

# Estado del arte del proyecto Detección de humor en textos en español

PROYECTO DE GRADO

Santiago CASTRO  
Matías CUBERO

*Tutores:*

Dr. Guillermo MONCECCHI  
MSc. Diego GARAT



Universidad de la República  
Facultad de Ingeniería

Abril de 2015



# Índice general

<b>1. Introducción</b>	<b>3</b>
<b>2. Humor</b>	<b>4</b>
2.1. Definición . . . . .	4
2.2. Teoría de la superioridad . . . . .	4
2.3. Teoría del alivio . . . . .	5
2.3.1. Teoría de Minsky . . . . .	5
2.4. Teoría de la resolución de incongruencias . . . . .	6
2.4.1. Desambiguación sorpresiva . . . . .	6
2.4.2. Modelo de dos fases . . . . .	7
2.5. Teoría de la violación . . . . .	8
2.6. Teorías sociológicas . . . . .	9
2.6.1. Teorías de mantenimiento . . . . .	9
2.6.2. Teorías de negociación . . . . .	9
2.6.3. Teorías de contextos . . . . .	10
2.7. Humor según la Lingüística . . . . .	10
2.7.1. Teoría de los guiones semánticos del humor . . . . .	10
2.7.2. Teoría general del humor verbal . . . . .	10
2.7.3. Pragmática . . . . .	11
<b>3. Detección de humor</b>	<b>12</b>
3.1. Características específicas para el humor . . . . .	12
3.1.1. Aliteración . . . . .	12
3.1.2. Ambigüedad . . . . .	13
3.1.3. Antonimia . . . . .	13
3.1.4. Centrado en personas . . . . .	14
3.1.5. Jerga sexual . . . . .	14
3.1.6. Negatividad . . . . .	14
3.1.7. Palabras clave . . . . .	15
3.1.8. Perplejidad del modelo de lenguaje . . . . .	15
3.2. Evaluación de los trabajos . . . . .	15
<b>4. Herramientas</b>	<b>17</b>
4.1. Corpus . . . . .	17
4.2. Otros recursos . . . . .	17
4.2.1. NLTK . . . . .	18
4.2.2. Wordnet . . . . .	18
4.2.3. FreeLing . . . . .	18

4.2.4. scikit-learn . . . . .	18
4.2.5. Apache OpenNLP . . . . .	19
<b>5. Resumen</b>	<b>20</b>

# Capítulo 1

## Introducción

El humor es un área hasta el momento poco explorada desde el punto de vista computacional, en especial desde el enfoque del Procesamiento de Lenguaje Natural. El humano, como ser social, convive a diario con el humor y por tanto se convierte en un tema a estudiar de fundamental interés para las Ciencias de la Computación.

En el presente documento se pretende abordar el estado del arte de la Detección de Humor para luego poder construir un clasificador binario supervisado y estadístico. Para esto se investigan diferentes trabajos relativos al área Humor Computacional. Primero se analiza el humor como concepto, introduciendo principalmente definiciones, teorías y puntos de vista. Luego se describen ciertas características intensamente relacionadas con el humor, que fueron sugeridas por distintos autores para ser utilizadas en un clasificador de humor. Por último se desarrollan temas sobre corpus en español sobre el área tratada, así como otros recursos disponibles que sirven de ayuda para el trabajo actual.

## Capítulo 2

# Humor

Es fundamental estudiar y entender el humor para poder trabajar con él e intentar detectarlo en textos. Se aborda la temática pasando por varios puntos y buscando diversidad de opiniones, pero sin profundizar mucho en cada aspecto.

### 2.1. Definición

Según el Diccionario de la Real Academia Española (2001) (DRAE), *humorismo* es un “modo de presentar, enjuiciar o comentar la realidad, resaltando el lado cómico, risueño o ridículo de las cosas”. Este documento se basa únicamente en aquel humor que pueda ser expresado por escrito. Por *chiste* se entiende por “dicho u ocurrencia aguda y graciosa”, aunque en este documento se usa esta palabra y la palabra *humor* de manera intercambiada, más allá que no signifiquen lo mismo. Cabe destacar también que *humor* puede causar *risa* pero no siempre.

Suelen confundirse los términos *ironía*, *sátira*, *sarcasmo* y *humor*. La *ironía*, a diferencia de los otros conceptos, según la DRAE es una “figura retórica que consiste en dar a entender lo contrario de lo que se dice”. *Sátira* en cambio es una “composición en verso o prosa cuyo objeto es censurar o ridiculizar a alguien o algo”. A esta última se la puede ver como forma de humor, aunque tiene un fin crítico, a diferencia del humor que tiene un fin cómico (causar risa). El *sarcasmo* es una “burla sangrienta, ironía mordaz y cruel con que se ofende o maltrata a alguien o algo”, teniendo un objetivo de burla. Muchas veces se mezcla al *ingenio* también en la confusión ya que varias figuras humorísticas que se escuchan frecuentemente vienen acompañadas de él, pero son obviamente conceptos distintos.

Existen varias teorías que explican qué es el humor. Se pueden clasificar en tres tipos según Raskin (1985): cognitivas-perceptivas, de comportamiento social o psicoanalíticas. Se destaca un reporte del estado del arte del año 2002 de Mulder y Nijholt (2002) respecto al humor en sí y respecto al Humor Computacional en el que se engloban algunas de ellas. En el presente capítulo se hace fuerte referencia a este trabajo.

### 2.2. Teoría de la superioridad

Gruner (2000) desarrolla una teoría que afirma que el humor está relacionado con la superioridad, que alguien gana sobre otro dentro del chiste. Al humor se lo ve como al deporte; es más agradable cuando se crea una tensión sobre el resultado, sobre cómo concluirá, y cuando a la vez se “gana” de repente y por sorpresa. Es decir, es más disfrutable ganar un partido parejo que una

victoria demasiado aplastante desde el principio. Lo mismo ocurre con el humor; en un chiste uno genera una expectativa sobre cómo concluirá y de forma repentina aparece el “ganador” y la potencial risa.

Cuando se encuentra humor en algo, hay risa sobre las desgracias o defectos de otro (del perdedor), hay un sentimiento de repentina superioridad, ya que momentáneamente no se está pasando por la situación que le está ocurriendo al otro. En este sentido, sentirse superior es “sentirse bien”, es conseguir lo que uno quiere, es ganar. Un ejemplo de chiste visto fácilmente según esta teoría es (1). El mismo tilda de perdedores a los políticos frente a esta situación, en contraste con los que no lo son, que se sienten ganadores.

- (1) — ¿En qué se parece Superman a un político honesto?  
— En que ninguno de los dos existe.

Risa es igual a ganar según el autor, pero subraya que no es siempre ganar en el sentido de vencer a alguien, sino más en general de conseguir lo que uno quiere, por ejemplo probar estar en lo cierto, o llegar a una situación que uno quería estar. Bajo este esquema uno podría intentar reconocer humor buscando dentro de un conjunto de oraciones por alguien que esté siendo el objetivo de una crítica, y buscar que sea el “perdedor”, siendo éste un requisito necesario para el humor bajo esta forma de pensar.

