentos con distintos planteamientos:

- ***Asignación única frente a una asignación por umbral.*** La asignación por umbral se propone como un proceso en el que cada palabra puede ser asignada a ninguno, uno o más términos; para ello, se han calculado las curvas *PR* o *Precision-Recall* de cada métrica con el propósito de situar un posible umbral. Estas curvas representan el rendimiento del sistema en relación con un umbral de pertenencia. La *precisión* es la fracción de gestos activados que son coherentes, mientras que el *recall* consiste en la fracción de gestos coherentes que son

[3] https://code.google.com/archive/p/word2vec/
[4] http://nlp.stanford.edu/projects/glove/

activados. Adicionalmente, se ha incluido la media de los valores de las similitudes término-palabra correctas, así como la media de las similitudes incorrectas. Por otro lado, en el proceso de asignación única se calcularán las tasas de acierto. En éste, cada palabra es asignada a un único gesto y el criterio es exclusivamente el valor de la métrica.

- ***Separación por categorías gramaticales.*** La restricción por categorías se llevará a cabo eliminando todas las relaciones entre palabras y términos de distintas categorías. De esta forma, únicamente competirán por la pertenencia de cada palabra los términos en consonancia con la categoría de la misma.

- ***Datos asimétricos por categorías gramaticales.*** La última condición experimental es el desbalanceo por categoría de los datos. Para ello, se eliminarán distintas palabras durante la fase final, reduciendo considerablemente el grupo de los adverbios. Los términos nominales mantendrán el mismo número de palabras relacionadas.

Capítulo 4

Experimentación y análisis de resultados

Este capítulo presenta los resultados obtenidos en la experimentación, a partir de los cuales se realiza la propuesta de la metodología para la asociación gestual por semántica.

4.1. Ejecución de los experimentos

En el capítulo anterior se han evaluado 10 métricas diferentes de similitud semántica en base a WordNet, y otras dos métricas con modelos *word embeddings*, sobre los términos más frecuentes extraídos del corpus *Word frequency data* y las palabras relacionadas con esos términos empleados en la experimentación. Además, se han empleado dos modelos *word embeddings* distintos, *word2vec* y *GloVe*, dando lugar a dos versiones diferentes de estas dos últimas métricas. Todos los valores de estas métricas caracterizan las relaciones semánticas entre las palabras y los términos, por lo que serán aplicados a los diferentes experimentos bajo las condiciones descritas anteriormente: la asignación por umbral en oposición de la asignación única, la fusión o distinción entre las categorías gramaticales y asimetría contra el balanceamiento de los datos en cada categoría gramatical. A continuación se va a analizar qué métrica tiene un mejor comportamiento ante la tarea descrita y cuál de las técnicas propuestas presenta una mejor adaptación a los requerimientos.

Podemos dividir los resultados en tres bloques, uno por cada experimen-

to.

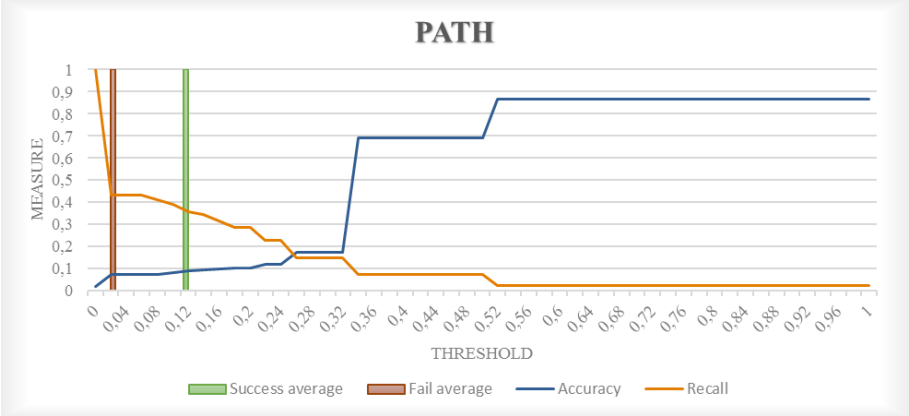
4.1.1. Asignación única vs asignación por umbral

En este bloque se presentan los resultados arrojados por un sistema con la misma proporción de palabras en cada categoría gramatical, pero sin esta separación a la hora de comparar similitudes; es decir, cada palabra puede ser asignada a cualquier etiqueta, independientemente de la categoría de los términos asociados.

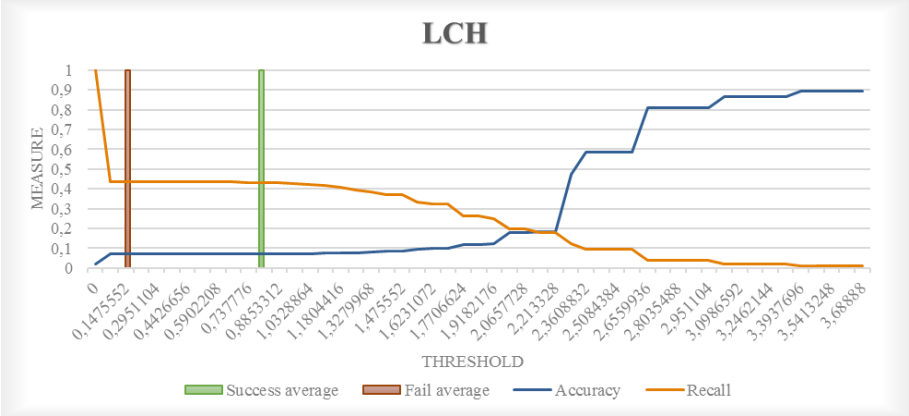
Las Figuras 4.1-4.6 muestran las curvas de *precisión* y *recall* de cada métrica para la asignación por umbral. Estas gráficas representan el rendimiento del proceso a través de la evolución de la *precisión* y el *recall* respecto a la variación del umbral de pertenencia. Además, se ha añadido tanto la media de los valores de similitud que se corresponden con los términos y palabras relacionados, como la media de los valores de similitud de aquellos que no tienen relación. Un umbral apropiado debería estar situado entre estas dos medias, filtrando la mayoría de las relaciones semánticas incorrectas y permitiendo, a su vez, todos los valores superiores.

La Figura 4.1 presenta las curvas *PR* de las métricas basadas en la *distancia entre conceptos* dentro de la taxonomía. La métrica *Path length* muestra un lento incremento de la precisión hasta el umbral 0.33; a partir de éste, aunque la precisión aumenta rápidamente y de forma escalonada hasta un 0.7, el escaso *recall* indica que apenas se activan gestos. La baja precisión entre la media de los aciertos y los fallos señala una mala respuesta a la hora de identificar los gestos correctos. La mayor separación entre las medias de aciertos y fallos de la métrica *LCH* favorece el intento de situar un umbral apropiado, aunque la escasa precisión que se mantiene hasta el valor 2.22 invalida cualquier iniciativa. En cuanto a la última métrica, a pesar de que la variación de *WUP* es más suave no presenta mayores cambios en el comportamiento. En general, las métricas de *distancia entre conceptos* no contienen ningún valor aconsejable para situar un umbral de pertenencia.

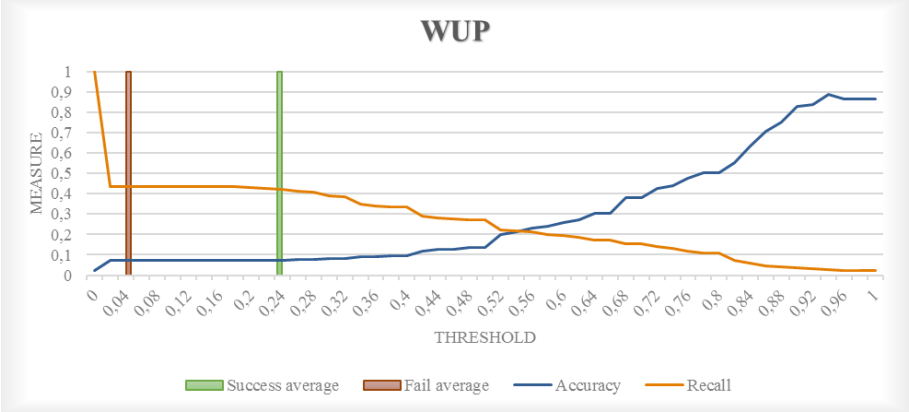
Respecto a las curvas *PR* de las métricas que emplean la *cantidad de información* (véase la Figura 4.2), todas presentan una rápida disminución del *recall* en favor de un leve incremento de la precisión. Aunque en la métrica *RES* se podría situar un umbral en el valor 0.61, relativamente cerca de la media de aciertos, el *recall* sería en torno a 0.3; como se ha comentado, en este contexto se prioriza la precisión. Otra posibilidad sería apurar el



(a) Path length. Precision and recall.

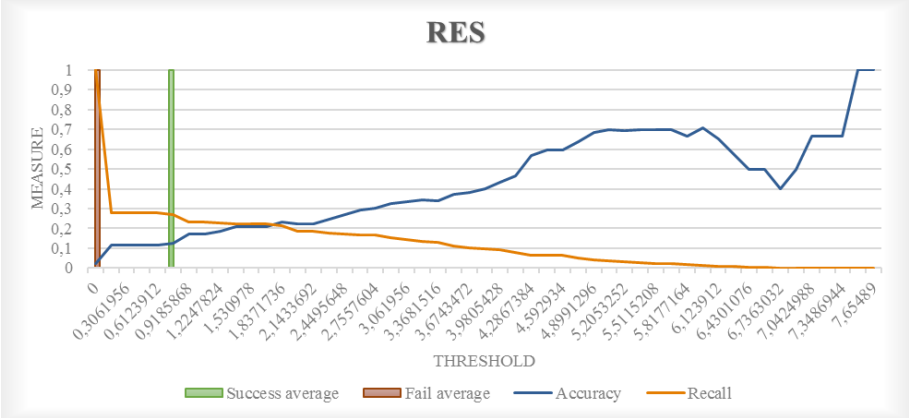


(b) Leacock & Chodorow. Precision and recall.

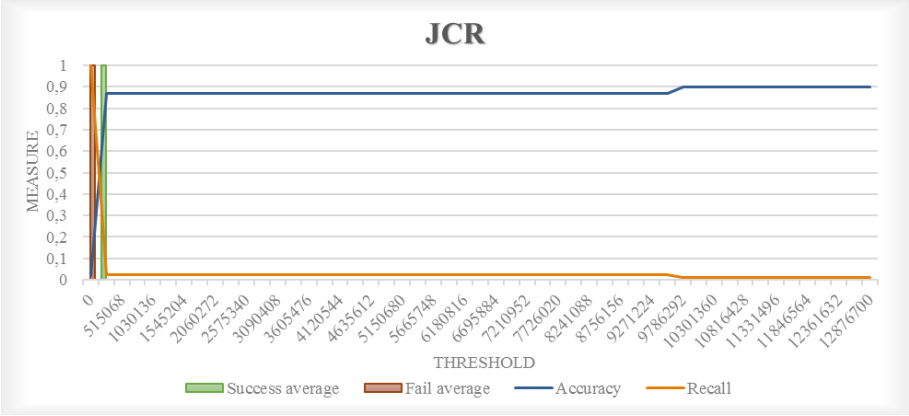


(c) Wu & Palmer. Precision and recall.

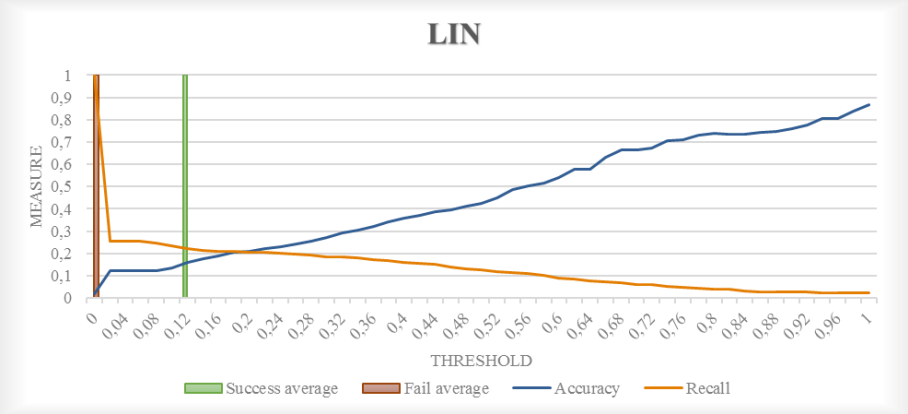
Figura 4.1: Gráficas de precisión y *recall* en función del umbral. Métricas basadas en la *distancia entre conceptos* de *WordNet*.



(a) Resnik. Precision and recall.



(b) Jiang & Conrath. Precision and recall.



(c) Lin. Precision and recall.

Figura 4.2: Gráficas de precisión y *recall* en función del umbral. Métricas en torno al *IC* basadas en *Wordnet*.

recall a 0.15, consiguiendo una precisión de 0.32. La métrica *LIN* tiene un comportamiento similar, empeorando ligeramente tanto la precisión como el *recall* entre las medias, y aumentando la precisión hasta casi 0.4 en el valor de *recall* 0.15. La métrica *JCR* parece no contener información acerca de la similitud, arrojando resultados nulos sobre casi todas las palabras.

