

Desambiguación de los sentidos de las palabras y representación lexical del mundo

WIESLAW BANYŚ, AGNIESZKA PALION-MUSIOL, ALEKSANDRA ŻŁOBIŃSKA-NOWAK, *Universidad de Silesia (Polonia)*

wieslaw.banys@us.edu.pl, agnieszka_palion@o2.pl

RESUMEN

En este artículo se aborda la polisemia del lexicón en las lenguas naturales como uno de los mayores escollos de la traducción automática. Se presenta una solución para el problema de la desambiguación de expresiones ambiguas, basada en la descripción orientada a objetos y en la transferencia de las posibles traducciones en una lengua meta a la lengua origen para resolver la posible ambigüedad. Finalmente, se muestran las diferencias en las representaciones del mundo en polaco y español a través de los diversos sentidos en polaco del verbo *ganar*.

Palabras clave: desambiguación, descripción orientada a objetos, clases de objetos, traducción automática, polisemia.

Disambiguation of the Meanings of Words and Lexical Representation of the World

ABSTRACT

This article discusses the natural polysemy of the lexicon of natural languages, perceived as one of the major problems in automatic translation. It also presents a solution for the problem of disambiguation of ambiguous expressions, based on object-oriented description, enriched with a transfer of various possible translations in a target language to resolve the possible ambiguity of a source language. Finally, the authors show differences in the representations of the world in Polish and Spanish, using the example of diverse Polish renderings of the Spanish verb *ganar*.

Keywords: disambiguation, object-oriented description, object classes, automatic translation, polysemy.

1. Desambiguación de los sentidos de las palabras y representación lexical del mundo

En el presente artículo vamos a centrarnos en uno de los mayores problemas de la traducción asistida por ordenador (en adelante TAO) que reside en la polisemia de las palabras en la lengua natural, y por consiguiente, la necesidad de la desambiguación de los significados de las palabras (en adelante WSD = word sense disambiguation); una desambiguación automatizable y eficaz en el tiempo de la interpretación del texto de partida para que un sistema pueda, entre otros, entregar la generación adecuada de sus equivalentes en la lengua meta.

Esto también nos permitirá observar, a base de la descripción de los significados del verbo ES *ganar*, las diferentes representaciones del mundo que nos aporta este verbo en cuestión y sus equivalentes en polaco.

En primer lugar, proponemos ejemplos de traducción real automática generados por algunos sistemas de traducción automática: *Unos ejemplos de las dificultades de partida (Sec.1)*.

Esto nos permitirá discutir:

- por una parte, la cuestión de las distinciones del significado en general, y después, de las que se encuentran en los diccionarios tradicionales de papel y las que hay que establecer por medio de procedimientos descriptivos propuestos,

- y, por otra, las consecuencias que se deducen de éstos para la confección de un diccionario bilingüe que podría servir para la traducción asistida por ordenador (TAO) del ES al PL en particular.

Después, nuestro objetivo será presentar un abanico de los sentidos del verbo ES ganar, definido en los siguientes diccionarios que hemos escogido: *Clave – diccionario del español actual* (CD), *Diccionario de la Lengua Española – Real Academia Española* (DRAE), *Diccionario de uso del Español – María Moliner* (DMM), *Gran diccionario de uso del Español Actual* (GDUES), *Diccionario Salamanca de la lengua española* (DS), *Nuevo “ganar”* (Sec. 2).

Terminaremos nuestra presentación con una descripción que satisfará las exigencias de una presentación contrastiva operatoria de los sentidos, hecha sobre el ejemplo del verbo *ganar* y sus equivalentes polacos: *Ganamos “ganar”* (Sec. 3).

2. Ejemplos de las dificultades de partida

Cuando observamos las traducciones propuestas por la mayoría, para no decir todos, de los sistemas de traducción llamada automática o asistida por ordenador (TAO), podemos notar inmediatamente (véase, por ejemplo, los trabajos de Y. Bar-Hillel (1960) sobre la TAO) que uno de los mayores problemas que estos sistemas tienen que afrontar, es la polisemia de las palabras que a menudo frena demasiado la traducción adecuada de los textos (véase los trabajos de la literatura abundante que analiza este problema, por ejemplo, N. Ide, J. Véronis, 1998).