## 2.3. Teoría del alivio

En este caso se tiene una naturaleza más psicológica. Está relacionada con la propuesta de Freud y Strachey (1905) de que la risa libera la tensión y la “energía síquica”. La misma se acumularía cuando ocurren sentimientos represivos sobre áreas tabú como la muerte o el sexo. Se muestra el chiste (2), que hace referencia a temas que generan tensión.

- (2) Cómo sería de mala aquella suegra que cuando murió, le pusieron este epitafio:  
“Aquí descansa doña Juana Baltasar García. En casa descansamos todos”.

Una versión de por qué uno se ríe en base a esta teoría es porque se experimenta una sensación placentera cuando el humor reemplaza sentimientos negativos como dolor o tristeza.

Rutter (1997) afirma que esta es más una teoría sobre la risa que sobre el humor. Establece que nos reímos para liberarnos de temas difíciles para nosotros, pero no dice cómo el humor que trata de esos temas nos hace reír, qué es lo que cumple ese humor.

Es interesante destacar que hay ciertos temas que, según esta teoría, son más graciosos ya que generan más tensión sobre las personas. Resulta cautivador fijarse si un texto habla por ejemplo de sexo, muerte o superación personal para encontrar humor.

### 2.3.1. Teoría de Minsky

Minsky (1980) sigue la Teoría del alivio, pero agrega que cuando es humor sin sentido (lógica absurda) nuestro subconsciente construye “sensores cognitivos” que actúan igual que los otros sensores mencionados por Freud ante temas tabú. A su vez dice que los sensores pueden aprender y mejorar, y es por eso que un chiste repetido es menos gracioso. Un ejemplo de chiste lógicamente absurdo es (3).

- (3) Un hombre entra al bar y pide una porción de pastel, pero la devuelve rápidamente y en lugar de ello pide una medida de whisky. Se emborracha y comienza a irse sin pagar. El propietario de lugar le dice:

- No pagaste la bebida.
- Pero te di la torta por la bebida.
- Tampoco pagaste por ella.
- Pero no me la comí.

También propone que pensamos en “marcos” o situaciones estereotipadas, y hacemos cambios entre estos marcos, encontrando conexiones entre dos situaciones. Un ejemplo particular es el juego de palabras. Hacer cambios sobre la situación “por defecto” nos hace experimentar humor.

## 2.4. Teoría de la resolución de incongruencias

La Teoría de la resolución de incongruencias (conocida en inglés como *Incongruity-Resolution Theory*) se basa en que la causa de la risa está en la percepción repentina de que hay incongruencia entre un concepto y los objetos reales que se tenían en mente. Según Rutter (1997), la incongruencia ocurre cuando dos objetos son presentados bajo un concepto, que parece aplicar a ambos y los hace similares, pero con el progreso del relato humorístico se vuelve aparente que el concepto sólo aplica a uno y la diferencia se hace visible. Puede ser a su vez que no es la incongruencia la que da la gracia, sino que es la resolución coherente de la aparente incoherencia que la da. Es decir, que cuando aparece una incoherencia en un razonamiento, lo gracioso sale de entender que era otro el camino del razonamiento. Se puede ver un ejemplo de aplicación de la teoría en el chiste (4).

- (4) — Mi amor llevamos 30 años juntos, ¿por qué no nos casamos?  
 — Mejor no, ¿quién se va a querer casar con nosotros?

En este caso, en el comienzo se piensa que “casamos” hace referencia a la pareja entre sí, pero termina haciendo referencia implícita a ellos casándose con otros. Es decir, la segunda oración no es congruente con el comienzo y se tiene que hacer un nuevo razonamiento sobre cómo realmente es.

Al igual que la Teoría de la superioridad, no explica por qué un chiste no es gracioso más de una vez, ni por qué no todas las incongruencias son graciosas. Una posible explicación es que necesitamos distinguir lo serio de los razonamientos equivocados, para saber cómo tratar dicha información, y el humor es una reacción del segundo caso, basándose en Suslov (1992). No se profundizará esta visión ni ninguna otra que fundamente los motivos de la risa ya que no se encuentran dentro del alcance de este proyecto.

La teoría, según Ritchie (1999), no se desarrolla en detalle y deja parcialmente abiertos los conceptos de incongruencia y de resolución. Se puede dividir según diferencias que se presentan a continuación.

### 2.4.1. Desambiguación sorpresiva

Ritchie (1999) desarrolla su visión de la teoría presentada anteriormente (IR) llamada Desambiguación sorpresiva (SD — Surprise Disambiguation, en inglés). Ésta asume que hay dos interpretaciones diferentes al inicio de un chiste, donde una es más obvia, pero la audiencia no se dará cuenta de la segunda hasta que el remate la haga entrar en conflicto y evoque a la otra; se precisa la presencia de ambigüedad antes del remate. Un ejemplo es (5), en donde la pregunta se refiere a por qué lo hacen y no por qué lo hacen volando, en lugar de a través de otro medio.

- (5) ¿Por qué las aves vuelan al norte en invierno?  
 Porque es muy lejos para ir caminando.

Esta teoría habla a su vez de la violación de “expectativas”: puede haber un conflicto entre el remate y lo esperado (que puede no coincidir con lo obvio), como también un conflicto entre la interpretación obvia y el remate. Estos serían dos subtipos de conflicto.

Se tienen las siguientes entidades en esta teoría:

- $M_1$ : la primera interpretación (más obvia) del inicio.
- $M_2$ : la segunda interpretación (oculta).
- $M_3$ : el significado del remate.

Bajo esta teoría se tienen que cumplir las siguientes propiedades:

- **OBVIEDAD:**  $M_1$  es más probable que sea notado que  $M_2$ .
- **CONFLICTO:**  $M_3$  no tiene sentido bajo  $M_1$ .
- **COMPATIBILIDAD:**  $M_3$  sí tiene sentido bajo  $M_2$ .
- **COMPARACIÓN:** existe alguna relación de contraste o contrariedad entre  $M_1$  y  $M_2$ .
- **INAPROPIADO:**  $M_2$  viola alguna norma: es excéntrico, absurdo, tabú, etc.

En realidad, **COMPARACIÓN** no parece necesaria. Por ejemplo no se cumple en el chiste (5). Tal vez su presencia le agrega más intensidad de humor.

Se propone un procedimiento general, y a priori nada fácil, de cómo se podría hallar humor según este modelo, al menos parcialmente. Primero se debe analizar el cuerpo del chiste y extraer su representación simbólica, seleccionando la interpretación más obvia. Luego hacer predicciones según dicha interpretación y al encontrar el remate, detectar el conflicto existente. Hasta el momento el procedimiento encuentra un mal entendido, ¿pero qué lo hace gracioso? Hay que encontrar qué combinación de la solución de los siguientes subproblemas hacen gracioso al texto:

- **OBVIEDAD:** ¿qué hace una potencial interpretación más obvia que otra?
- **CONFLICTO/COMPATIBILIDAD:** ¿qué tan difícil de interpretar debe ser un texto para estimular la búsqueda de otra interpretación? ¿Cómo esta búsqueda puede ser guiada por el mismo texto?
- **COMPARACIÓN:** ¿qué significa que dos interpretaciones sean muy diferentes?
- **INAPROPIADO:** ¿qué hace más divertida a una interpretación?