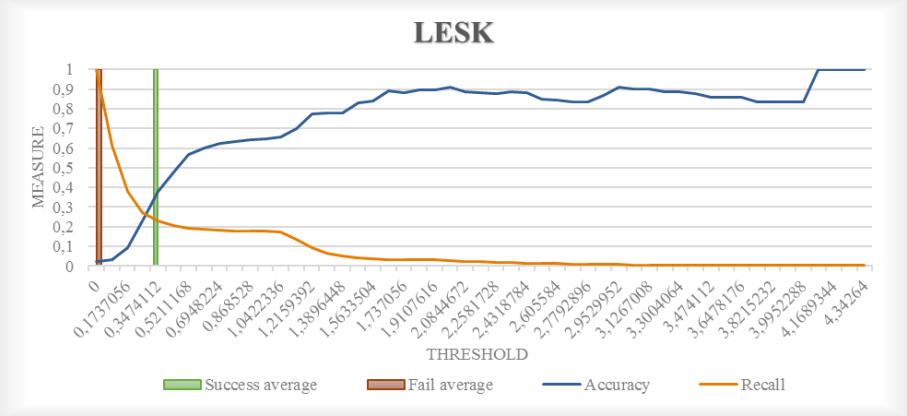
En cuanto a las gráficas de las métricas basadas en la *similitud entre glosas* de la Figura 4.3, presentan evoluciones de la precisión en forma de raíz cúbica. La métrica *Lesk* consigue una evolución exponencial de la precisión, pudiendo situar un umbral cerca de la media de aciertos en torno al valor 0.6, con una precisión de 0.6 y 0.2 de *recall*. Respecto a la métrica *Vector*, presenta precisiones más bajas en relación a los valores de *recall*, así como un limite máximo de 0.8 para la precisión. *Vector pairs* muestra cierto solapamiento entre los valores que representan similitudes entre términos y palabras no relacionadas y los valores que presentan similitud semántica, según la cercanía de las medias. Con ello, no parece albergar ningún umbral válido.

La Figura 4.4 vuelve a contener una gráfica escalonada, señal de la existencia de agrupaciones de relaciones semánticas en torno a unos valores. La media de aciertos es cercana al 3.2, dónde la precisión y el *recall* coinciden en 0.3.

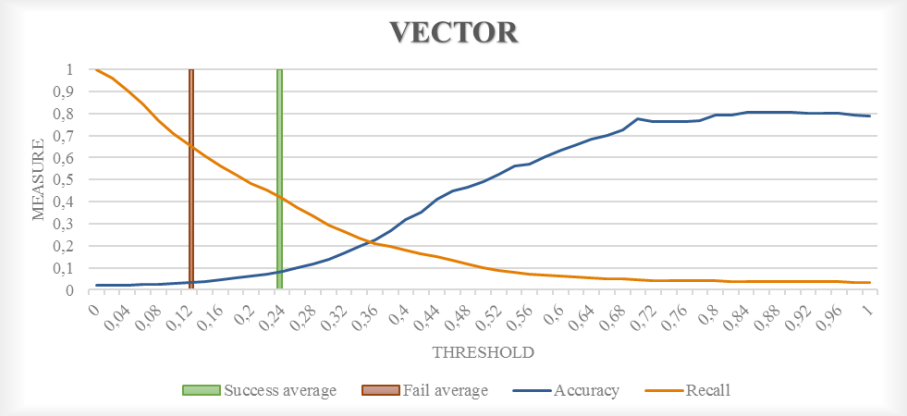
La *distancia euclídea* de cada modelo *word embedding* aparece en la Figura 4.5. El comportamiento de la precisión es nefasto, apenas alcanzando el valor de 0.3 en ambas gráficas. En cambio, la *similitud coseno* presente en la Figura 4.6 muestra mayores incrementos. En particular, el modelo *word2vec* permite la ubicación de umbrales en torno a 0.35 con una precisión y *recall* de 0.3.

Como queda reflejado, cuanto mayor es el umbral, menos palabras son asignadas, suponiendo una disminución del *recall* y un incremento de la precisión. Del mismo modo, un umbral bajo supone la asignación de la mayoría de las palabras, relacionadas o no, por lo que afecta negativamente a la precisión. El *recall* se maximiza con umbrales nulos, donde se asignan todas las palabras, incluyendo aquellas realmente relacionadas.

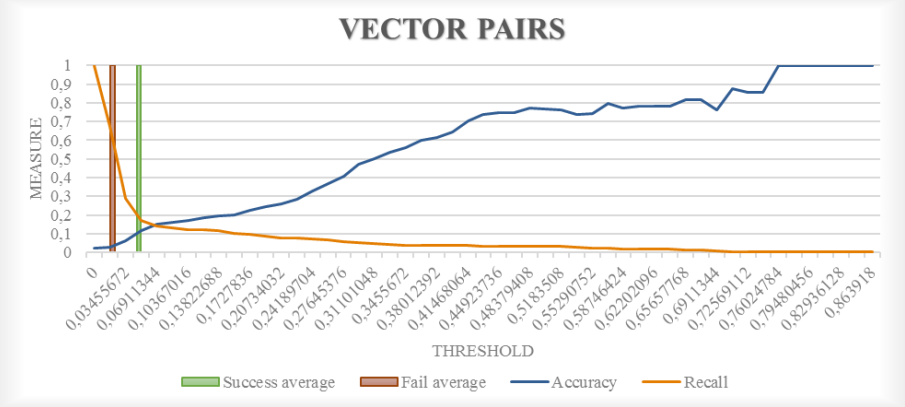
La asignación única de las palabras se ha llevado a cabo con las etiquetas con mayor valor de similitud. De esta forma, una palabra es asignada a aquella etiqueta cuyo término presenta el valor más alto. Se han calculado los porcentajes de acierto por cada métrica en la Tabla 4.1, así como la



(a) Adapted Lesk. Precision and recall.



(b) Gloss Vector. Precision and recall.



(c) Gloss Vector (pairwise). Precision and recall.

Figura 4.3: Gráficas de precisión y *recall* en función del umbral. Métricas basadas en la *similitud entre glosas* en *WordNet*.

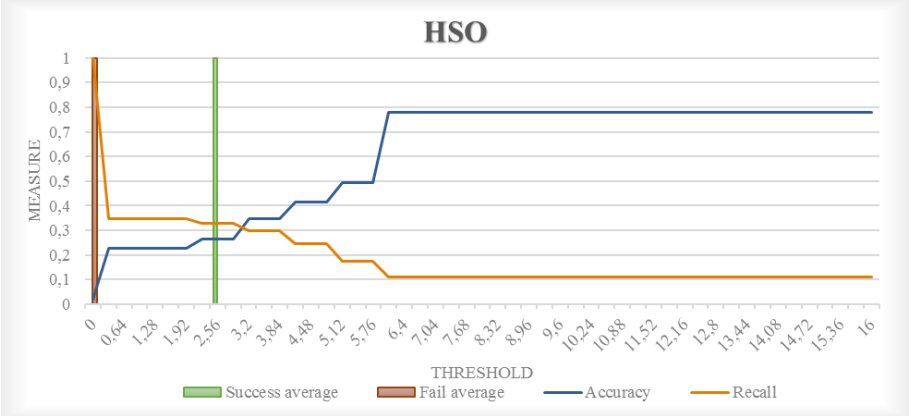
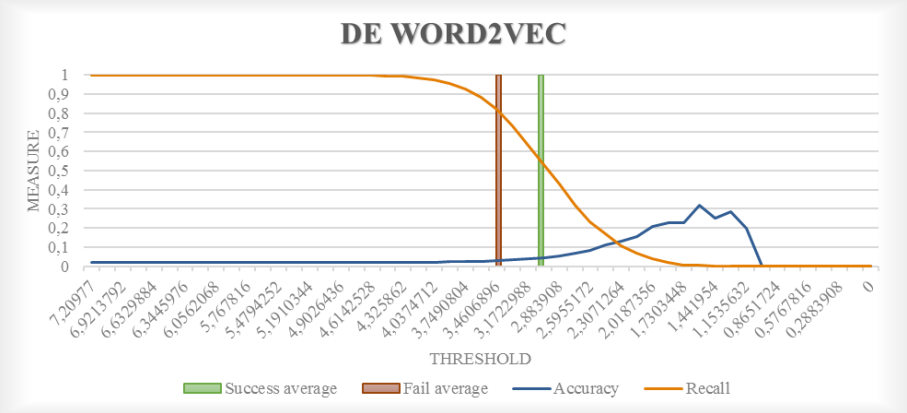


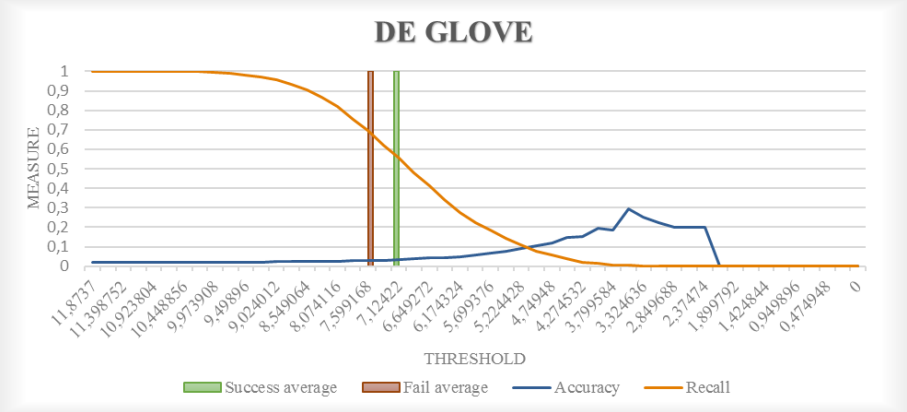
Figura 4.4: Gráfica de precisión y *recall* en función del umbral. Métrica Hirst & St-Onge.

mejora de cada métrica respecto a su inferior. Sin duda, la *similitud coseno* aplicada a los vectores de *word2vec* es la métrica que ofrece los resultados más prometedores, con una tasa de acierto del 42 % y un incremento respecto a la segunda del 34 %. Cabe destacar la semejanza de resultados entre las métricas *LESK*, *HSO* y la *distancia euclídea* de *word2vec*. Otro salto cuantitativo se produce entre *Vector* y *WUP*, con un 23 % de incremento. La métrica con peor tasa de acierto es *Vector pairs*, resultando el resultado de la primera en una mejora acumulada del 162 % respecto a este.

Con estas gráficas de precisión y *recall*, la idea de establecer un umbral solo parece factible en cinco de ellas. En base a la propuesta de este proyecto, la precisión es más relevante que el *recall*; aun así, es necesario cierto *recall* para que existan aciertos, pues representa la proporción de asignaciones correctas realizadas. Según la evolución de las curvas, no puede establecerse un umbral con una buena precisión sin que el *recall* prácticamente desaparezca; en ese caso, el número de gestos asociados es mínimo. La precisión que se obtiene con un mínimo de *recall* de 0.3 no supera el 0.4; esto quiere decir que se asocian un 30 % del total de los gestos, y de todos los gestos asociados, un 40 % han sido correctos. En cambio, aunque con el proceso de asignación única se alcanzan precisiones máximas del 0.4, garantizan la asignación de todas las palabras. Por lo tanto, ante los dos procesos con semejantes porcentajes aciertos, el *recall* de la asignación única determina su elección en la metodología.

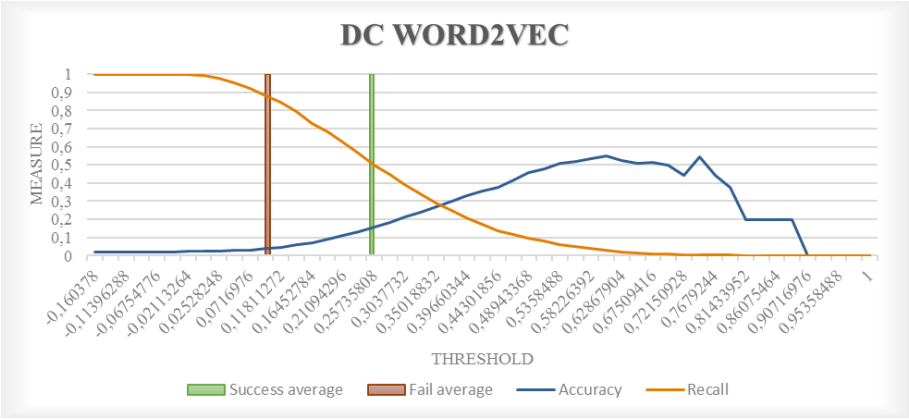


(a) Euclidean distance from word2vec. Precision and recall.

