

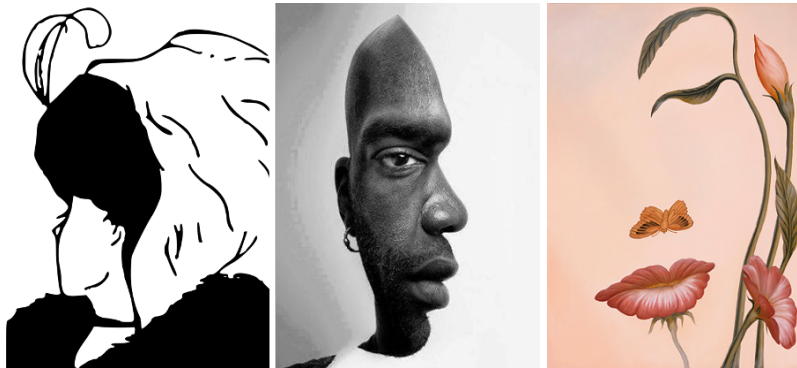
Ambigüedad en el lenguaje natural

... y desambiguación en procesamiento de lenguaje natural

Dr. Ramón Soto de la Cruz (ramon.soto@unison.mx)

Se entiende por ambigüedad la cualidad de una situación, como un dicho, un evento, la actitud de una persona, etc., "que puede entenderse de varios modos o admitir distintas interpretaciones y dar, por consiguiente, motivo a dudas, incertidumbre o confusión" ([drae](#)).

La ambigüedad es un fenómeno común que puede ocasionar problemas de comunicación (particularmente cuando es intencional y escrita en letras pequeñas, al final de un contrato). Para las personas, es un problema relativamente pequeño, ya que, en la mayoría de los casos, el contexto permite eliminar la ambigüedad, incluso sin siquiera notarla. Sin embargo, en el caso de la lingüística computacional, el contexto es difícil de establecer, por lo que este problema se ha convertido en uno de los temas principales de investigación en inteligencia artificial.



En el caso del lenguaje natural, la ambigüedad puede ser *fonética*, *léxica*, *sintáctica*, *semántica*, o *pragmática*.

La ambigüedad fonética ocurre en la lengua hablada cuando una determinada combinación de sonidos puede originar diferentes palabras. En algunos casos, la ambigüedad ocurre debido a imprecisiones en el habla, como por ejemplo en la frase "¿Me diste la caja?" contra "¿Mediste la caja?" o "Dámela vacía" frente a "Dame la vacía". La ambigüedad fonética también puede ocurrir debido a la pronunciación idéntica de una o varias palabras, lo que se conoce como homofonía, como ocurre en las siguientes frases: "Ellos van a su casa" contra "Ellos van a su caza" o "Las bayas están muy altas", contra "Las vallas están muy altas"



En el caso del lenguaje escrito, la forma principal de ambigüedad es la de tipo léxico. Este tipo de ambigüedad ocurre cuando una palabra tiene múltiples significados o usos. Existen dos tipos de ambigüedad léxica: la homografía (que en conjunto con la homofonía conforman la llamada homonimia) y la polisemia.

La polisemia se presenta cuando una misma palabra o signo lingüístico tiene varias acepciones o significados que se relacionan entre sí. Por ejemplo, la palabra "*sierra*" puede tener los siguientes significados:

1. Herramienta para cortar madera u otros objetos duros, que generalmente consiste en una hoja de acero dentada sujeta a una empuñadura.
2. Subconjunto de montañas cuya línea de cumbres tiene forma aserrada, quebrada o bastante pronunciada, por lo general es más larga que ancha.
3. Varios tipos de peces.



En los casos de los significados 2 y 3, la palabra se utiliza en forma metafórica, como remembranza de la primera; el perfil del conjunto de montañas se asemeja a los dientes de una sierra (herramienta), lo mismo que alguna característica en los peces sierra.