Surgen más preguntas a partir de esto: ¿cómo medir las propiedades mencionadas por este modelo? ¿Qué intensidad de las mismas las convierte en humor?

#### 2.4.2. Modelo de dos fases

Suls (1972), mediante el Modelo de dos fases (Two-Stage Model, en inglés), interpreta de forma distinta la Teoría de la resolución de incongruencias: asume que el remate crea la incongruencia y que una regla cognitiva debe ser encontrada que permita darle sentido al resto cuando no se cumplen las predicciones en el texto. Es decir, esta teoría afirma, a diferencia de la Desambiguación Sorpresiva, que no es necesaria ninguna ambigüedad al inicio del chiste; simplemente se encuentra o no una regla cognitiva al final, indicando con su presencia si se trata de humor. El chiste (6) es un ejemplo que no presenta ambigüedad lingüística antes del remate (inclusive tampoco en su totalidad), y por tanto no está contemplado por SD, pero sí por esta teoría. Lo mismo ocurre con (7).

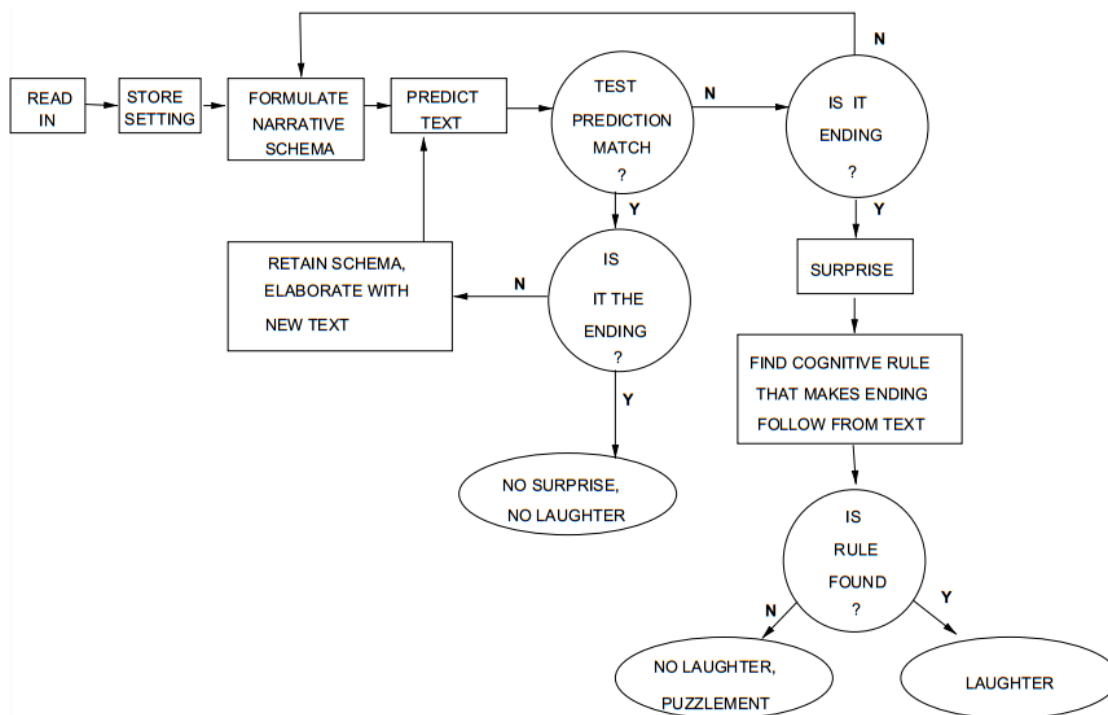


Figura 2.1: Esquema Two-Stage Model de Ritchie (1999)

- (6) Juan está siendo enjuiciado por robo a mano armada.  
 — Inocente — anuncia el juez.  
 — Excelente. ¿Esto quiere decir que me puedo quedar con la plata?
- (7) Un hombre muy gordo va a una pastelería y pide una torta entera.  
 — ¿La corto en cuatro u ocho porciones? — dice la persona que lo atiende.  
 — Cuatro, porque estoy a dieta.

Bajo este modelo se plantea el esquema algorítmico de la Figura 2.1, que se puede encontrar también en el trabajo de Ritchie (1999). Para obtener humor entonces hay que llegar al remate desviándose de una predicción, pero poder encontrar una regla cognitiva que haga tener sentido una segunda interpretación.

De manera parecida a la teoría anterior, se ha planteado un esquema general de detección de humor bajo este modelo, sin embargo no es fácil trabajar a partir de estas ideas. A pesar de que logran una primera partición del problema, lo cual no es trivial, en base a las teorías planteadas, los subproblemas presentados son muy difíciles de resolver y parece que es por ello que no se han encontrado trabajos que se basen en estos esquemas.

## 2.5. Teoría de la violación

Veatch (1998) propone que la condición necesaria y suficiente para la percepción de humor es que se cumplan los siguientes tres puntos:

- **V:** que se viole cierto compromiso de cómo las cosas deben ser según quien percibe.

- **N:** el perceptor tiene un sentimiento dominante de que la situación es normal.
- **Simultaneidad:** V y N están presentes en la mente del perceptor al mismo tiempo.

Es decir la situación parece normal (N) pero al mismo tiempo parece que hay algo mal (V). Por ejemplo, se tiene el chiste (8). La situación parece normal, está mencionando que cuida su celular, que lo tiene en la caja la cual seguramente vino cuando lo compró. Pero luego se viola el pensamiento sobre que lo compró al decir que está en la tienda, simultáneamente, ya que habla de lo mismo que antes, que era normal hasta el momento.

- (8) Mi iPhone 6 Plus lo cuido tanto que todavía está en su caja... en la tienda... en el shopping...

Si se cumple V y no se cumple N, lo que está pasando no es normal entonces, pero si hay una violación. El perceptor en este caso está muy comprometido con la situación y lo ve como algo ofensivo. Esto pasa por ejemplo con los chistes racistas, cuando el que lo escucha pertenece a la raza objetivo del chiste.

## 2.6. Teorías sociológicas

El humor además puede ser visto desde el punto de vista de la Sociología, es decir los comportamientos que aparecen en el humor según las relaciones humanas. Lo que dicen estas teorías no es lo que pasa dentro del ser humano frente al humor, sino entre los seres humanos. El trabajo de Rutter (1997) separa en las siguientes 3 subramas.