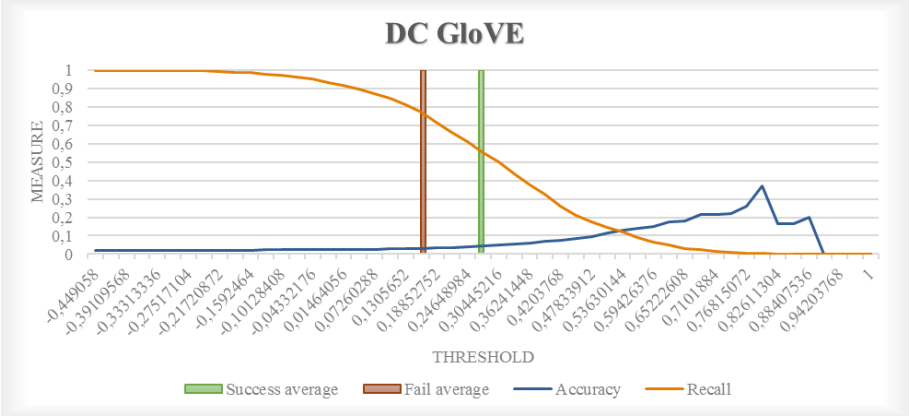


(b) Euclidean distance from GloVe. Precision and recall.

Figura 4.5: Gráficas de precisión y *recall* en función del umbral. Métricas basadas en la *distancia euclídea* de vectores *word embeddings*.



(a) Cosine distance from word2vec. Precision and recall.



(b) Cosine distance from GoVe. Precision and recall.

Figura 4.6: Gráficas de precisión y *recall* en función del umbral. Métricas basadas en la *similitud coseno* de vectores *word embeddings*.

Métrica	**Tasa de acierto**	**Mejora incremental**
Cosine similarity word2vec	0,41651865	0,347701149
Euclidean distance word2vec	0,30905861	0,026548673
LESK	0,30106572	0,01497006
HSO	0,29662522	0,021406728
Cosine similarity GloVe	0,29040853	0,086378738
Vector	0,26731794	0,233606557
WUP	0,21669627	0,016666667
Euclidean distance GloVe	0,21314387	0,057268722
LCH	0,20159858	0
Path length	0,20159858	0,03652968
JCR	0,19449378	0,00921659
LIN	0,19271758	0,058536585
RES	0,18206039	0,145251397
Vector pairs	0,1589698	

Tabla 4.1: Asignación única de datos balanceados sin separación por categorías gramaticales.

4.1.2. Separación por categorías gramaticales

Una vez determinada la asignación al mejor valor, se ha querido comprobar cómo afecta la separación por categorías gramaticales. Así, cada palabra ha sido aislada de los términos con categorías gramaticales distintas, y a efectos de la experimentación, esas relaciones son ignoradas. La Tabla 4.2 contiene las nuevas tasas de acierto. Todas las métricas incrementan su tasa de aciertos de forma considerable respecto a la asociación sin categorías. La *similitud coseno* parece especialmente sensible a las categorías, incrementándose hasta en un 20 % en ambos modelos. De esta forma, el modelo *word2vec* se sitúa en una tasa de acierto del 0.5, y el modelo *GloVe* se ubica en segunda posición con una tasa del 0.44, con una diferencia entre ambos del 14 %. Por otro lado, tanto *Vector* como *Vector pairs* muestran una alta dependencia de las categorías, incrementándose un 42 % y un 59 % respectivamente. Con ello, *Vector pairs* se sitúa junto a las demás métricas basadas en la *similitud entre glosas*.

Observando detenidamente el orden de las métricas respecto a la tasa de aciertos, se pueden agrupar en base a los principios que utilizan. Las métricas con peores resultados son las que se basan en la *cantidad de información* o *IC*, con una tasa del 0.21. Posteriormente se encuentran con una tasa similar las métricas basadas en la *distancia entre conceptos*. Entre tasas del 0.32 y

Métrica	**Tasa de acierto**	**Mejora incremental**
Cosine similarity word2vec	0,50534759	0,136272545
Cosine similarity GloVe	0,44474153	0,043933054
Euclidean distance word2vec	0,42602496	0,124705882
Vector	0,37878788	0,05721393
Euclidean distance GloVe	0,35828877	0,049608355
LESK	0,34135472	0,05801105
HSO	0,32263815	0,274647887
Vector pairs	0,25311943	0,075757576
WUP	0,23529412	0,068825911
LCH	0,2201426	0
PATH	0,2201426	0,008163265
LIN	0,21836007	0,020833333
JCR	0,21390374	0,030042918
RES	0,20766488	

Tabla 4.2: Asignación única de datos balanceados con separación por categorías gramaticales.

0.38 se ubican las métricas de *similitud entre glosas*. Las tasas más altas se producen con las métricas geométricas, encabezadas por la *similitud coseno*.

Haciendo un desglose de los aciertos por categorías, se observa en las Tablas 4.3, 4.4, 4.5 y 4.6 una tasa de aciertos de casi el 70 % para los nombres, que difiere bastante de la tasa del 36 % de los adverbios. Los adjetivos y verbos tienen tasas similares, del 50 % y 45 % respectivamente. Esto indica que la similitud semántica entre los adverbios del experimento no es lo suficientemente significativa para arrojar uno de cada tres resultados aceptables. Una causa de ello podría ser la saturación de los datos en relación a los adverbios; es decir, que los datos contengan más adverbios relacionados con un mismo término de los existan debido a lazos de similitud débiles.

En la similitud semántica entre nombres se observa un considerable incremento de las métricas basadas en la *distancia entre conceptos* e *IC*. Mientras que métricas en relación a la distancia prácticamente han doblado su tasa de aciertos, pasando al 0.4, las métricas de *similitud entre glosas* permanecen inmutables. Sin duda, lo más destacable son los niveles de aciertos que alcanzan las métricas de *similitud coseno* de ambos modelos, casi el 0.7 y el 0.64. La distancia entre las tasas de la primera métrica y la última se ha ensanchado, así como las distancias entre cada métrica geométrica. La última *distancia euclídea* dista un 18 % de la primera métrica de *distancia*

Métrica	**Tasa de acierto**	**Mejora incremental**
Cosine similarity word2vec	0,695652174	0,08900524
Cosine similarity GloVe	0,638795987	0,17177914
Euclidean distance word2vec	0,545150502	0,14788732
Euclidean distance GloVe	0,474916388	0,18333333
WUP	0,401337793	0,02564103
LCH	0,391304348	0
PATH	0,391304348	0,0733945
LIN	0,364548495	0,00925926
JCR	0,361204013	0,03846154
HSO	0,347826087	0,10638298
Vector	0,314381271	0
RES	0,314381271	0,03296703
LESK	0,304347826	0,16666667
Vector pairs	0,260869565	

Tabla 4.3: Asignación única de datos balanceados con separación por categorías gramaticales: Nombres.

entre conceptos.

La similitud entre verbos concentra las tasas de aciertos de todas las métricas en un rago de 0.32 a 0.45, a excepción de la métrica *Vector pairs* que permanece igual. En esta categoría no parece favorecerse ningún tipo de métrica, situándose en 3ª, 4ª, 5ª y 6ª posiciones las métricas *Vector*, *WUP*, *Similitud conseno* y *RES*, una de cada tipo.

Los adjetivos no muestran similitudes semánticas a través de medidas basadas en la *distancia entre conceptos* o el *IC*. La estructura de *WordNet* soporta grupos de nodos interconectados menos numerosos en la taxonomía de adjetivos, dando lugar a un mayor aislamiento que dificulta este tipo de métricas. Sin embargo, las métricas basadas en *similitud entre glosas* como *Vector*, *LESK* y *HSO* presentan tasas de acierto similares al 0.5 de la *similitud coseno*.

De manera similar, las métricas basadas en *IC* o *distancia entre conceptos* son inviables en los adverbios. *HSO* y *LESK* descienden hasta el 0.14 y 0.23 respectivamente. Se mantiene la *similitud coseno* como la métrica con una mayor tasa de aciertos con 0.36, seguida de la métrica *Vector* con un 0.31. La métrica *Vector pairs* permanece en 0.26, al igual que en el resto de categorías.

Por lo tanto, parece que la restricción por categorías consigue aumentar

Métrica	**Tasa de acierto**	**Mejora incremental**
Cosine similarity word2vec	0,45054945	0,09821429
Euclidean distance word2vec	0,41025641	0,03703704
Vector	0,3956044	0,03846154
WUP	0,38095238	0,00970874
Cosine similarity GloVe	0,37728938	0,04040404
RES	0,36263736	0,03125
LIN	0,35164835	0,0212766
LESK	0,34432234	0,02173913
HSO	0,33699634	0
JCR	0,33699634	0,02222222
LCH	0,32967033	0
PATH	0,32967033	0,02272727
Euclidean distance GloVe	0,32234432	0,22222222
Vector pairs	0,26373626	

Tabla 4.4: Asignación única de datos balanceados con separación por categorías gramaticales: Verbos.

Métrica	**Tasa de acierto**	**Mejora incremental**
Cosine similarity word2vec	0,50378788	0
Vector	0,50378788	0,01526718
LESK	0,49621212	0,0234375
Cosine similarity GloVe	0,48484848	0,00787402
HSO	0,48106061	0,03252033
Euclidean distance word2vec	0,46590909	0,26804124
Euclidean distance GloVe	0,36742424	0,67241379
Vector pairs	0,21969697	1,9
JCR	0,07575758	0
LCH	0,07575758	0
LIN	0,07575758	0
PATH	0,07575758	0
RES	0,07575758	0
WUP	0,07575758	

Tabla 4.5: Asignación única de datos balanceados con separación por categorías gramaticales: Adjetivos.

Métrica	**Tasa de acierto**	**Mejora incremental**
Cosine similarity Google	0,36013986	0,14444444
Vector	0,31468531	0,125
Euclidean distance word2vec	0,27972028	0,03896104
Cosine similarity GloVe	0,26923077	0,01315789
Vector pairs	0,26573427	0,01333333
Euclidean distance GloVe	0,26223776	0,11940299
LESK	0,23426573	0,71794872
HSO	0,13636364	0,95
JCR	0,06993007	0
LCH	0,06993007	0
LIN	0,06993007	0
PATH	0,06993007	0
RES	0,06993007	0
WUP	0,06993007	

Tabla 4.6: Asignación única de datos balanceados con separación por categorías gramaticales: Adverbios.

la tasa de aciertos global, intensificándose con los nombres. Mientras que las métricas basadas en la *distancia entre conceptos* e *IC* se adaptan bien a los nombres y verbos pero no sirven para adjetivos y adverbios, las métricas basadas en la *similitud entre glosas* arroja buenos resultados para estos últimos y no es tan útil para los nombres y verbos. La métrica que siempre arroja las mejores tasas es la *similitud coseno*; la *distancia euclídea* queda por detrás de ésta.

4.1.3. Datos asimétricos por categorías gramaticales

Hasta ahora hemos asumido un balanceo en el número de palabras relacionadas con cada término asociado a un gesto. Sin embargo, partiendo de la idea de que en el lenguaje hay una mayor abundancia de nombres, seguidos de adjetivos y verbos, y en menor cantidad adverbios, esta última configuración del experimento consiste en la modificación de los datos para llegar a las mismas proporciones. Para ello, se han eliminado todos los adverbios, adjetivos y verbos no arrojados en las primeras búsquedas de las páginas web de palabras relacionadas. De esta forma, los nombres permanecen con 20 palabras relacionadas, los adjetivos bajan a una media de 18, los verbos a 15 y los adverbios en torno a 11. Estos últimos presentan menos palabras relacionadas por cada término.

Bajo estas condiciones, se consigue incrementar la tasa de aciertos global al 0.53. Mientras que los resultados de los nombres y adjetivos permanecen constantes y la tasa de los verbos apenas sube un punto, los adverbios alcanzan el 0.4. Aunque se consigue una mejora no despreciable, las tasas parecen no superar esos valores. Por lo tanto, cabe afirmar que los métodos empleados son más eficaces con los nombres, y menos con los adverbios, lo que podría hacer pensar en establecer un orden de prioridades a la hora de calcular similitudes.

4.2. Propuesta definitiva

En base a los resultados anteriores, se determina que la metodología que más se adecua a la similitud semántica entre las palabras y un gesto consiste en:

- la búsqueda de vectores de representación semántica en modelos tipo *word embeddings*.
- el cálculo de la *similitud coseno* entre estos vectores.
- la exclusión de similitudes entre palabras y términos de distinta categoría gramatical.
- la asignación exclusiva del término más cercano a la palabra, siempre que se encuentre lo suficientemente cerca; es decir, supere cierto umbral.
- la priorización de los nombres, verbos, adjetivos y adverbios, en este orden.

La tarea de la propuesta comienza con el procesamiento de un texto mediante la separación en oraciones, continúa con la tokenización de las palabras y su clasificación gramatical, y finalmente calcula la distancia semántica de los nombres, verbos, adjetivos y adverbios a los términos asociados a los gestos para devolver el gesto más cercano.

En la experimentación no se ha tenido en cuenta la posibilidad de que no haya ningún término relacionado con la palabra. Con el objetivo de mejorar la precisión, se ha decidido incluir un umbral mínimo para la asignación, de forma que el gesto más cercano semánticamente a la palabra no será anotado

si esa distancia supera un valor mínimo. Este umbral no se especifica en base a los resultados de la experimentación, sino a partir de los resultados prácticos una vez implementado el algoritmo que se especifica en el siguiente capítulo.