Aquí tenemos unos ejemplos de traducción propuesta por algunos sistemas actuales en los que la polisemia muestra algunas facetas:

A heavily loaded car really came in → Ciężko załadowany samochód rzeczywiście przychodził
 [Un coche muy cargado en realidad se produjo en]
 The car crashed into the building → Samochód huknął do budującego [El coche se estrelló contra el edificio]
 He always wears a jacket and tie to work → Siempre lleva una chaqueta y corbata para trabajar
 He offered to walk her home → Se ofreció a acompañarla hasta su casa
 El cristal se ha roto → Szklanka jest w podziale [The glass is broken]

La WSD es necesaria no solamente en las tareas de traducción asistida por ordenador (TAO) sino también en otras tareas unidas a un tratamiento computacional de la lengua ampliamente concebido, por ejemplo, la búsqueda de la información y la navegación hipertextual en Internet, el análisis automático de las palabras en partes de la oración, el resumen automático y el tratamiento automático de la palabra.

Dada la complejidad de la tarea es natural que se deba, recurrir a diferentes parámetros «fuentes de información», para intentar sacar adelante la WSD (McRoy, 1992; Ng y Lee, 1996; Wilks et al., 1996; Wilks y Stevenson, 1998b).

Las características generales que nos guían entonces en la WSD son de tipos diferentes, p. ej. partes de la oración y sus géneros (por ejemplo, *un capital* vs. *una capital*, *un pez* vs. *una pez*, *lamentar* vs. *el lamentar*, *poder* vs. *el poder*), sinónimos, antónimos, esquemas

sintáctico-semánticos (por ejemplo, *estar a* vs. *estar de* vs. *estar por*), estructuras predicativo-argumentativas diferentes (por ejemplo, *leer un libro* vs. *tomar un libro*) de las que derivan las restricciones de selección semántica sobre los argumentos de los predicados (por ejemplo, el tipo de abstractos: *observar un partido* (PL: *oglądać mecz*) vs. *observar una regla/una ley* (PL: *przestrzegać reguły/prawa*) lo que es el resultado de la pertenencia a las clases semánticas y las diferentes reglas de herencia semántica, y es, por consiguiente, una manifestación de relaciones semánticas, sintagmáticas y paradigmáticas diferentes cuyo caso extremo, siendo todo un continuum, está representado por las colocaciones o expresiones más o menos fijas (entre otros, Gross (1996) para una explicación de muchos puntos que habitualmente se confunden en lo relativo a las expresiones fijas).

Vamos a añadir también, como factor importante para tomar en consideración, los precedentes desde el punto de vista de la operacionalidad de los diccionarios electrónicos utilizados en la TAO bilingüe, los equivalentes en otra lengua o lengua meta (Dagan e Itai, 1994). Digamos que, desde el punto de vista presentado, hay tantos significados diferentes de una palabra en la lengua de origen como de sus diferentes traducciones a la lengua de llegada y de un conjunto de marcadores sintagmáticos diferentes que los distinguen unos de otros en la lengua de partida.

En un enfoque del tipo «eficacia en primer lugar», la importancia del segundo parámetro puede ser reducida de manera significativa en la medida en que una palabra puede tener diferentes conjuntos de marcadores sintagmáticos en la lengua de partida, sin embargo ser traducido de la misma manera en la lengua meta, lo que es, por lo demás, un reflejo especular dialéctico del enfoque aplicado cambiando de óptica en el camino. Esto supone, desde el punto de vista de cada una de las lenguas, una comunidad semántica operacional de las construcciones a un nivel más abstracto del análisis, tipo de esquema invariante (Desclés y Banyś, 1997) o regla heurística (Banyś, 2002a y 2002b) o bien, un corte semántico de la palabra que lleva en verdad al tratamiento monosémico según A. Wierzbicka (1989) que se pueden sacar más adelante, por ejemplo, ES: *ganar dinero en un juego/ganar un pleito* vs. PL: *wygrywać*.

Este tipo de enfoque se parece también en cierta manera, sin ser idéntico, a la posición en esta materia de A. Kilgarriff (1993) cuyo famoso *I don't believe in word senses* (Kilgarriff, 1997) es la coronación que negaba la posibilidad objetiva del establecimiento de un inventario definido de sentidos de una palabra en la lengua dada, porque, entre otros, los lexicógrafos tienen que, según A. Kilgarriff, describir las palabras como si tuvieran todas un conjunto de sentidos discretos, sin intersecciones. Como consecuencia de esto, observamos que a menudo es arbitrario obtener el sentido de la palabra en el uso concreto en el discurso de uno de los significados de dicha palabra incluidos en el diccionario.