### 2.6.1. Teorías de mantenimiento

Estas teorías afirman que el humor mantiene grupos sociales: se fortalece el vínculo entre quien cuenta el chiste y quienes están en su grupo. Por ejemplo si se relata un chiste, los que se ríen pertenecen al mismo grupo de quien lo relata — porque lo entendieron — mientras que los otros están fuera. O también aquellas personas a las que son alcanzadas por un chiste son separadas en un grupo y al resto en otro, como los chistes racistas por ejemplo. Fortalecen vínculos entre quien cuenta el chiste y quienes están en el mismo grupo que él.

### 2.6.2. Teorías de negociación

Miran el rol del humor como medios de interacción o pasatiempos. El perceptor define si es gracioso según su contexto social y cultural. Él está de acuerdo con lo que dice el chiste riéndose, o en desacuerdo si no lo hace. Por ejemplo, alguien podría haber estado mirando con otra persona el partido Brasil-Alemania de la edición 2014 de la Copa Mundial de Fútbol, donde el locatario Brasil pierde con Alemania siete a uno en semifinales, dejando en la etapa anterior afuera a Colombia, y decirle el chiste (9).

- (9) ¿Para eso eliminaron a Colombia?

La otra persona podría reírse, lo cual significaría que está de acuerdo con él, que está perdiendo Brasil de una mala manera. O podría no reírse por algún motivo. Puede no estar de acuerdo por ejemplo porque fueron partidos jugados de manera totalmente distinta, o se lo podría tomar muy en serio el chiste y enojarse. O tal vez la persona, porque ya estaba enojada desde antes con él, simplemente elige no estar de acuerdo y no reírse, por decir motivos.

### 2.6.3. Teorías de contextos

Estas teorías afirman que el humorista hace un salto desde el contexto serio al humorístico, puede criticar sin miedo y hasta puede hablar de temas tabú. Los chistes se alejan de lo que es un discurso normal. Al cambiar a un contexto humorístico introduciendo un tema de tratamiento delicado, uno siempre puede decir “era sólo un chiste”.

## 2.7. Humor según la Lingüística

### 2.7.1. Teoría de los guiones semánticos del humor

Esta teoría (SSTH — Semantic Script Theory of Humor, en inglés) asume que siempre un chiste está relacionado con dos guiones opuestos entre sí en algún sentido. Dice que hasta antes del remate el chiste no es ambiguo, y que éste es el que dispara la ambigüedad. Postula también que hay tres niveles de abstracción de oposición de guiones:

- **El más alto:** real vs irreal.
- **El del medio:** actual vs no actual, normal vs anormal o posible vs imposible.
- **El más bajo:** oposiciones temáticas como bueno vs malo, muerte vs vida, sexual vs no sexual, etc.

El remate puede traer ambigüedad. Si trae contradicción semántica (oposición), se trata de *ironía*.

### 2.7.2. Teoría general del humor verbal

Esta teoría (GTVH — The General Theory of Verbal Humor, en inglés) es formada a partir de SSTH, entre otras. Consiste de parámetros llamados Recursos de Conocimiento (KR — Knowledge Resources):

- **Oposición de guiones:** definida igual que en SSTH.
- **Mecanismo lógico:** la forma en la cual juntar dos guiones diferentes en un chiste.
- **Situación:** el contexto, ya sea los involucrados, objetos, actividades, etc.
- **Víctima:** la víctima del chiste, si la hay.
- **Estrategia narrativa:** organización narrativa (cómo se cuenta) el chiste.
- **Lenguaje:** conjunto de componentes lingüísticos elegidos para formar el texto del chiste.

El orden en que fueron presentados fue estudiado por Ruch, Attardo y Raskin (1993) y se basa en que hay un incremento lineal de similitud entre chistes según esta jerarquía. Primero, cuanto más parámetros tengan en común dos chistes, más parecidos son. Además, que dos chistes tengan iguales sus parámetros menos Lenguaje, por decir un ejemplo, los hace mucho más parecidos entre sí que dos chistes que tengan iguales sus parámetros menos Oposición de guiones. Se puede entender mejor con los siguientes ejemplos:

- (10) — ¿Cuántos gallegos se precisan para cambiar una lamparita?  
— Cinco. Uno para agarrarla y cuatro para girar la mesa.

- (11) — ¿Cuántos gallegos se precisan para lavar un auto?  
— Dos. Uno para agarrar la esponja y otro para mover el auto adelante y hacia atrás.

Ambos chistes tienen la misma Oposición de guiones (tonto/inteligente), el mismo Mecanismo lógico (se intercambia qué cosa se mueve), la misma Víctima (los gallegos), la misma Estrategia narrativa (en forma de pregunta), mismo Lenguaje pero difieren en la Situación: en el chiste (10) está la lamparita y en el chiste (11) está el lavado del auto. En conclusión, y como se puede observar, estos chistes son muy similares entre sí. Más aún, reforzando la jerarquía, podemos notar que el chiste (10) comparado con el mismo pero contado con ingenieros en lugar de gallegos, tienen entre sí un nivel de similitud que es más alto que el de los dos chistes anteriores.

Este orden a su vez indica determinación: un recurso de conocimiento (KR) determina a los de abajo. Es decir, un KR limita las opciones de los KR bajo el mismo, según el orden jerárquico presentado. Por ejemplo, Oposición de guiones afecta la elección del Mecanismo lógico.

Se pone en duda si el Mecanismo lógico es un KR o no, ya que si se comparan dos chistes cualesquiera, la similitud de cada KR es creciente siguiendo el orden jerárquico, exceptuando el mismo. Éste a su vez se puede ver como la parte de resolución de la incongruencia según las teorías que afirman su existencia. Asimismo, los chistes “sin sentido” (que fueron mencionados anteriormente) no tienen Mecanismo lógico, lo cual refuerza la opinión de que se trata de un parámetro opcional.

### 2.7.3. Pragmática

Rachel Giora, según Mulder y Nijholt (2002, pág. 10), dice que: “si un estímulo es óptimamente innovador sería categorizado como más placentero que un estímulo familiar o que uno puramente innovador”. Un estímulo es óptimamente innovador si es una respuesta nueva a un estímulo familiar: aquello que innova y a la vez es familiar, gusta más.

Marlene Dolitsky, citada en el mismo reporte, sostiene que lo gracioso de los chistes está en lo “no dicho”. Por ejemplo en cómo habla el locutor (gestos, pronunciaciones, etc.), pero también en cómo el sentido común hace dar por verdaderas ciertas suposiciones (por ejemplo, saber cómo termina un guion estándar conocido, como los proverbios) y luego el locutor hace ir a los oyentes por otro hilo de razonamiento. Asimismo lo gracioso está en romper aquellas reglas sociales o lingüísticas, sobre todo las no dichas.

## Capítulo 3

# Detección de humor

La temática del humor ha sido cubierta por varias publicaciones y resulta de interés para la comunidad. En este estudio del estado del arte se relevan diferentes trabajos, similares al trabajo que se realiza, que sirven como base.

Se analizan diferentes trabajos que se enfocan en la detección del humor en oraciones generalmente en el idioma inglés, de usualmente un promedio de quince palabras, denominadas *one-liners*. En dichos trabajos se intentan utilizar características (muchas veces denominadas *features*, por su traducción del inglés) específicas del humor y en todas se utilizan métodos de clasificación estadísticos. Primero se analizan características específicas para reconocer humor y luego se evalúan los trabajos estudiados.