Capítulo 5

Aplicación práctica de la metodología

Este capítulo describe la implementación de un algoritmo para la anotación gestual, siguiendo la metodología configurada tras el análisis, el desarrollo de un proceso de integración de *gestos co-verbales* y su aplicación sobre un cuento de *Teo* en español. La narración anotada será integrada junto con la librería de animaciones en un robot *Nao* para su visualización.

5.1. Implementación del algoritmo

El algoritmo a desarrollar parte de un texto de entrada que debe *tokenizar*; posteriormente tiene que etiquetar las palabras en función de su categoría gramatical, consiguiendo así extraer las palabras más relevantes (coincidentes con los nombres, verbos, adjetivos y adverbios); tras el etiquetado, deberá medir las similitudes semánticas en base a las métricas *similitud coseno* o *distancia euclídea* entre dichas palabras y los términos asociados a los gestos; finalmente decidirá que gestos son los más cercanos en base a un umbral y asignará dichos gestos a las palabras. El resultado final del algoritmo consistirá en un listado de las palabras que han sido seleccionadas, especificando su posición en el texto, el gesto asociado y el valor semántico calculado.

Los procesos de división en sentencias, tokenización de las palabras y detección de su clase gramatical han sido resueltos mediante el uso de la herramienta de análisis morfosintáctico del lenguaje *FreeLing*. Con ello, el

algoritmo del sistema se ha desarrollado en *C++*, integrando las librerías de *FreeLing* y el cálculo de los valores de las métricas geométricas. Se ha fijado la indicación del fichero que contiene el modelo vectorial pre-entrenado de *word embeddings* como parámetro, así como el índice de priorización de las categorías gramaticales. Como primera aproximación se analizarán en primer lugar las palabras de la oración que pertenezcan a la primera categoría; en el caso de que no se encontrase ninguna relación, se analizarán las palabras de la segunda categoría, y así progresivamente. Además del umbral mínimo, es necesario indicar al algoritmo la lista de gestos y términos con la que se pretende asociar las palabras relevantes de las oraciones. Para ello, se ha elegido el formato expuesto en la Fórmula 5.1, donde TAG es la etiqueta de cada gesto, *key* son los términos significativos asociados a ese gesto y el *postag* de cada término es su categoría gramatical.

$$\begin{aligned} LISTA_{GESTOS} = \quad [\quad & TAG_1 = [key_{1,1}\{postag_{1,1}\},\ key_{1,2}\{postag_{1,2}\},\ ...], \\ & TAG_2 = [key_{2,1}\{postag_{2,1}\},\ key_{2,2}\{postag_{2,2}\},\ ...], \\ & \qquad\qquad\qquad\qquad \vdots \\ & TAG_n = [key_{n,1}\{postag_{n,1}\},\ key_{n,2}\{postag_{n,2}\},\ ...], \\ ] & \end{aligned} \tag{5.1}$$

Finalmente, la salida del algoritmo se incluirá en un módulo de integración con el texto independiente del mismo; de esta forma, el algoritmo implementado en base a la propuesta se ha encapsulado, posibilitando una mejora futura progresiva de este último módulo con el propósito de mejorar la naturalidad. El formato de salida empleado en el algoritmo será el mostrado en la Fórmula 5.2; donde $WORD_n$ es la palabra designada, $POSITION_n$ es el número de letras que precede a ésta, $VALUE_n$ es el valor de la métrica empleada y TAG_n es la etiqueta del gesto.

$$OUTPUT = TAG_n\ VALUE_n\ POSITION_n\ WORD_n \tag{5.2}$$

Para facilitar el acceso al algoritmo desde cualquier entorno y a todos los usuarios, se ha encapsulado en un servicio web, junto al modulo de integración desarrollado hasta la fecha. Se ha diseñado una interfaz web para

acceder a dicho servicio con el objetivo de facilitar la ejecución del algoritmo sobre un cuento propuesta en este capítulo.

5.2. Módulo de integración de etiquetas en texto

Como se ha mencionado, este módulo engloba un conjunto de heurísticas para gestionar e integrar todos los gestos detectados en el texto, y es susceptible de sufrir futuras modificaciones para mejorar la naturalidad de los gestos en su sincronía con el discurso.

Entre las posibles modificaciones, se prevé la introducción de un método para calcular el período de tiempo consumido por cada frase durante el discurso. De esta forma, es sería posible una mejor gestión de las gesticulaciones y su duración, pudiendo ejecutarlas parcialmente o eliminarlas si es preciso.

Hasta el momento de realización de este trabajo, el módulo de integración se ha desarrollado para introducir el gesto durante la pronunciación de la palabra a la que ha sido asociado, suspendiendo el resto del discurso hasta su finalización. Además, gestiona la posibilidad de anotar una palabra con distintos gestos debido a una similitud semántica similar, asignando uno de ellos aleatoriamente.

5.3. Integración en *Nao*

Una vez presentada la experimentación y su análisis, así como la propuesta definitiva y su implementación en un algoritmo, se va a realizar una demostración del potencial que aquí se plantea. Para ello, se ha reescrito una versión libre del cuento *Teo el valiente* en la que se incluyen temas relacionados con los gestos proporcionados por la librería *Animations* de *Nao*. Esta librería la componen un conjunto de gesto pre-programados que se han incluido de manera oficial en el nuevo *framework NaoQi 2.x.*, como consecuencia de la alta popularidad del prototipo.

El robot *Nao* es un prototipo, diseñado por la empresa francesa Aldebaran-Robotics [1], que cuenta con 25 grados de libertad (14 para la parte inferior y 11 para la superior). La Figura 5.1 muestra el aspecto general del prototipo y los elementos que lo componen. Las capacidades cognitivas de un

[1] Empresa Aldebaran-Robotics. URL:www.aldebaran-robotics.com

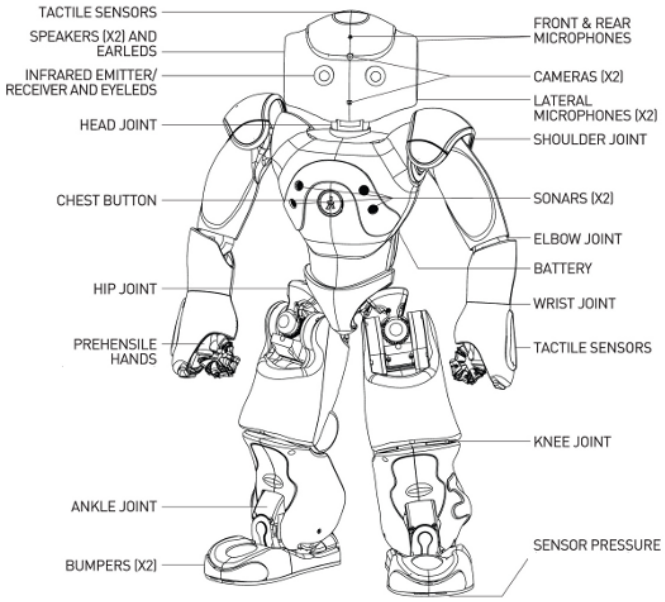


Figura 5.1: Prototipo NAO H25.

robot dependen en gran medida del software integrado en su *framework*. La arquitectura *NaoQi* conforma un entorno modular y distribuido dirigido a eventos, sustentando ejecución paralela y secuencial de distintos métodos. Así, cada funcionalidad, como la visión, el movimiento o el reconocimiento de voz, está modularizada, pudiendo ejecutarse de forma independiente. Un mayor detalle se encuentra en la revisión del diseño y la arquitectura realizada por los autores D. Gouaillier, V. Hugel y P. Blazevic (Gouaillier et al., 2008). Aunque existen otros robots con diseños e interfaces más avanzados, no son accesibles a la mayoría de investigadores; su escaso mantenimiento y bajo coste han fomentado la adquisición de este prototipo en numerosos centros educativos y de investigación.

El cuento empleado en esta demostración es una versión libre de *Teo el valiente*. El texto anotado mediante el algoritmo se ha empleado sobre el módulo *ALAnimatedSpeech* de *Nao*. Éste detecta las etiquetas *^start()* y *^wait()*, ejecutando el gesto indicado en sincronización con el texto acotado entre ambas. La etiqueta *^wait()* señala que la gesticulación tiene prioridad y se debe suspender el discurso posterior hasta su ejecución.

5.3.1. Visualización

Se ha grabado un vídeo para la visualización de la integración gestual en el cuento. Éste se puede visualizar en la siguiente dirección *https://www.youtube.com/embed/lzViIUvNyJM?rel=0&hl=es&cc_lang_pref=es&cc_load_policy=1*. Se han incluido subtítulos para mostrar qué palabras activan gestos; para ello, se detalla entre paréntesis qué término asociado al gesto es detectado y su valor de similitud semántica según la métrica *similitud coseno* con el formato $(TERM_n =\sim VALUE_n)$.

Capítulo 6

Conclusiones y trabajo futuro

Este capítulo recopila las diferentes conclusiones extraídas del trabajo, proponiendo algunas líneas de desarrollo futuras.

6.1. Conclusiones

La robótica pretende ser uno de los pilares de la vida diaria de las personas, facilitando todas aquellas tareas imaginables. Aún quedan grandes retos por superar, como mejorar la flexibilidad con la creación de prototipos de propósito general. Conforme se incrementa la complejidad en el comportamiento de los diseños, mayor es el interés por simplificar y facilitar su control. Prácticamente todos los estudios señalan que el camino más eficiente hacia el acceso universal a la robótica es la integración de interfaces sociales para la interacción humana en el ámbito *HRI*. En este punto cobra importancia la expresión emocional y, por extensión, la expresión gestual; las emociones han sido estudiadas por su función evolutiva en la mejora comunicativa. Además, con la pretensión de mantener una interacción prolongada, no carece de interés abordar el nivel de empatía que despiertan los robots. Distintos estudios indican que los seres humanos son capaces de desarrollar sentimientos afectivos por las máquinas, con una mayor aceptación cuanto más alto es el parecido visual o conductual. Una interfaz gestual comprende ambos ámbitos, consiguiendo tanto aumentar la naturalidad como transmitir información no verbal.

Este trabajo ha desarrollado el diseño de una metodología para la anotación gestual en textos con el afán de mejorar la integración de los gestos co-verbales. Se ha logrado configurar esta metodología mediante varios experimentos, en los que se han aplicado diferentes técnicas de procesamiento del lenguaje natural. El método propuesto consiste en un primer filtrado del texto, señalando las palabras más significativas. Para ello, se ha asumido que los nombres, verbos, adjetivos y adverbios son los elementos más influyentes en el significado de una oración. Partiendo de una asignación de términos a cada gesto y definiendo su representación semántica, se ha pretendido encontrar el gesto con el significado más cercano a cada palabra relevante detectada.

Tras el análisis de los resultados de la experimentación, se ha llegado a la conclusión de que los modelos basados en *word embeddings* representan técnicas de comparación semántica más factibles en el ámbito de la robótica, no sólo por su mayor eficacia en la asociación de términos en cada categoría gramatical, sino por el mínimo coste computacional que supone: búsquedas en ficheros y operaciones geométricas. La comparación entre el modelo de datos de *WordNet* y los *word embeddings* supone el estudio de los modelos basados en datos experienciales frente a aquellos generados por datos distributivos. Mientras que los primeros se han diseñado manualmente, representando distintos tipos de datos y su distribución estructural, los *word embeddings* son modelos estadísticos que se establecen de forma automática a partir de una gran cantidad de datos empíricos, generalmente de grandes repositorios online. Por lo tanto, se puede afirmar que los métodos estadísticos muestran mayores capacidades en este contexto debido a que sus propiedades semánticas son descritas en términos de parámetros, al contrario que *WordNet*, en las que se presentan en forma de relaciones distribuidas. El empleo de modelos pre-entrenados elimina el tiempo de entrenamiento del proceso, por lo que la comparación entre las propiedades de dos términos es inmediata; en cambio, *WordNet* requiere algoritmos de búsqueda en taxonomías para navegar entre nodos.

Los resultados también muestran un incremento del porcentaje de acierto al comparar únicamente palabras y términos de la misma categoría gramatical. De ello se deduce que los *word embeddings* tienen cierta dependencia de las categorías gramaticales; esto contrasta con el hecho de que los vectores contienen información sobre el contexto de las palabras, y éste varía en

función de la categoría gramatical. Las métricas de similitud entre nodos de *WordNet* se basan en tres principios diferentes: la *distancia entre conceptos*, la *cantidad de información* y la *similitud entre glosas*. El análisis de los resultados por cada categoría pone de manifiesto la disparidad de las métricas para calcular la similitud. Las métricas que emplean la *distancia entre conceptos* o la *cantidad de información* son apropiadas para los nombres y verbos, mientras que las métricas de *similitud entre glosas* captan mejor la semejanza entre adjetivos y adverbios. En cambio, las métricas geométricas mantienen cierta independencia de la categoría en el cálculo de similitud semántica; específicamente, la similitud coseno consigue resultados tan buenos como las métricas basadas en la *distancia entre conceptos* en nombres y verbos, y las métricas de *similitud entre glosas* en adjetivos y adverbios. A pesar de ello, se aprecia un decremento en la tasa de aciertos de los adverbios en comparación con los nombres.