Las palabras homógrafas, por otra parte, son aquellas que se escriben de forma idéntica pero tienen diferente etimología y, por lo tanto, distinto significado, por ejemplo, la palabra "*vela*" puede interpretarse como la acción de velar o como un cilindro de cera con una mecha para iluminar, ambos sentidos derivados del latín *vigilāre* y por lo tanto relacionados entre sí; la misma palabra también puede significar una tela grande que aprovecha la fuerza del viento, especialmente en un barco y proviene del latín *vela*, plural de *velum*, derivado a su vez del proto-indo-europeo **weg* (navegar) o **weg^h* (transportar), por lo que su significado es por completo diferente al de *vigilāre*. Lo mismo puede decirse de palabras como "*vino*" (bebida fermentada/conjugación de venir) o "*nada*" (del verbo nadar o ninguna cosa).



La ambigüedad sintáctica ocurre cuando una frase puede analizarse de diferentes formas, como en el caso de la frase "Le disparé a un elefante en mi patio"; esta frase tiene dos posibles significados: "Le disparé a un elefante que se encontraba en mi patio" o "Le disparé a un elefante estando yo en mi patio". Este problema suele ser frecuente con los adjetivos que pueden funcionar como adverbios, como en el caso de "solo" o "rápido": En la frase "Arregló el camión rápido", la palabra "rápido" puede hacer referencia a que el camión arreglado fue "el rápido" o a que la reparación fue "rápida". En ambos casos el significado de la palabra es el mismo y la ambigüedad se deriva del uso que se le da a la palabra en la construcción de la frase.

La ambigüedad semántica se presenta cuando una palabra o una frase tienen por sí mismas un significado difuso, generado frecuentemente por un uso informal o generalizante del lenguaje. Así, por ejemplo, en la frase "*todos los programadores deben conocer tres lenguajes de programación*", no queda claro si se habla de tres lenguajes específicos (C, Java y Python, por ejemplo), como conocimiento básico, o si se refiere a un nivel de diversidad (tres lenguajes cualesquiera). La misma frase pudiera reescribirse como "*todos los programadores deben conocer los lenguajes de programación C, Java y Python*" (por ejemplo) o "*todos los programadores deben conocer al menos tres lenguajes de programación diferentes*". Otro ejemplo lo vemos en la frase "*teme al hombre de un solo libro*", atribuida a Tomás de Aquino (y en algunos casos a Agustín de Hipona). Esta frase puede interpretarse (y quizás esa fue la intención del autor) de dos maneras: "Un hombre que limita su razonar a un solo libro (un solo punto de vista) es una persona peligrosa, intransigente", o bien, "una persona que conoce solo un libro, pero lo conoce muy bien, es un contrincante de cuidado".

Finalmente, la ambigüedad pragmática se presenta cuando una frase tiene dos o más posibles interpretaciones, dependiendo del contexto en que se presenta, como en la frase "*¿me puedes pasar la sal?*" que puede interpretarse como "*¿me haces el favor de pasarme la sal?*" o como "*¿tienes la capacidad de pasarme la sal?*". Sin embargo, a diferencia de los casos de ambigüedad semántica, la ambigüedad pragmática suele resolverse al conocer el contexto. Otro ejemplo de ambigüedad pragmática se presenta en la frase "*José le compró un ramo de flores*", la ambigüedad provocada por el complemento indirecto (*le*), que puede referirse al destinatario de las flores o a la persona que las vende, se resuelve -usualmente- al conocer a quién se señala con "*le*".

La desambiguación es una de las tareas más importantes en procesamiento de lenguaje natural, particularmente en tareas como la minería de opiniones. Aunque aún es un problema abierto (es decir, no existe una solución ni siquiera aproximada a este problema), se han hecho avances importantes, particularmente para la desambiguación de palabras polisémicas (y homónimas). Para realizar esta tarea se utilizan métodos basados en diccionarios, así como métodos de aprendizaje automático (*machine learning*) y, más recientemente, mediante combinaciones de ambas aproximaciones.