### 3.1. Características específicas para el humor

Para poder construir un clasificador binario estadístico de humor, se estudian varias características de estilo específicas en base a dos trabajos. El primero es «Making Computers Laugh: Investigations in Automatic Humor Recognition» de Mihalcea y Strapparava (2005), el segundo trabajo es «Recognizing Humor Without Recognizing Meaning» de Sjöbergh y Araki (2007), que está fuertemente basado en el anterior. También existen diferentes trabajos donde se estudian ciertas características específicas del humor como la ambigüedad, las palabras fuera del vocabulario, etc., sin el objetivo de construir un clasificador como «An Analysis of the Impact of Ambiguity on Automatic Humour Recognition» Reyes, Buscaldi y Rosso 2009 y «Características y rasgos afectivos del humor: un estudio de reconocimiento automático del humor en textos escolares en catalán» Reyes, Rosso et al. 2009, entre otros. A continuación se describen las características específicas para la detección de humor detalladas en los trabajos analizados.

#### 3.1.1. Aliteración

La aliteración es la “repetición notoria del mismo o de los mismos fonemas, sobre todo consonánticos, en una frase”, basándose en la Real Academia Española (2001). Según Mihalcea y Strapparava (2005) las propiedades estructurales y fonéticas de los chistes son tan importantes como el contenido del mismo; muchas veces esta característica importa más que el contenido en sí. Los chistes de una línea son graciosos por aliteración, rimas y repetición de palabras aunque no están pensados para leerlos en voz alta. Un ejemplo de esta característica está dado en el chiste (12). Es importante notar que similarmente la retórica juega un importante rol en los chistes, pero también en otros campos como por ejemplo en los titulares de una noticia.

- (12) Cuando pasé por tu puerta me tiraste una flor, la próxima vez que pase, ¡sin maceta por favor!

Esta característica parece no vincularse con ninguna de las teorías estudiadas. Sin embargo, la aliteración es un condimento que da gusto no sólo al humor, sino a los textos literarios, a las narraciones, canciones, etc.

### 3.1.2. Ambigüedad

Algunas teorías, como por ejemplo Desambiguación sorpresiva (Subsección 2.4.1), afirman enfáticamente que la ambigüedad juega un papel fundamental en el humor. Un ejemplo visto es el chiste (5). La forma más fácil de calcular la ambigüedad es mirando la cantidad de significados de cada palabra en una oración, siguiendo el trabajo de Sjöbergh y Araki (2007). En el mismo también se presenta otra forma de detección de ambigüedad, que viene dada por la cantidad de posibles árboles de parsing encontrados para una oración, aunque una oración bajo un mismo árbol sintáctico puede ser semánticamente ambigua igualmente, como ocurre en (13).

- (13) La perra de mi vecina me ladró.

Existe una medida de la complejidad de una oración, que representa el promedio de algunas de las dependencias sintácticas de la misma, sugerida por Basili y Zanzotto (2002). La fórmula es

$$\frac{N + V}{C}$$

$N$  representa la cantidad de complementos nominales. Ellos son las relaciones sintácticas entre sustantivos y grupos preposicionales (N-GP), como en “té de tilo”.  $V$  representa el número de complementos verbales, es decir, las relaciones entre verbos y grupo preposicionales (V-GP), como en “ir al parque”. Por último,  $C$  representa el número de cláusulas. Las cláusulas son oraciones atómicas. Por ejemplo, la oración (14) tiene 2 cláusulas.

- (14) Voy a lo de Juan y vuelvo más tarde.

Según Reyes, Buscaldi y Rosso (2009) esta medida puede ser utilizada para reconocer la ambigüedad en chistes, dado que, siguiendo su estudio, los chistes tienen una estructura sintáctica más compleja teniendo así un mayor grado de ambigüedad.

Trivialmente estas estructuras suelen ser muy ambiguas. Considerar el ejemplo (15): ¿quién tiene la escopeta? Mirando el contenido, es obvio. Sin embargo, sintácticamente tanto el cazador como la liebre pueden tener la escopeta.

- (15) Cacé a la liebre con la escopeta.

### 3.1.3. Antonimia

La antonimia es la relación de oposición entre los significados de dos palabras (útil — inútil). El humor muchas veces proviene de la incongruencia, oposición u otras formas de contradicción, según Sjöbergh y Araki (2007). Esto es comprobado por la Teoría de los guiones semánticos del humor (Subsección 2.7.1) en su esencia, como así también por la teoría de Minsky (Subsección 2.3.1) que argumenta que el humor puede ser causado por lógica sin sentido. Ésta se puede dar si se presentan antónimos simultáneamente en una oración. Como ejemplo de chiste que utiliza antonimia está el ejemplo (16). Esta característica se puede medir buscando, para

cada palabra en la oración, los antónimos. Si está presente un antónimo en la oración para una palabra, se suma uno obteniendo así el total de antónimos presentes de la palabra. En el trabajo de Mihalcea y Strapparava (2005) se utiliza, para lograr lo anteriormente descrito, la relación antonimia brindada por la herramienta WordNet (Subsección 4.2.2) entre sustantivos, verbos, adjetivos y adverbios. Como características interesantes para decir que hay antonimia en un texto se pueden mirar tanto la cantidad de antónimos de las palabras que componen la oración o también la palabra que tiene más antónimos presentes.

- (16) — ¿Qué le dice Tarzán a un ratón?  
 — ¡Tan pequeño y con bigote!  
 — ¿Y qué le dice el ratón a Tarzán?  
 — ¡Tan grandote y con pañal!

### 3.1.4. Centrado en personas

Los textos de humor parecen referenciar constantemente a escenarios relacionados con las personas, según Mihalcea y Pulman (2007). Se comprobó en el trabajo de Mihalcea y Strapparava (2005) que referencias humanas como “you”, “I”, “person”, “woman”, “man” y “my” son altamente utilizadas en chistes. También en otro estudio realizado por Mihalcea y Strapparava (2010) se comprueba que la palabra “you” se encuentra en un 25 % y la palabra “I” se encuentra un 15 %, en un corpus de chistes de one-liners.

### 3.1.5. Jerga sexual

La Teoría del alivio (Subsección 2.3) argumenta que el humor libera la tensión y energía síquica que se acumula cuando ocurren sentimientos represivos sobre áreas tabú como el sexo. Es por esto que palabras con un tinte sexual podrían constituir una característica específica del humor. Según Mihalcea y Strapparava (2005) el humor basado en jerga adulta es altamente popular, un ejemplo de esto se da en el chiste (17). Para poder medir esta característica se debe tener un conjunto de palabras de orientación sexual. Para esto los autores formaron un léxico extrayendo, de los dominios de la herramienta WordNet, todas las oraciones etiquetadas con el dominio sexualidad. Luego se procesa la lista quitando todas las palabras que tienen más de cuatro significados distintos (*polisemia*  $\geq 4$ ).