La metodología resultante se basa inicialmente en un etiquetado gramatical para filtrar todas aquellas palabras no relevantes en el significado. Posteriormente, utiliza la similitud coseno, siendo la métrica con una mejor respuesta en todos los ámbitos, para calcular los gestos más cercanos semánticamente a los nombres, verbos, adjetivos y adverbios, en ese orden. El análisis ha demostrado que la similitud semántica es captada en mayor grado por los nombres, de ahí la priorización. Esta metodología es una buena primera aproximación a la integración de gestos co-verbales desde la semántica. Ejemplo de ello es la aplicación al cuento *Teo el valiente* del algoritmo implementado a partir de las librerías de *FreeLing* para el *POS tagging* y los modelos de *word2vec*. La ejecución del cuento anotado sobre un robot *Nao* deja de manifiesto su utilidad en el campo de las interfaces sociales.

Por otro lado, esta metodología presenta una facilidad para el multiidomado: la introducción de un nuevo idioma depende exclusivamente del uso de una herramienta de etiquetado gramatical acorde a éste y el modelo preentrenado de *word embeddings* correspondiente. Actualmente la herramienta *FreeLing* contempla los siguientes idiomas: inglés, español, francés, italiano, alemán, ruso, noruego, catalán, gallego, croata, esloveno, asturiano y galés. En cuanto a los *word embeddings*, hay una gran disponibilidad de modelos

pre-entrenados en diferentes idiomas (inglés[1], francés[2], alemán[3], español[4], italiano[5]...) y ámbitos (histórico[6], académico[7], *Twitter*[7]...); sin embargo, en todo caso, es posible generar un modelo propio a partir de los datos de cualquier repositorios online como *Wikipedia*.

La representación y similitud semántica de palabras es un campo muy estudiado en el ámbito del procesamiento del lenguaje natural; sin embargo, prácticamente no se ha abordado en la integración de gestos co-verbales en el ámbito *HRI*. Este trabajo constata la importancia de explotar la correlación entre el significado verbal del interlocutor y sus expresiones, independientemente de la existencia de otro tipo de gesticulaciones exentas de base semántica. La metodología diseñada es un primer paso hacia procesos de integración gestual más complejos y avanzados que tengan en cuenta otros factores extralingüísticos.

6.2. Trabajo futuro

Este trabajo arroja una nueva metodología para la integración de gestos a través de la semántica, contribuyendo al estado del arte con una nueva línea de investigación. Además, la factibilidad de esta metodología ha sido validada con su implementación en una herramienta de anotación gestual; y ésta, a su vez, ha sido empleada junto con un robot *Nao* sobre un cuento. Sin embargo, tanto la metodología como la herramienta están lejos de reproducir los mismos mecanismos que el ser humano, por lo que requieren nuevas técnicas para su mejora.

La herramienta de anotación presenta diferentes posibles mejoras para dotar de una mayor naturalidad a los gestos durante la ejecución posterior. Por ejemplo, la fluidez en el habla puede ser incrementada señalando aquellos gestos ya realizados anteriormente para impedir su ejecución completa. Otra posibilidad es integrar un planificador de tiempo, empleado por otras aproximaciones. De esta forma, se realiza en paralelo el cálculo de la duración oral de la sentencia, filtrando los gestos en función del tiempo que

[1] https://code.google.com/archive/p/word2vec
[2] http://fauconnier.github.io/index.html
[3] http://devmount.github.io/GermanWordEmbeddings/
[4] http://crscardellino.me/SBWCE/
[5] http://hlt.isti.cnr.it/wordembeddings/
[6] http://nlp.stanford.edu/projects/histwords/
[7] http://nlp.stanford.edu/projects/glove/

ocupan. Por último, también se puede gestionar el solapamiento de gestos en una misma frase reduciendo la duración de alguno de ellos, e incluso de todos.

En cuanto a la metodología, sería interesante la introducción de una capa heurística para definir distintas reglas. Por ejemplo, es posible configurar una alternancia de prioridades sobre las categorías gramaticales, de forma que no siempre se analicen nombres. Incluso cabría estudiar la competencia ponderada entre las categorías; así, si se detectan varios nombres con un margen muy limitado de similitud y al mismo tiempo un verbo o adjetivo muy significativo, se podría dar prioridad a este último. Por otro lado, se podría modificar el sistema actual que rige la asociación considerando un empate técnico entre los gestos con una similitud similar o cuya métrica responda dentro de un rango. Así se incluiría una mayor variabilidad. En relación con ésta, la memorización de gestos para evitar repeticiones podría ser otra técnica. Como se mostró en el estado del arte, otras aproximaciones emplean parte del contexto de las oraciones para su propio sistema de reglas. En este caso, cabría estudiar la desambiguación semántica mediante el contexto más a largo plazo, evitando así dobles significados. Por último, la inclusión de detectores de negación podría ser una buena técnica para matizar los gestos. Otro posible estudio podría consistir en la combinación probabilística de múltiples métricas para complementar la información semántica. Se podría emplear alguna métrica de *similitud entre glosas* junto con la *similitud coseno* para determinar la pertenencia de adverbios y adjetivos, así como el uso de métricas de *distancia entre conceptos* para los nombres y verbos.

Además de estas reglas heurísticas propuestas en la metodología, se ha planteado cómo futuras líneas de investigación la integración de otras técnicas de procesamiento del lenguaje natural fuera del ámbito semántico. Inicialmente se planteó una mejora de las interfaces sociales asentada sobre tres ámbitos del procesamiento del lenguaje natural: el análisis semántico, el *análisis de sentimiento* y un *análisis del discurso*. Sin embargo, debido a la gran envergadura de la propuesta, este trabajo se ha orientado únicamente hacia el análisis semántico, dejando abiertas las líneas de investigación en dirección a las otras propuestas.

El *análisis de sentimiento* se basa en el estudio del carácter emocional de las oraciones. Las técnicas de este ámbito radican en una clasificación de la polaridad oracional a través de modelos estadísticos y de análisis sintácti-

cos basados en reglas, como los detectores de negación. Su integración en este proyecto arrojaría un matiz relevante a la hora de modificar los gestos para transmitir más o menos efusividad y entusiasmo. Un ejemplo de ello consistiría en ralentizar los movimientos ante frases negativas.

En cuanto al tercer ámbito, está enfocado desde el punto de vista del *análisis del discurso* en base a la *Teoría de la Estructura Retórica* (*Rhetorical Structure Theory*) o *RST* (Bateman y Delin, 2005). Ésta concibe los textos como secuencias lógicas de elementos con funciones específicas que sustentan la coherencia; distintas estructuras relacionadas entre sí, generalmente del estilo núcleo-satélite, conforman las oraciones. Este tipo de gestos es más frecuente en discursos prolongados para evitar la pérdida de atención de los oyentes. De esta forma, al localizar relaciones de causalidad, numeración, justificación o contraste, se podrían activar animaciones de tránsito entre significados.

Bibliografía

Bibliografía

[and others1953] and others. 1953. Remote-control manipulator, Marzo 24. US Patent 2,632,574.

[Andrews, Vigliocco, y Vinson2009] Andrews, Mark, Gabriella Vigliocco, y David Vinson. 2009. Integrating experiential and distributional data to learn semantic representations. *Psychological review*, 116(3):463.

[Asfour et al.2008] Asfour, Tamim, Pedram Azad, Florian Gyarfas, y Rüdiger Dillmann. 2008. Imitation learning of dual-arm manipulation tasks in humanoid robots. *International Journal of Humanoid Robotics*, 5(02):183–202.

[Atkeson y Schaal1997a] Atkeson, Christopher G y Stefan Schaal. 1997a. Learning tasks from a single demonstration. En *Robotics and Automation, 1997. Proceedings., 1997 IEEE International Conference on*, volumen 2, páginas 1706–1712. IEEE.

[Atkeson y Schaal1997b] Atkeson, Christopher G y Stefan Schaal. 1997b. Robot learning from demonstration. En *ICML*, volumen 97, páginas 12–20.

[Ballard et al.2012] Ballard, Leslie Anne, Selma Sabanovic, Jasleen Kaur, y Stasa Milojevic. 2012. George charles devol, jr.[history]. *Ieee Robotics & Automation Magazine*, 19(3):114–119.

[Banerjee y Pedersen2003] Banerjee, Satanjeev y Ted Pedersen. 2003. Extended gloss overlaps as a measure of semantic relatedness. En *Ijcai*, volumen 3, páginas 805–810.

[Bar-Cohen2004] Bar-Cohen, Yoseph. 2004. *Electroactive polymer (EAP) actuators as artificial muscles: reality, potential, and challenges*, volumen 136. SPIE press.

[Baroni y Lenci2010] Baroni, Marco y Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

[Baroni y Zamparelli2010] Baroni, Marco y Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: Representing adjective-noun constructions in semantic space. En *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, páginas 1183–1193. Association for Computational Linguistics.

[Bateman y Delin2005] Bateman, John y Judy Delin. 2005. Rhetorical structure theory.

[Belew2007] Belew, B. 2007. Japan's ifbot understands language and emotional tone, gets rejected. rising sun of nihon.

[Bengio y Heigold2014] Bengio, Samy y Georg Heigold. 2014. Word embeddings for speech recognition. En *INTERSPEECH*, páginas 1053–1057.

[Bennewitz et al.2007] Bennewitz, Maren, Felix Faber, Dominik Joho, y Sven Behnke. 2007. Fritz-a humanoid communication robot. En *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*, páginas 1072–1077. IEEE.

[Benyon y Mival2008] Benyon, David y Oli Mival. 2008. Landscaping personification technologies: from interactions to relationships. En *CHI'08 extended abstracts on Human factors in computing systems*, páginas 3657–3662. ACM.

[Bergmann, Kahl, y Kopp2013] Bergmann, Kirsten, Sebastian Kahl, y Stefan Kopp. 2013. Modeling the semantic coordination of speech and gesture under cognitive and linguistic constraints. En *International Workshop on Intelligent Virtual Agents*, páginas 203–216. Springer.

[Billard2002] Billard, Aude. 2002. Imitation: a means to enhance learning of a synthetic protolanguage in autonomous robots, imitation in animals and artifacts.

[Billingsley, Visala, y Dunn2008] Billingsley, John, Arto Visala, y Mark Dunn. 2008. Robotics in agriculture and forestry. En *Springer handbook of robotics*. Springer, páginas 1065–1077.

[Bird2006] Bird, Steven. 2006. Nltk: the natural language toolkit. En *Proceedings of the COLING/ACL on Interactive presentation sessions*, páginas 69–72. Association for Computational Linguistics.

[Blacoe y Lapata2012] Blacoe, William y Mirella Lapata. 2012. A comparison of vector-based representations for semantic composition. En *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, páginas 546–556. Association for Computational Linguistics.

[Blanco-Fernández et al.2008] Blanco-Fernández, Yolanda, José J Pazos-Arias, Alberto Gil-Solla, Manuel Ramos-Cabrer, Martín López-Nores, Jorge García-Duque, Ana Fernández-Vilas, Rebeca P Díaz-Redondo, y Jesús Bermejo-Muñoz. 2008. A flexible semantic inference methodology to reason about user preferences in knowledge-based recommender systems. *Knowledge-Based Systems*, 21(4):305–320.

[Blei, Ng, y Jordan2003] Blei, David M, Andrew Y Ng, y Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

[Bogue2009] Bogue, Robert. 2009. Exoskeletons and robotic prosthetics: a review of recent developments. *Industrial Robot: An International Journal*, 36(5):421–427.

[Bouma2009] Bouma, Gerlof. 2009. Normalized (pointwise) mutual information in collocation extraction. En *Proceedings of the Biennial GSCL Conference*, volumen 156.

[Breazeal2004] Breazeal, Cynthia. 2004. Social interactions in hri: the robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 34(2):181–186.

[Breazeal y Aryananda2002] Breazeal, Cynthia y Lijin Aryananda. 2002. Recognition of affective communicative intent in robot-directed speech. *Autonomous robots*, 12(1):83–104.

[Breazeal, Hoffman, y Lockerd2004] Breazeal, Cynthia, Guy Hoffman, y Andrea Lockerd. 2004. Teaching and working with robots as a collaboration. En *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*, páginas 1030–1037. IEEE Computer Society.

[Breazeal y others1998] Breazeal, Cynthia y others. 1998. A motivational system for regulating human-robot interaction. En *Aaai/iaai*, páginas 54–61.

[Breazeal y Scassellati2002] Breazeal, Cynthia y Brian Scassellati. 2002. f 4 challenges in building robots that imitate people. *Imitation in animals and artifacts*, 363.

[Briscoe2011] Briscoe, C Ted. 2011. Introduction to formal semantics for natural language.

[Broadbent, Stafford, y MacDonald2009] Broadbent, Elizabeth, Rebecca Stafford, y Bruce MacDonald. 2009. Acceptance of healthcare robots for the older population: review and future directions. *International Journal of Social Robotics*, 1(4):319.

[Broekens, Heerink, y Rosendal2009] Broekens, Joost, Marcel Heerink, y Henk Rosendal. 2009. Assistive social robots in elderly care: a review. *Gerontechnology*, 8(2):94–103.

[Bruce, Nourbakhsh, y Simmons2002] Bruce, Allison, Illah Nourbakhsh, y Reid Simmons. 2002. The role of expressiveness and attention in human-robot interaction. En *Robotics and Automation, 2002. Proceedings. ICRA'02. IEEE International Conference on*, volumen 4, páginas 4138–4142. IEEE.

[Calinon y Billard2006] Calinon, Sylvain y Aude Billard. 2006. Teaching a humanoid robot to recognize and reproduce social cues. En *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, páginas 346–351. IEEE.

[Calinon y Billard2007] Calinon, Sylvain y Aude Billard. 2007. Incremental learning of gestures by imitation in a humanoid robot. En *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, páginas 255–262. ACM.

[Calinon, Guenter, y Billard2007] Calinon, Sylvain, Florent Guenter, y Aude Billard. 2007. On learning, representing, and generalizing a task in a humanoid robot. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 37(2):286–298.

[Cañamero y Fredslund2001] Cañamero, Lola y Jakob Fredslund. 2001. I show you how i like you-can you read it in my face?[robotics]. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and humans*, 31(5):454–459.

[Cannata et al.2008] Cannata, Giorgio, Marco Maggiali, Giorgio Metta, y Giulio Sandini. 2008. An embedded artificial skin for humanoid robots. En *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, páginas 434–438. IEEE.

[Cassell et al.] Cassell, J, T Bickmore, L Campbell, H Vilhjalmsson, y H Yan. Human conversation as a system framework: Designing emobodied conversational agents. *Embodied Conversational Agents*, páginas 29–63.

[Cassell y others2000] Cassell, Justine y others. 2000. Nudge nudge wink wink: Elements of face-to-face conversation for embodied conversational agents. *Embodied conversational agents*, 1.

[Cassell, Vilhjálmsson, y Bickmore2004] Cassell, Justine, Hannes Högni Vilhjálmsson, y Timothy Bickmore. 2004. Beat: the behavior expression animation toolkit. En *Life-Like Characters*. Springer, páginas 163–185.

[Chella y Macaluso2009] Chella, Antonio y Irene Macaluso. 2009. The perception loop in cicerobot, a museum guide robot. *Neurocomputing*, 72(4):760–766.

[Chen y Manning2014] Chen, Danqi y Christopher D Manning. 2014. A fast and accurate dependency parser using neural networks. En *EMNLP*, páginas 740–750.

[Chiu, Morency, y Marsella2015] Chiu, Chung-Cheng, Louis-Philippe Morency, y Stacy Marsella. 2015. Predicting co-verbal gestures: a deep and temporal modeling approach. En *International Conference on Intelligent Virtual Agents*, páginas 152–166. Springer.

[Chu, Kazerooni, y Zoss2005] Chu, Andrew, Hami Kazerooni, y Adam Zoss. 2005. On the biomimetic design of the berkeley lower extremity exoskeleton (bleex). En *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, páginas 4345–4352. IEEE.

[Chu et al.2012] Chu, Won-Shik, Kyung-Tae Lee, Sung-Hyuk Song, Min-Woo Han, Jang-Yeob Lee, Hyung-Soo Kim, Min-Soo Kim, Yong-Jai Park, Kyu-Jin Cho, y Sung-Hoon Ahn. 2012. Review of biomimetic underwater robots using smart actuators. *International journal of precision engineering and manufacturing*, 13(7):1281–1292.

[Collins y Loftus1975] Collins, Allan M y Elizabeth F Loftus. 1975. A spreading-activation theory of semantic processing. *Psychological review*, 82(6):407.

[Collins y Quillian1969] Collins, Allan M y M Ross Quillian. 1969. Retrieval time from semantic memory. *Journal of verbal learning and verbal behavior*, 8(2):240–247.

[Courty y Marchand2003] Courty, Nicolas y Eric Marchand. 2003. Visual perception based on salient features. En *Intelligent Robots and Systems, 2003.(IROS 2003). Proceedings. 2003 IEEE/RSJ International Conference on*, volumen 1, páginas 1024–1029. IEEE.

[Crystal2012] Crystal, David. 2012. *English as a global language*. Cambridge University Press.

[Dahlbäck, Jönsson, y Ahrenberg1993] Dahlbäck, Nils, Arne Jönsson, y Lars Ahrenberg. 1993. Wizard of oz studies-why and how. *Knowledge-based systems*, 6(4):258–266.

[Dardas y Georganas2011] Dardas, Nasser H y Nicolas D Georganas. 2011. Real-time hand gesture detection and recognition using bag-of-features and support vector machine techniques. *IEEE Transactions on Instrumentation and Measurement*, 60(11):3592–3607.

[Dario y Susani1996] Dario, P y G Susani. 1996. Physical and psychological interactions between humans and robots in the home environment. En *Proceedings of the First International Symposium on Humanoid Robots (HURO96), Tokyo, Japan*, páginas 5–16.

[Dario, Guglielmelli, y Laschi2001] Dario, Paolo, Eugenio Guglielmelli, y Cecilia Laschi. 2001. Humanoids and personal robots: Design and experiments. *Journal of Field Robotics*, 18(12):673–690.

[Darwin, Ekman, y Prodger1998] Darwin, Charles, Paul Ekman, y Phillip Prodger. 1998. *The expression of the emotions in man and animals*. Oxford University Press, USA.

[Derimis y Hayes2002] Derimis, Y y G Hayes. 2002. Imitations as a dual-route process featuring predictive and learning components: a biologically plausible computational model. *Imitation in animals and artifacts*, páginas 327–361.

[Devol1948] Devol, George C. 1948. Coaxial line coupling, Septiembre 28. US Patent 2,449,983.

[Díaz, Gil, y Sánchez2011] Díaz, Iñaki, Jorge Juan Gil, y Emilio Sánchez. 2011. Lower-limb robotic rehabilitation: literature review and challenges. *Journal of Robotics*, 2011.

[Dumais et al.1988] Dumais, Susan T, George W Furnas, Thomas K Landauer, Scott Deerwester, y Richard Harshman. 1988. Using latent semantic analysis to improve access to textual information. En *Proceedings of the SIGCHI conference on Human factors in computing systems*, páginas 281–285. ACM.

[Ekman1992] Ekman, Paul. 1992. Are there basic emotions?

[Ekman y Oster1982] Ekman, Paul y Harriet Oster. 1982. Review of research, 1970-1980. *Emotion in the human face*, páginas 147–173.

[Endrass et al.2010] Endrass, Birgit, Ionut Damian, Peter Huber, Matthias Rehm, y Elisabeth André. 2010. Generating culture-specific gestures for virtual agent dialogs. En *International Conference on Intelligent Virtual Agents*, páginas 329–335. Springer.

[Fellbaum1998] Fellbaum, Christiane. 1998. *WordNet*. Wiley Online Library.

[Forlizzi y DiSalvo2006] Forlizzi, Jodi y Carl DiSalvo. 2006. Service robots in the domestic environment: a study of the roomba vacuum in the home.

En *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, páginas 258–265. ACM.

[Foulk2007] Foulk, E. 2007. Lonely robots ignored by elderly ludites. *The New Zealand Herald available from http://www. nzherald. co. nz/section/story. cfm.*

[Francke, Ruiz-del Solar, y Verschae2007] Francke, Hardy, Javier Ruiz-del Solar, y Rodrigo Verschae. 2007. Real-time hand gesture detection and recognition using boosted classifiers and active learning. En *Pacific-Rim Symposium on Image and Video Technology*, páginas 533–547. Springer.

[Frijda y others1994] Frijda, Nico H y others. 1994. Emotions are functional, most of the time.

[Garcia et al.2007] Garcia, Elena, Maria Antonia Jimenez, Pablo Gonzalez De Santos, y Manuel Armada. 2007. The evolution of robotics research. *IEEE Robotics & Automation Magazine*, 14(1):90–103.

[Gennaro y McConnell-Ginet2000] Gennaro, Chierchia y Sally McConnell-Ginet. 2000. Meaning and grammar: An introduction to semantics.

[Geppert2004] Geppert, Linda. 2004. Qrio, the robot that could. *Ieee Spectrum*, 41(5):34–37.

[Germak et al.2015] Germak, Claudio, Maria Luce Lupetti, Luca Giuliano, Kaouk Ng, y Miguel Efrain. 2015. Robots and cultural heritage: New museum experiences. *Journal of Science and Technology of the Arts*, 7(2):47–57.

[Gonzalez-Agirre, Laparra, y Rigau2012] Gonzalez-Agirre, Aitor, Egoitz Laparra, y German Rigau. 2012. Multilingual central repository version 3.0. En *LREC*, páginas 2525–2529.

[Gouaillier et al.2008] Gouaillier, David, Vincent Hugel, Pierre Blazevic, Chris Kilner, Jerome Monceaux, Pascal Lafourcade, Brice Marnier, Julien Serre, y Bruno Maisonnier. 2008. The nao humanoid: a combination of performance and affordability. *CoRR abs/0807.3223*.

[Graf y Staab2009] Graf, Birgit y Harald Staab. 2009. Service robots and automation for the disabled/limited. En *Springer Handbook of Automation*. Springer, páginas 1485–1502.

[Gross et al.2009] Gross, H-M, H Boehme, Ch Schroeter, Steffen Müller, Alexander König, Erik Einhorn, Ch Martin, Matthias Merten, y Andreas Bley. 2009. Toomas: interactive shopping guide robots in everyday use-final implementation and experiences from long-term field trials. En *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, páginas 2005–2012. IEEE.

[Haasch et al.2004] Haasch, Axel, Sascha Hohenner, Sonja Hüwel, Marcus Kleinehagenbrock, Sebastian Lang, Ioannis Toptsis, Gernot A Fink, Jannik Fritsch, Britta Wrede, y Gerhard Sagerer. 2004. Biron–the bielefeld robot companion. En *Proc. Int. Workshop on Advances in Service Robotics*, páginas 27–32. Stuttgart, Germany.

[Han et al.2005] Han, Shihui, Yi Jiang, Glyn W Humphreys, Tiangang Zhou, y Peng Cai. 2005. Distinct neural substrates for the perception of real and virtual visual worlds. *NeuroImage*, 24(3):928–935.

[Hara y Kobayashi1996] Hara, Fumio y Hiroshi Kobayashi. 1996. A face robot able to recognize and produce facial expression. En *Intelligent Robots and Systems'96, IROS 96, Proceedings of the 1996 IEEE/RSJ International Conference on*, volumen 3, páginas 1600–1607. IEEE.

[Hara et al.1998] Hara, Fumio, Hiroshi Kobayashi, Fumiya Iida, y Massoki Tabata. 1998. Personality characterization of animate face robot through interactive communication with human. *Proceedings of the 1998 International Advanced Robotics Program (IARP98), Tsukuba, Japan, pp. IV–1.*

[Harris1954] Harris, Zellig S. 1954. Distributional structure. *Word*, 10(2-3):146–162.

[Hartmann, Mancini, y Pelachaud2005] Hartmann, Björn, Maurizio Mancini, y Catherine Pelachaud. 2005. Implementing expressive gesture synthesis for embodied conversational agents. En *International Gesture Workshop*, páginas 188–199. Springer.

[Hato et al.2010] Hato, Yasuhiko, Satoru Satake, Takayuki Kanda, Michita Imai, y Norihiro Hagita. 2010. Pointing to space: modeling of deictic interaction referring to regions. En *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, páginas 301–308. IEEE.

[Hirst, St-Onge, y others1998] Hirst, Graeme, David St-Onge, y others. 1998. Lexical chains as representations of context for the detection and correction of malapropisms. *WordNet: An electronic lexical database*, 305:305–332.

[Hofmann1999] Hofmann, Thomas. 1999. Probabilistic latent semantic indexing. En *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 50–57. ACM.

[Itoh et al.2004] Itoh, Kazuko, Hiroyasu Miwa, Munemichi Matsumoto, Massimiliano Zecca, Hideaki Takanobu, Stefano Roccella, Maria Chiara Carrozza, Paolo Dario, y Atsuo Takanishi. 2004. Various emotional expressions with emotion expression humanoid robot we-4rii. En *Robotics and Automation, 2004. TExCRA'04. First IEEE Technical Exhibition Based Conference on*, páginas 35–36. IEEE.

[Izard2001] Izard, C. 2001. Human emotions,(1977).

[Izard1993] Izard, Carroll E. 1993. Four systems for emotion activation: cognitive and noncognitive processes. *Psychological review*, 100(1):68.

[Jayawardena et al.2010] Jayawardena, Chandimal, I Han Kuo, Ulrike Unger, Aleksandar Igic, Richie Wong, Catherine I Watson, RQ Stafford, Elizabeth Broadbent, Priyesh Tiwari, Jim Warren, y others. 2010. Deployment of a service robot to help older people. En *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, páginas 5990–5995. IEEE.

[Jenkins et al.2007] Jenkins, Marie-Claire, Richard Churchill, Stephen Cox, y Dan Smith. 2007. Analysis of user interaction with service oriented chatbot systems. En *International Conference on Human-Computer Interaction*, páginas 76–83. Springer.