- (17) — Che, ¿qué significa “let’s fuck”?  
 — Hagamos el amor.  
 — Bueno, haremos el amor. Pero después me dices qué significa.

### 3.1.6. Negatividad

Según Reyes, Rosso et al. (2009) se destaca que el humor tiene una tendencia hacia las connotaciones negativas. Ellos intentan evaluar si existen verbos o adjetivos que denotan un carácter negativo. Como ejemplo podemos encontrar el chiste (18). Según Mihalcea y Pulman (2007) la característica negatividad también es una característica específica del humor y la subdividen en: palabras de negación como “not” o su abreviación “n’t”, orientación negativa con palabras que tienen polaridad negativa como “bad”, “illegal”, “wrong” y debilidades humanas que refiere a todos los eventos o entidades que están asociadas con un momento de debilidad humana como por ejemplo los sustantivos “ignorance”, “stupidity”, “trouble”, “beer”, “alcohol” o verbos como “quit”, “steal”, “lie” o “drink”. Esto último se relaciona con la Teoría de la superioridad (Subsección 2.2) anteriormente explicada, porque presumen la existencia de un perdedor (como “ignorance” y

“stupidity”). También se puede expresar en función de la Teoría del alivio (Subsección 2.3), ya que estos temas, como “alcohol”, “illegal” y “lie”, generan tensión.

- (18) — Dr., ¿cómo hago para vivir 100 años?  
— Nada de sexo, alcohol ni vicios.  
— ¿Y funciona?  
— No sé, pero seguro que se le va a hacer largo.

### 3.1.7. Palabras clave

Según Sjöbergh y Araki (2007) existen palabras que son más referenciadas en textos de humor que en textos de no humor, y palabras que son más aludidas en textos de no humor que en textos de humor. Un ejemplo son los animales, que usualmente son mencionados en chistes, como en el chiste (19), así como los profesionales como abogados (ver chiste (20)) y programadores, según Mihalcea y Pulman (2007). Este razonamiento se puede, no sólo realizar sobre palabras, sino que también se puede realizar con bigramas y trigramas, y más genéricamente, con modelo de lenguaje. Sjöbergh y Araki (2007) construyen dos listas de palabras, bigramas, trigramas de humor y no humor a partir del corpus de entrenamiento y una vez construida se le asigna un peso a cada palabra, bigrama o trigrama. Todas las palabras, bigramas o trigramas que no estén en las listas tienen peso nulo; son neutras.

- (19) — ¿Cómo se me mete una jirafa en una heladera?  
— Abrís la heladera, metés la jirafa y cerrás la heladera.
- (20) — ¿En qué se diferencia un abogado a un cuervo?  
— En que uno es rapaz, ladrón y traicionero, y si puede te saca los ojos, y el otro es un inocente pajarito negro.

### 3.1.8. Perplejidad del modelo de lenguaje

En el trabajo realizado por Reyes, Rosso et al. (2009) se construye un modelo del lenguaje a partir de narraciones. Luego al graficar la perplejidad de un conjunto de prueba de narraciones y otro de textos de humor se puede ver que la perplejidad del conjunto de textos de humor es mayor a la de las narraciones. La perplejidad es una medida que permite evaluar un modelo de lenguaje: cuanto más bajo sea este factor indica que más se adapta al modelo de lenguaje. Esto indica, entonces, que en los textos de humor se dicen cosas “menos predecibles” comparado con un texto representativo, como las narraciones. Por lo tanto, es menos predecible saber qué palabra sigue a otra cuando aparece en los chistes que cuando aparece en otros textos. Esto también es comprobado en Reyes, Buscaldi y Rosso (2009). La perplejidad puede medir la sorpresa de un chiste, lo que es analizado por la mayoría de las teorías analizadas en el capítulo anterior, como las Teorías de la resolución de incongruencias, SSTH y GTVH, entre otras. También según el trabajo de Reyes, Rosso et al. (2009) y el de Reyes, Buscaldi y Rosso (2009) en los textos de humor es más común encontrar palabras que están fuera del vocabulario (OOV — Out Of Vocabulary) que en textos de no humor.

## 3.2. Evaluación de los trabajos

En el trabajo realizado por Mihalcea y Strapparava (2005) utilizan como características específicas del humor la aliteración, antonimia y jerga adulta. Se puede observar que la aliteración es la característica que parece tener más precisión en la detección de humor. En la propuesta

de Sjöbergh y Araki (2007) utilizan como características específicas del humor las palabras clave, ambigüedad, aliteración, antonimia y palabras centradas en las personas, y se llega a que las palabras clave en chistes es la característica que parece tener más precisión con la detección de humor.

En ambos trabajos mencionados anteriormente resaltan que se obtienen resultados similares con los clasificadores Naïve Bayes y Support Vector Machine, siendo no preferible uno en vez del otro. También comprueban que una mayor cantidad de datos no mejora la precisión al usar los clasificadores luego que la cantidad de datos supera un umbral.

La gran mayoría de los trabajos encontrados utilizan como datos negativos (de no humor) titulares de noticias, proverbios y el British National Corpus (BNC). Tanto Mihalcea y Strapparava (2005) como Sjöbergh y Araki (2007) tenían la intuición de que las frases graciosas son similares a textos creativos, como por ejemplo proverbios, y por lo tanto más difícil de clasificar contra éstos. Sin embargo, se observa que la tarea más difícil es distinguir entre textos de humor y texto regular (BNC).

Combinando las características específicas del humor y los clasificadores basados en contenido del trabajo Mihalcea y Strapparava (2005) se tiene una precisión de 96,95% con el corpus de los titulares de diario, 79,15% con el corpus BNC y 84,82% con el corpus de proverbios. Mientras que Sjöbergh y Araki (2007) tienen una precisión global del 85,4%. Vale aclarar que entre los dos trabajos anteriormente mencionados no se utilizan los mismos conjuntos de chistes y de oraciones dentro de los corpus de no humor por lo tanto los resultados no son comparables.

Los autores en general son positivos en cuanto a los resultados, teniendo en cuenta la precisión obtenida. Hay que considerar que se tratan de versiones simplificadas y muy controladas de la realidad, como se ha mencionado. Clasificar por características consideradas esenciales y por algunas características accidentales en el humor parece haber sido buena idea. Se pueden mejorar los resultados encontrando características que revelen nuevos puntos de vista, o mejorando la forma de buscar ciertos rasgos difíciles de medir como la ambigüedad.

## Capítulo 4

# Herramientas

En este capítulo se hace un breve pasaje sobre las herramientas y recursos disponibles que hay en la actualidad para trabajar sobre la detección de humor.

### 4.1. Corpus

Para poder construir un clasificador de aprendizaje supervisado de humor de tweets en español, se precisa mínimamente un corpus con cada ejemplo etiquetado como *humor* o *no humor*. No se conoce la existencia de un corpus con dichas características. Publicaciones, como la de Mihalcea y Strapparava (2005), han armado su propio corpus. *Twitter* provee una API para la extracción de tweets que puede ser de utilidad.

Corpus en español no específicamente de tweets y no humorísticos existen. Algunos ejemplos son: AnCora de Taulé, Martí y Recasens (2008), IULA Spanish LSP Treebank de Arias et al. (2014) que está anotado sintácticamente inclusive, Wikicorpus de Reese et al. (2010) que contiene 120 millones de palabras y además está etiquetado con información sintáctica y semántica, también está disponible CORIN de Grassi et al. (2001) que es un corpus del español del Uruguay, y puede agregar texto más cotidiano para los uruguayos.

Parece necesario también basarse en alguna herramienta que modele el lenguaje o modelarlo a partir del corpus. Como modelo se puede utilizar, por ejemplo, el modelo de los N-gramas, como así también sería interesante explorar otros, como el sugerido por Mikolov, Le y Sutskever (2013).

### 4.2. Otros recursos

Es fundamental aprovechar los recursos existentes en el área de la Inteligencia Artificial, en particular para Procesamiento del Lenguaje Natural y Aprendizaje Automático.

Al trabajar con textos es esencial primero contar con un analizador léxico o tokenizador; una herramienta que permita obtener los tokens que se desean de un texto, en este caso trabajar por lo menos a nivel de palabras. Es una herramienta en principio sencilla, pero puede requerir que se definan ciertos detalles como por ejemplo cómo tratar a las fechas dentro del texto o a los números. Existen otros puntos a definir, como por ejemplo si es de interés reconocer entidades con nombre, como “San José”, “Juan” o “Ciudad de Buenos Aires”.

Luego hay otro trabajo que suele ser de interés: tener un etiquetador gramatical (POS-tagger, en inglés), que a partir de las palabras en la oración indiquen si se trata de un verbo, sustantivo,

determinante, adverbio, etc., e inclusive puedan dar otra información como el género o el número, qué tipo de sustantivo es, entre otros. Es útil también contar con un analizador sintáctico o parser que dé (o al menos intente dar) posibles árboles sintácticos para una oración, y así medir la ambigüedad, que es una de las características que muchas veces aparecen en los chistes.

Siguiendo la teoría SSTH, es necesario tener un diccionario de antónimos. A un diccionario de sinónimos también se le sacaría provecho para poder asociar palabras con otras y así mirar características de negatividad o jerga sexual, por ejemplo, en el texto.

Respecto al clasificador en sí, es necesario contar con herramientas que permitan construirlo a partir de un corpus de entrenamiento, utilizando alguna técnica en particular, como Naïve Bayes.

A continuación se hace un breve pasaje sobre algunos recursos que existen en la comunidad de forma libre que pueden ser explotados para la detección de humor que cubren los objetivos mencionados e inclusive otros más.

#### **4.2.1. NLTK**

NLTK (2014) (Natural Language ToolKit) es una biblioteca para el lenguaje de programación Python que provee un conjunto grande de herramientas para múltiples idiomas. Cuenta con tokenizadores, etiquetadores gramaticales, analizadores sintácticos, reconocedores de entidades con nombre, corpus enteros de diverso tipo, entre muchos otros recursos.

#### **4.2.2. Wordnet**

Es un corpus de referencia léxica iniciado en Princeton, que según Mulder y Nijholt (2002) está basado en cómo la memoria léxica humana funciona. Las palabras se organizan en conjuntos de sinónimos según cada lexema que componen. A su vez se tienen las relaciones de hiponimia e hiperonimia entre conceptos, es decir las relaciones de tipo “es un”, como cuando se dice que un gato *es un* animal. También guarda lugar para las relaciones de antonimia, la distancia entre dos conceptos (según ciertas medidas definidas), entre otras cosas. Está incluido en NLTK.

Está basado en el idioma inglés, aunque existen alternativas en otros idiomas, como por ejemplo EuroWordNet Vossen 1998 y MultiWordNet Pianta, Bentivogli y Girardi 2002. Existe incluso una asociación, Global WordNet Association (2015), que se encarga de estandarizar, vincular y mantener las versiones de WordNet existentes en los distintos idiomas.

#### **4.2.3. FreeLing**

FreeLing, de Padró y Stanilovsky (2012), es una biblioteca para C++ que provee recursos para análisis lingüístico de distinto tipo: tokenización, separación en oraciones, diccionarios morfológicos, etiquetación gramatical, entre muchos otros.

#### **4.2.4. scikit-learn**

scikit-learn (2014) es una biblioteca para Python que da facilidades en el área Aprendizaje Automático. Ayuda en las tareas clásicas, como clasificación (decir a qué conjunto conocido pertenece una muestra), regresión (asignar un valor numérico a un suceso), agrupamiento o clustering (agrupar entidades similares en conjuntos), entre otras. Es de interés para la detección de humor ya que se puede ver como la clasificación binaria de un texto, es decir, distinguir si pertenece o no al humor.

#### **4.2.5. Apache OpenNLP**

OpenNLP de Apache (2014) es una biblioteca para Java para el Procesamiento del Lenguaje Natural mediante Aprendizaje Automático. Soporta las tareas de análisis lingüístico típicas que se han estado mencionando, utilizando Aprendizaje Automático.

# Capítulo 5

## Resumen

Se estudian varias visiones de lo que es el humor y qué lo hace gracioso. A grandes rasgos algunas ponen el foco en la ambigüedad, otras en el sentimiento de superioridad o en los aspectos sociológicos, y otras en la liberación de cierta tensión existente. Ninguna de ellas cubre todo el espectro de la temática, ni demuestran certeza completa en lo que plantean, pero permiten tener una idea de qué se busca al intentar detectarlo. Se estudia también el trabajo de otros investigadores que buscan el mismo objetivo o similares. Esto sirve de gran referencia en cuanto a resultados esperados y recomendaciones a seguir.

En particular no es un área del Procesamiento del Lenguaje Natural en el que abunden recursos especializados como corpus de humor, menos teniendo en cuenta que se está trabajando para el idioma español. Pero sí hay muchos puntapié de trabajos, por ejemplo teorías que pretenden ser el inicio de un esquema algorítmico como en las teorías de la Resolución de Incongruencias, ideas para obtener más recursos lingüísticos para trabajar como en el trabajo de Mihalcea y Strapparava (2005), entre otras mencionadas en este documento. Sí se cuenta también con herramientas básicas como etiquetadores gramaticales y analizadores sintácticos.

En todos los trabajos estudiados sobre la detección de humor se utilizan métodos de clasificación estadísticos (en particular Naïve Bayes y SVM, con resultados similares indistintamente), enriqueciendo las características con conocimientos lingüísticos. Sin embargo, se enfoca en el estudio del reconocimiento de chistes de tipo “one-liner”, que como se ha mencionado son chistes de aproximadamente 15 palabras. El trabajo de Mihalcea y Strapparava (2005) indica que se deben investigar más características específicas del humor para lograr una mayor precisión. Mientras que Sjöbergh y Araki (2007) mencionan que se debe realizar un estudio más profundo en la característica ambigüedad para lograr mayor precisión.