[Jiang y Conrath1997] Jiang, Jay J y David W Conrath. 1997. Semantic similarity based on corpus statistics and lexical taxonomy. *arXiv preprint cmp-lg/9709008*.

[Kajita et al.2003] Kajita, Shuuji, Fumio Kanehiro, Kenji Kaneko, Kiyoshi Fujiwara, Kensuke Harada, Kazuhito Yokoi, y Hirohisa Hirukawa.

2003. Biped walking pattern generation by using preview control of zero-moment point. En *Robotics and Automation, 2003. Proceedings. ICRA'03. IEEE International Conference on*, volumen 2, páginas 1620–1626. IEEE.

[Kaneko et al.2008] Kaneko, Kenji, Kensuke Harada, Fumio Kanehiro, Go Miyamori, y Kazuhiko Akachi. 2008. Humanoid robot hrp-3. En *Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on*, páginas 2471–2478. IEEE.

[Kang et al.2011] Kang, HeeSu, Kiwon Rhee, Kyoung-jin You, y Hyun-chool Shin. 2011. Intuitive robot navigation using wireless emg and acceleration sensors on human arm. En *Intelligent Signal Processing and Communications Systems (ISPACS), 2011 International Symposium on*, páginas 1–4. IEEE.

[Kawamura et al.1996] Kawamura, Kazuhiko, D Mitch Wilkes, Todd Pack, Magued Bishay, y Jason Barile. 1996. Humanoids: Future robots for home and factory. En *International symposium on humanoid robots*, páginas 53–62.

[Kendon1986] Kendon, Adam. 1986. Current issues in the study of gesture. *The biological foundations of gestures: Motor and semiotic aspects*, 1:23–47.

[Kendon2004] Kendon, Adam. 2004. *Gesture: Visible action as utterance*. Cambridge University Press.

[Kiesler y Goetz2002] Kiesler, Sara y Jennifer Goetz. 2002. Mental models of robotic assistants. En *CHI'02 extended abstracts on Human Factors in Computing Systems*, páginas 576–577. ACM.

[Kiguchi y Fukuda2004] Kiguchi, Kazuo y Toshio Fukuda. 2004. A 3 dof exoskeleton for upper limb motion assist: Consideration of the effect of bi-articular muscles. En *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, volumen 3, páginas 2424–2429. IEEE.

[Kilgarriff y Fellbaum2000] Kilgarriff, Adam y Christiane Fellbaum. 2000. Wordnet: An electronic lexical database.

[Kim, Kwak, y Kim2008] Kim, Heeyoung, Sonya S Kwak, y Myungsuk Kim. 2008. Personality design of sociable robots by control of gesture design factors. En *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on*, páginas 494–499. IEEE.

[Klein et al.2015] Klein, Benjamin, Guy Lev, Gil Sadeh, y Lior Wolf. 2015. Associating neural word embeddings with deep image representations using fisher vectors. En *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, páginas 4437–4446.

[Komura et al.2005] Komura, Taku, Howard Leung, Shunsuke Kudoh, y James Kuffner. 2005. A feedback controller for biped humanoids that can counteract large perturbations during gait. En *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, páginas 1989–1995. IEEE.

[Kopp y Wachsmuth2004] Kopp, Stefan y Ipke Wachsmuth. 2004. Synthesizing multimodal utterances for conversational agents. *Computer animation and virtual worlds*, 15(1):39–52.

[Kozima1994] Kozima, Hideki. 1994. *Computing lexical cohesion as a tool for text analysis*. Ph.D. tesis, Ph. D. thesis, University of Electro-Communications.

[Kozima, Michalowski, y Nakagawa2009] Kozima, Hideki, Marek P Michalowski, y Cocoro Nakagawa. 2009. Keepon. *International Journal of Social Robotics*, 1(1):3–18.

[Landauer y Dumais1997] Landauer, Thomas K y Susan T Dumais. 1997. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological review*, 104(2):211.

[Lanfranco et al.2004] Lanfranco, Anthony R, Andres E Castellanos, Jaydev P Desai, y William C Meyers. 2004. Robotic surgery: a current perspective. *Annals of surgery*, 239(1):14–21.

[Leacock y Chodorow1998] Leacock, Claudia y Martin Chodorow. 1998. Combining local context and wordnet similarity for word sense identification. *WordNet: An electronic lexical database*, 49(2):265–283.

[Lee y Kim1999] Lee, Hyeon-Kyu y Jin-Hyung Kim. 1999. An hmm-based threshold model approach for gesture recognition. *IEEE Transactions on pattern analysis and machine intelligence*, 21(10):961–973.

[Levenson1994] Levenson, RW. 1994. Human emotion: A functional view. in. p. ekman & rj davidson (eds.), the nature of emotion: Fundamental questions (pp. 123–126). *Hillsdale, NJ: Erlbaum. Levenson, RW (1999). Intrapersonal functions of emotion. Cognition and Emotion*, 13:481504.

[Levine, Theobalt, y Koltun2009] Levine, Sergey, Christian Theobalt, y Vladlen Koltun. 2009. Real-time prosody-driven synthesis of body language. En *ACM Transactions on Graphics (TOG)*, volumen 28, página 172. ACM.

[Levy y Goldberg2014] Levy, Omer y Yoav Goldberg. 2014. Neural word embedding as implicit matrix factorization. En *Advances in neural information processing systems*, páginas 2177–2185.

[Li, Bandar, y McLean2003] Li, Yuhua, Zuhair A Bandar, y David McLean. 2003. An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions on knowledge and data engineering*, 15(4):871–882.

[Lin y others1998] Lin, Dekang y others. 1998. An information-theoretic definition of similarity. En *ICML*, volumen 98, páginas 296–304. Citeseer.

[Liu, Joty, y Meng2015] Liu, Pengfei, Shafiq R Joty, y Helen M Meng. 2015. Fine-grained opinion mining with recurrent neural networks and word embeddings. En *EMNLP*, páginas 1433–1443.

[Loper et al.2009] Loper, Matthew M, Nathan P Koenig, Sonia H Chernova, Chris V Jones, y Odest C Jenkins. 2009. Mobile human-robot teaming with environmental tolerance. En *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, páginas 157–164. ACM.

[Mataric2000] Mataric, Maja J. 2000. Getting humanoids to move and imitate. *IEEE Intelligent Systems and their Applications*, 15(4):18–24.

[McNeill1992] McNeill, David. 1992. *Hand and mind: What gestures reveal about thought*. University of Chicago press.

[McNeill2008] McNeill, David. 2008. *Gesture and thought*. University of Chicago press.

[McNeill y Levy1980] McNeill, David y Elena Levy. 1980. *Conceptual representations in language activity and gesture*. ERIC Clearinghouse.

[Meena, Jokinen, y Wilcock2012] Meena, Raveesh, Kristiina Jokinen, y Graham Wilcock. 2012. Integration of gestures and speech in human-robot interaction. En *Cognitive Infocommunications (CogInfoCom), 2012 IEEE 3rd International Conference on*, páginas 673–678. IEEE.

[Mikolov et al.2013] Mikolov, Tomas, Ilya Sutskever, Kai Chen, Greg S Corrado, y Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. En *Advances in neural information processing systems*, páginas 3111–3119.

[Minato et al.2004] Minato, Takashi, Michihiro Shimada, Hiroshi Ishiguro, y Shoji Itakura. 2004. Development of an android robot for studying human-robot interaction. En *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, páginas 424–434. Springer.

[Mitra y Acharya2007] Mitra, Sushmita y Tinku Acharya. 2007. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 37(3):311–324.

[Mnih y Kavukcuoglu2013] Mnih, Andriy y Koray Kavukcuoglu. 2013. Learning word embeddings efficiently with noise-contrastive estimation. En *Advances in neural information processing systems*, páginas 2265–2273.

[Morea1992] Morea, Saverio F. 1992. The lunar roving vehicle: Historical perspective. En *Lunar Bases and Space Activities of the 21st Century*.

[Mori, MacDorman, y Kageki2012] Mori, Masahiro, Karl F MacDorman, y Norri Kageki. 2012. The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2):98–100.

[Mubin et al.2013] Mubin, Omar, Catherine J Stevens, Suleman Shahid, Abdullah Al Mahmud, y Jian-Jie Dong. 2013. A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, 1:209–0015.

[Mukai et al.2008] Mukai, Toshiharu, Masaki Onishi, Tadashi Odashima, Shinya Hirano, y Zhiwei Luo. 2008. Development of the tactile sensor system of a human-interactive robot ri-man". *IEEE Transactions on Robotics*, 24(2):505–512.

[Murthy y Jadon2009] Murthy, GRS y RS Jadon. 2009. A review of vision based hand gestures recognition. *International Journal of Information Technology and Knowledge Management*, 2(2):405–410.

[Nakaoka et al.2007] Nakaoka, Shiníchiro, Atsushi Nakazawa, Fumio Kanehiro, Kenji Kaneko, Mitsuharu Morisawa, Hirohisa Hirukawa, y Katsushi Ikeuchi. 2007. Learning from observation paradigm: Leg task models for enabling a biped humanoid robot to imitate human dances. *The International Journal of Robotics Research*, 26(8):829–844.

[Nam, Wohn, y others1996] Nam, Yanghee, K Wohn, y others. 1996. Recognition of space-time hand-gestures using hidden markov model. En *ACM symposium on Virtual reality software and technology*, páginas 51–58.

[Nam et al.2014] Nam, Yunjun, Bonkon Koo, Andrzej Cichocki, y Seungjin Choi. 2014. Gom-face: Gkp, eog, and emg-based multimodal interface with application to humanoid robot control. *IEEE Transactions on Biomedical Engineering*, 61(2):453–462.

[Narahara y Maeno2007] Narahara, Hisayuki y Takashi Maeno. 2007. Factors of gestures of robots for smooth communication with humans. En *Proceedings of the 1st international conference on Robot communication and coordination*, página 44. IEEE Press.

[Neff et al.2008] Neff, Michael, Michael Kipp, Irene Albrecht, y Hans-Peter Seidel. 2008. Gesture modeling and animation based on a probabilistic re-creation of speaker style. *ACM Transactions on Graphics (TOG)*, 27(1):5.

[Nehaniv et al.2005] Nehaniv, Chrystopher L, Kerstin Dautenhahn, Jens Kubacki, Martin Haegele, Christopher Parlitz, y Rachid Alami. 2005. A methodological approach relating the classification of gesture to identification of human intent in the context of human-robot interaction. En *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, páginas 371–377. IEEE.

[Nespoulous y Lecours1986] Nespoulous, Jean-Luc y A Roch Lecours. 1986. Gestures: Nature and function. *The biological foundations of gestures: motor and semiotic aspects*, páginas 49–62.

[Ng-Thow-Hing, Luo, y Okita2010] Ng-Thow-Hing, Victor, Pengcheng Luo, y Sandra Okita. 2010. Synchronized gesture and speech production for humanoid robots. En *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*, páginas 4617–4624. IEEE.

[Nickel y Stiefelhagen2003] Nickel, Kai y Rainer Stiefelhagen. 2003. Pointing gesture recognition based on 3d-tracking of face, hands and head orientation. En *Proceedings of the 5th international conference on Multimodal interfaces*, páginas 140–146. ACM.

[Nickel y Stiefelhagen2007] Nickel, Kai y Rainer Stiefelhagen. 2007. Visual recognition of pointing gestures for human–robot interaction. *Image and vision computing*, 25(12):1875–1884.

[Niewiadomski et al.2009] Niewiadomski, Radoslaw, Elisabetta Bevacqua, Maurizio Mancini, y Catherine Pelachaud. 2009. Greta: an interactive expressive eca system. En *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, páginas 1399–1400. International Foundation for Autonomous Agents and Multiagent Systems.

[Niles y Pease2001] Niles, Ian y Adam Pease. 2001. Towards a standard upper ontology. En *Proceedings of the international conference on Formal Ontology in Information Systems-Volume 2001*, páginas 2–9. ACM.

[Ozkil et al.2009] Ozkil, Ali Gurcan, Zhun Fan, Steen Dawids, Henrik Aanes, Jens Klestrup Kristensen, y Kim Hardam Christensen. 2009. Service robots for hospitals: A case study of transportation tasks in a hospital. En *Automation and Logistics, 2009. ICAL'09. IEEE International Conference on*, páginas 289–294. IEEE.

[Palangi et al.2016] Palangi, Hamid, Li Deng, Yelong Shen, Jianfeng Gao, Xiaodong He, Jianshu Chen, Xinying Song, y Rabab Ward. 2016. Deep sentence embedding using long short-term memory networks: Analysis and application to information retrieval. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP)*, 24(4):694–707.

[Patwardhan2003] Patwardhan, Siddharth. 2003. *Incorporating dictionary and corpus information into a context vector measure of semantic relatedness*. Ph.D. tesis, University of Minnesota, Duluth.

[Patwardhan, Banerjee, y Pedersen2003] Patwardhan, Siddharth, Satanjeev Banerjee, y Ted Pedersen. 2003. Using measures of semantic relatedness for word sense disambiguation. En *International Conference on Intelligent Text Processing and Computational Linguistics*, páginas 241–257. Springer.

[Pedersen, Patwardhan, y Michelizzi2004] Pedersen, Ted, Siddharth Patwardhan, y Jason Michelizzi. 2004. Wordnet:: Similarity: measuring the relatedness of concepts. En *Demonstration papers at HLT-NAACL 2004*, páginas 38–41. Association for Computational Linguistics.

[Pennington, Socher, y Manning2014] Pennington, Jeffrey, Richard Socher, y Christopher D Manning. 2014. Glove: Global vectors for word representation. En *EMNLP*, volumen 14, páginas 1532–1543.

[Perani et al.2001] Perani, Daniela, Ferruccio Fazio, Nunzio Alberto Borghese, Marco Tettamanti, Stefano Ferrari, Jean Decety, y Maria Carla Gilardi. 2001. Different brain correlates for watching real and virtual hand actions. *Neuroimage*, 14(3):749–758.

[Plutchik1991] Plutchik, Robert. 1991. *The emotions*. University Press of America.

[Quek1994] Quek, Francis KH. 1994. Toward a vision-based hand gesture interface. En *Virtual Reality Software and Technology Conference*, volumen 94, páginas 17–29.

[Quek1995] Quek, Francis KH. 1995. Eyes in the interface. *Image and vision computing*, 13(6):511–525.

[Rada et al.1989] Rada, Roy, Hafedh Mili, Ellen Bicknell, y Maria Blettner. 1989. Development and application of a metric on semantic nets. *IEEE transactions on systems, man, and cybernetics*, 19(1):17–30.

[Ramamoorthy et al.2003] Ramamoorthy, Aditya, Namrata Vaswani, Santanu Chaudhury, y Subhashis Banerjee. 2003. Recognition of dynamic hand gestures. *Pattern Recognition*, 36(9):2069–2081.

[Reeves y Nass1996] Reeves, B y C Nass. 1996. The media equation. csli.

[Resnik1995] Resnik, Philip. 1995. Using information content to evaluate semantic similarity in a taxonomy. *arXiv preprint cmp-lg/9511007*.

[Rickel y Johnson2000] Rickel, Jeff y W Lewis Johnson. 2000. Task-oriented collaboration with embodied agents in virtual worlds. *Embodied conversational agents*, páginas 95–122.

[Riek et al.2010] Riek, Laurel D, Tal-Chen Rabinowitch, Paul Bremner, Anthony G Pipe, Mike Fraser, y Peter Robinson. 2010. Cooperative gestures: Effective signaling for humanoid robots. En *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on*, páginas 61–68. IEEE.

[Riek et al.2009a] Riek, Laurel D, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, y Peter Robinson. 2009a. Empathizing with robots: Fellow feeling along the anthropomorphic spectrum. En *Affective Computing and Intelligent Interaction and Workshops, 2009. ACII 2009. 3rd International Conference on*, páginas 1–6. IEEE.

[Riek et al.2009b] Riek, Laurel D, Tal-Chen Rabinowitch, Bhismadev Chakrabarti, y Peter Robinson. 2009b. How anthropomorphism affects empathy toward robots. En *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, páginas 245–246. ACM.

[Rogalla et al.2002] Rogalla, O, M Ehrenmann, R Zollner, R Becher, y R Dillmann. 2002. Using gesture and speech control for commanding a robot assistant. En *Robot and Human Interactive Communication, 2002. Proceedings. 11th IEEE International Workshop on*, páginas 454–459. IEEE.

[Roussel, Canudas-de Wit, y Goswami1998] Roussel, L, C Canudas-de Wit, y Ambarish Goswami. 1998. Generation of energy optimal complete gait cycles for biped robots. En *Robotics and Automation, 1998. Proceedings. 1998 IEEE International Conference on*, volumen 3, páginas 2036–2041. IEEE.

[Sakagami et al.2002] Sakagami, Yoshiaki, Ryujin Watanabe, Chiaki Aoyama, Shinichi Matsunaga, Nobuo Higaki, y Kikuo Fujimura. 2002. The intelligent asimo: System overview and integration. En *Intelligent Robots*

and Systems, 2002. IEEE/RSJ International Conference on, volumen 3, páginas 2478–2483. IEEE.

[Salem et al.2009] Salem, Maha, Stefan Kopp, Ipke Wachsmuth, y Frank Joublin. 2009. Towards meaningful robot gesture. En *Human Centered Robot Systems*. Springer, páginas 173–182.

[Salem et al.2010] Salem, Maha, Stefan Kopp, Ipke Wachsmuth, y Frank Joublin. 2010. Towards an integrated model of speech and gesture production for multi-modal robot behavior. En *RO-MAN, 2010 IEEE*, páginas 614–619. IEEE.

[Sales et al.2016] Sales, Jorge, Jose V Martí, Raúl Marín, Enric Cervera, y Pedro J Sanz. 2016. Comparob: The shopping cart assistance robot. *International Journal of Distributed Sensor Networks*.

[Sauppé y Mutlu2014] Sauppé, Allison y Bilge Mutlu. 2014. Robot deictics: How gesture and context shape referential communication. En *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, páginas 342–349. ACM.

[Scassellati, Admoni, y Matarić2012] Scassellati, Brian, Henny Admoni, y Maja Matarić. 2012. Robots for use in autism research. *Annual review of biomedical engineering*, 14:275–294.

[Schaal1999] Schaal, Stefan. 1999. Is imitation learning the route to humanoid robots? *Trends in cognitive sciences*, 3(6):233–242.

[Schaal y others1997] Schaal, Stefan y others. 1997. Learning from demonstration. *Advances in neural information processing systems*, páginas 1040–1046.

[Scheeff y others2000] Scheeff, M y others. 2000. Experiences with sparky, a social robot, in proceedings of the workshop on interactive robotics and entertainment (wire-2000).

[Scherer1994] Scherer, Klaus Rainer. 1994. *Evidence for both universality and cultural specificity of emotion elicitation.*

[Schraft1994] Schraft, Rolf Dieter. 1994. Mechatronics and robotics for service applications. *IEEE Robotics & Automation Magazine*, 1(4):31–35.

[Scott et al.2015] Scott, Gregory P, C Glen Henshaw, Ian D Walker, y Bryan Willimon. 2015. Autonomous robotic refueling of an unmanned surface vehicle in varying sea states. En *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, páginas 1664–1671. IEEE.

[Shahinpoor et al.1998] Shahinpoor, Mohsen, Yoseph Bar-Cohen, JO Simpson, y J Smith. 1998. Ionic polymer-metal composites (ipmcs) as biomimetic sensors, actuators and artificial muscles-a review. *Smart materials and structures*, 7(6):R15.

[Shima2014] Shima, Hideki. 2014. Ws4j-wordnet similarity for java.

[Springer2013] Springer, Paul J. 2013. *Military robots and drones: a reference handbook*. ABC-CLIO.

[Srihari, Zhang, y Rao2000] Srihari, Rohini K, Zhongfei Zhang, y Aibing Rao. 2000. Intelligent indexing and semantic retrieval of multimodal documents. *Information Retrieval*, 2(2-3):245–275.

[Stiefelhagen et al.2004] Stiefelhagen, Rainer, C Fugen, R Gieselmann, Hartwig Holzapfel, Kai Nickel, y Alex Waibel. 2004. Natural human-robot interaction using speech, head pose and gestures. En *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volumen 3, páginas 2422–2427. IEEE.

[Stone et al.2004] Stone, Matthew, Doug DeCarlo, Insuk Oh, Christian Rodriguez, Adrian Stere, Alyssa Lees, y Chris Bregler. 2004. Speaking with hands: Creating animated conversational characters from recordings of human performance. En *ACM Transactions on Graphics (TOG)*, volumen 23, páginas 506–513. ACM.

[Sugiyama et al.2007] Sugiyama, Osamu, Takayuki Kanda, Michita Imai, Hiroshi Ishiguro, y Norihiro Hagita. 2007. Natural deictic communication with humanoid robots. En *Intelligent robots and systems, 2007. IROS 2007. IEEE/RSJ international conference on*, páginas 1441–1448. IEEE.

[Takanobu et al.1998] Takanobu, H, A Takanishi, S Hirano, I Kato, K Sato, y T Umetsu. 1998. Development of humanoid robot heads for natural human-robot communication. En *Proceedings of HURO98*, páginas 21–28.

[Takeuchi y Nagao1993] Takeuchi, Akikazu y Katashi Nagao. 1993. Communicative facial displays as a new conversational modality. En *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, páginas 187–193. ACM.

[Takubo, Inoue, y Arai2005] Takubo, Tomohito, Kenji Inoue, y Tatsuo Arai. 2005. Pushing an object considering the hand reflect forces by humanoid robot in dynamic walking. En *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*, páginas 1706–1711. IEEE.

[Tang et al.2014] Tang, Duyu, Furu Wei, Nan Yang, Ming Zhou, Ting Liu, y Bing Qin. 2014. Learning sentiment-specific word embedding for twitter sentiment classification. En *ACL (1)*, páginas 1555–1565.

[Tepper, Kopp, y Cassell2004] Tepper, Paul, Stefan Kopp, y Justine Cassell. 2004. Content in context: Generating language and iconic gesture without a gestionary. En *Proceedings of the Workshop on Balanced Perception and Action in ECAs at AAMAS*, volumen 4, página 8.

[Tresadern y Reid2004] Tresadern, Phil y Ian Reid. 2004. Uncalibrated and unsynchronized human motion capture: A stereo factorization approach. En *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volumen 1, páginas I–I. IEEE.

[Upadek2010] Upadek, Christine. 2010. Socially interactive robots.

[Valin, Michaud, y Rouat2007] Valin, Jean-Marc, François Michaud, y Jean Rouat. 2007. Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. *Robotics and Autonomous Systems*, 55(3):216–228.

[Valin et al.2007] Valin, Jean-Marc, Shuníchi Yamamoto, Jean Rouat, François Michaud, Kazuhiro Nakadai, y Hiroshi G Okuno. 2007. Robust recognition of simultaneous speech by a mobile robot. *IEEE Transactions on Robotics*, 23(4):742–752.

[Vossen1998] Vossen, Piek. 1998. *A multilingual database with lexical semantic networks*. Springer.

[Vukobratovic, Frank, y Juricic1970] Vukobratovic, Miomir, AA Frank, y Davor Juricic. 1970. On the stability of biped locomotion. *IEEE Transactions on Biomedical Engineering*, (1):25–36.

[Vulić y Moens2015] Vulić, Ivan y Marie-Francine Moens. 2015. Monolingual and cross-lingual information retrieval models based on (bilingual) word embeddings. En *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, páginas 363–372. ACM.

[Wang, Yang, y Xie2012] Wang, Baocheng, Chenguang Yang, y Qing Xie. 2012. Human-machine interfaces based on emg and kinect applied to teleoperation of a mobile humanoid robot. En *Intelligent Control and Automation (WCICA), 2012 10th World Congress on*, páginas 3903–3908. IEEE.

[Wang et al.2015] Wang, Xin, Yuanchao Liu, Chengjie Sun, Baoxun Wang, y Xiaolong Wang. 2015. Predicting polarities of tweets by composing word embeddings with long short-term memory. En *ACL (1)*, páginas 1343–1353.

[Wu, Fassert, y Rigaud2012] Wu, Ya-Huei, Christine Fassert, y Anne-Sophie Rigaud. 2012. Designing robots for the elderly: appearance issue and beyond. *Archives of gerontology and geriatrics*, 54(1):121–126.

[Wu y Palmer1994] Wu, Zhibiao y Martha Palmer. 1994. Verbs semantics and lexical selection. En *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, páginas 133–138. Association for Computational Linguistics.

[Yamamoto et al.2006] Yamamoto, Shuníchi, Kazuhiro Nakadai, Mikio Nakano, Hiroshi Tsujino, Jean-Marc Valin, Kazunori Komatani, Tetsuya Ogata, y Hiroshi G Okuno. 2006. Real-time robot audition system that recognizes simultaneous speech in the real world. En *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, páginas 5333–5338. IEEE.

[Yamazaki et al.2009] Yamazaki, Keiichi, Akiko Yamazaki, Mai Okada, Yoshinori Kuno, Yoshinori Kobayashi, Yosuke Hoshi, Karola Pitsch, Paul Luff, Dirk vom Lehn, y Christian Heath. 2009. Revealing gauguin:

engaging visitors in robot guide's explanation in an art museum. En *Proceedings of the SIGCHI conference on human factors in computing systems*, páginas 1437–1446. ACM.

[Yip y Niemeyer2015] Yip, Michael C y Günter Niemeyer. 2015. High-performance robotic muscles from conductive nylon sewing thread. En *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, páginas 2313–2318. IEEE.

[Zheng y Meng2012] Zheng, Minhua y Max Q-H Meng. 2012. Designing gestures with semantic meanings for humanoid robot. En *Robotics and Biomimetics (ROBIO), 2012 IEEE International Conference on*, páginas 287–292. IEEE.

[Zou et al.2013] Zou, Will Y, Richard Socher, Daniel M Cer, y Christopher D Manning. 2013. Bilingual word embeddings for phrase-based machine translation. En *EMNLP*, páginas 1393–1398.