El conjunto de todas las características tenidas en cuenta por los distintos trabajos son: Aliteración, Ambigüedad, Antonimia, Centrado en personas, Jerga Sexual, Negatividad, Palabras clave y Perplejidad del modelo de lenguaje. En un trabajo la mejor resulta ser la Aliteración y en el otro Palabras clave. A su vez, identificar humor da mejores resultados si se compara con noticias, y peor si se compara con texto cotidiano. Los resultados obtenidos son buenos, pero el problema se trata de forma simplificada por los trabajos anteriores.

# Bibliografía

- Apache (julio de 2014). *Apache OpenNLP*. URL: <https://opennlp.apache.org> (accedido en julio de 2014).
- Arias, Blanca et al. (2014). «The IULA Spanish LSP Treebank: building and browsing». En: *Ninth International Conference on Language Resources and Evaluation (LREC)*.
- Basili, Roberto y Fabio Massimo Zanzotto (2002). «Parsing Engineering and Empirical Robustness». En: *Natural Language Engineering* 8.3, págs. 97-120.
- Freud, S. y J. Strachey (1905). *Jokes and Their Relation to the Unconscious*. Complete Psychological Works of Sigmund Freud. ISBN: 9780393001457.
- Global WordNet Association (enero de 2015). *The Global WordNet Association*. URL: <http://globalwordnet.org> (accedido en enero de 2015).
- Grassi, Mariela et al. (2001). «Corpus informatizado: textos del español del Uruguay (CORIN)». En: *SLPLT-2-Second International Workshop on Spanish Language Processing and Language Technologies-Jaén, España*.
- Gruner, C.R. (2000). *The Game of Humor: A Comprehensive Theory of Why We Laugh*. Transaction Publishers. ISBN: 9780765806598.
- Mihalcea, Rada y Stephen Pulman (2007). *Characterizing Humour: An Exploration of Features in Humorous Texts*.
- Mihalcea, Rada y Carlo Strapparava (2005). «Making Computers Laugh: Investigations in Automatic Humor Recognition». En: *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing. HLT '05*. Vancouver, British Columbia, Canada: Association for Computational Linguistics, págs. 531-538.
- (19 de enero de 2010). «Learning to Laugh (automatically): Computational Models for Humor Recognition». En: *Computational Intelligence* 22.2, págs. 126-142. URL: <http://dblp.uni-trier.de/db/journals/ci/ci22.html#MihalceaS06>.
- Mikolov, Tomas, Quoc V. Le e Ilya Sutskever (2013). «Exploiting Similarities among Languages for Machine Translation». En: *CoRR* abs/1309.4168.
- Minsky, Marvin (1980). *Jokes and the Logic of the Cognitive Unconscious*.
- Mulder, M.P. y Antinus Nijholt (septiembre de 2002). *Humour Research: State of Art*. Imported from CTIT. Enschede, the Netherlands. URL: <http://doc.utwente.nl/63066>.
- NLTK (julio de 2014). *NLTK*. URL: <http://www.nltk.org> (accedido en julio de 2014).
- Padró, Lluís y Evgeny Stanilovsky (mayo de 2012). «FreeLing 3.0: Towards Wider Multilinguality». En: *Proceedings of the Language Resources and Evaluation Conference (LREC 2012)*. ELRA. Istanbul, Turkey.
- Pianta, Emanuele, Luisa Bentivogli y Christian Girardi (2002). «Developing an Aligned Multilingual Database». En: *Proc. 1st Int'l Conference on Global WordNet*.
- Raskin, Victor (1985). *Semantic Mechanisms of Humor*. Studies in Linguistics and Philosophy. Springer. ISBN: 9789027718211.

- Real Academia Española (2001). *Diccionario de la lengua española*. URL: <http://www.rae.es/recursos/diccionarios/drae> (accedido en julio de 2014).
- Reese, Samuel et al. (mayo de 2010). «Wikicorpus: A Word-Sense Disambiguated Multilingual Wikipedia Corpus». En: *Proceedings of 7th Language Resources and Evaluation Conference (LREC'10)*. La Valleta, Malta.
- Reyes, Antonio, Davide Buscaldi y Paolo Rosso (2009). «An Analysis of the Impact of Ambiguity on Automatic Humour Recognition». En: *TSD*. Ed. por Václav Matousek y Pavel Mautner. Vol. 5729. Lecture Notes in Computer Science. Springer, págs. 162-169. ISBN: 978-3-642-04207-2.
- Reyes, Antonio, Martin Potthast et al. (19–21 de mayo de 2010). «Evaluating Humour Features on Web Comments». Inglés. En: *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*. Ed. por Nicoletta Calzolari (Conference Chair) et al. Valletta, Malta: European Language Resources Association (ELRA). ISBN: 2-9517408-6-7.
- Reyes, Antonio, Paolo Rosso et al. (2009). «Características y rasgos afectivos del humor: un estudio de reconocimiento automático del humor en textos escolares en catalán». En: *Procesamiento del Lenguaje Natural* 43, págs. 235-243.
- Ritchie, Graeme (1999). «Developing the Incongruity-Resolution Theory». En: *Proceedings of the AISB Symposium on Creative Language: Stories and Humour*, págs. 78-85.
- Ruch, Willibald, Salvatore Attardo y Victor Raskin (1993). «Toward an Empirical Verification of the General Theory of Verbal Humor». En: *HUMOR: the International Journal of Humor Research*.
- Rutter, Jason (1997). «Stand-up as Interaction: Performance and Audience in Comedy Venues». Tesis doct. Citeseer.
- scikit-learn (julio de 2014). *scikit-learn*. URL: <http://scikit-learn.org> (accedido en julio de 2014).
- Sjöbergh, Jonas y Kenji Araki (3 de septiembre de 2007). «Recognizing Humor Without Recognizing Meaning». En: *WILF*. Ed. por Francesco Masulli, Sushmita Mitra y Gabriella Pasi. Vol. 4578. Lecture Notes in Computer Science. Springer, págs. 469-476. ISBN: 978-3-540-73399-7. URL: <http://dblp.uni-trier.de/db/conf/wilf/wilf2007.html#SjoberghA07>.
- Suls, Jerry M. (1972). «Chapter 4 - A Two-Stage Model for the Appreciation of Jokes and Cartoons: An Information-Processing Analysis». En: *The Psychology of Humor*. Ed. por Jeffrey H. Goldstein y Paul E. McGhee. San Diego: Academic Press, págs. 81-100. ISBN: 978-0-12-288950-9.
- Suslov, I.M. (1992). «Computer Model of a "Sense of Humor" I. General Algorithm». En: *Biofizika* 37.2, págs. 318-324.
- Taulé, Mariona, Maria Antònia Martí y Marta Recasens (2008). «AnCora: Multilevel Annotated Corpora for Catalan and Spanish». En: *LREC*.
- Veatch, Thomas (1998). «A Theory of Humor». En: *HUMOR: The International Journal of Humor Research* 11-2, págs. 161-215.
- Vossen, Piek (1998). *A Multilingual Database with Lexical Semantic Networks*. Springer.