Por cierto, es por eso que, según A. Kilgarriff, por ejemplo, P. Hanks (1994), que analizaba los sentidos de *climb* en inglés o Ch. Fillmore y B.T.S. Atkins (1992) analizando el campo semántico del sustantivo inglés *risk*, propagaban el recurso al análisis prototípico o al análisis en los términos de *frames* como cuadros más detallados en la descripción de los sentidos de las palabras. Sabemos que para este tipo de enfoques había razones más fundamentales: cambio del paradigma de ver la esencia de la lengua en todo: desde el lógico hasta el cognitivo, del parcial hasta el holístico.

Según la posición de A. Kilgarriff (1993), los sentidos de las palabras son definidos sólo

en relación a un conjunto de necesidades de un lexicógrafo, lo que mencionaba, entre otros, de forma mucho más ponderada, por ejemplo, Y. Wilks (1972) y, por otra parte, el conjunto de sentidos de una palabra definida en un diccionario puede o no puede corresponder al conjunto de sentidos que es relevante para la aplicación de un tratamiento computacional de la lengua (véase una discusión crítica esclarecedora de las tesis radicales de A. Kilgarriff (1993, 1997) en Y. Wilks (1997)). Los dos, pero también muchos otros, se ponen de acuerdo en que la asignación de un sentido dado a una ocurrencia de una palabra en el texto es relativa a la elección de los sentidos que presenta un diccionario, a diferencia de A. Kilgarriff, que sostiene que muy a menudo es imposible hacer tal asignación, deslizándose sin duda así desde la pregunta cuáles y cuántos son los sentidos incluidos en un diccionario hasta la pregunta por sentidos nuevos o sentidos puntualmente creados en un discurso por los procedimientos retóricos como metáfora, metonimia etc. Por otro lado, Y. Wilks (1972) distingue netamente los dos problemas, pero, de hecho, la pregunta es saber cómo se establecen los diferentes sentidos, y su número, en un diccionario.

Se puede imaginar que la respuesta será diferente según las tareas a realizar, de las que una de las facetas importantes es saber si se habla de un diccionario monolingüe, bilingüe, papel, electrónico, etc. Nuestra posición se parece al enfoque relativista de A. Kilgarriff (1993, 1997) en el sentido de que insiste en el análisis del corpus para extraer los usos reales de una palabra y relativizar el número de los sentidos de una palabra dada en la lengua de origen en un diccionario electrónico de «interés» particular de la descripción, que es una traducción a otra lengua, lo que se reduce al número de traducciones diferentes que la palabra en cuestión recibe en la lengua meta, pero difiere de éste, porque añade a los elementos de decisión el parámetro de los conjuntos de características sintagmáticas.

Esto nos permite tratar este tipo de enfoque como una cierta ayuda a la hora de analizar la pertenencia de las distinciones de sentidos presentadas en un diccionario tradicional de la lengua de origen.

Todo esto quiere decir que nuestro enfoque se reduce a unos trámites de decisión:

- se establece la concordancia de usos de una palabra analizada en un corpus amplio,
- se reparten las concordancias encontradas en unos conjuntos en los que los elementos tienen más rasgos sintácticos, semánticos y léxicos en común, que con los elementos que pertenecen a otros conjuntos,
- se aplica el enfoque orientado al objeto (G. Gross, 1997, 1999a, 1999b; Banyś, 1997, 2002a, 2002b) para analizar y agrupar las características comunes,
- se les asignan las traducciones en la lengua meta,
- se reorganizan, si es necesario, los conjuntos establecidos conforme a los resultados de la traducción,
- se codifican las conclusiones del análisis en uno de los formatos descriptivos (de acuerdo con las propuestas del enfoque orientado al objeto, se aplica aquí una de los variantes de sus fichas descriptivas).

3. ¿«Ganar» en ganar?

Vamos a aplicar este enfoque a la descripción de los sentidos del verbo ES *ganar*. El

corpus ha sido limitado, teniendo en cuenta el objetivo de nuestra presentación, a las entradas de los siguientes diccionarios: *Nuevo diccionario esencial de la lengua española* (NDE), *Diccionario de la Lengua Española – Real Academia Española* (DRAE), *Diccionario de uso del Español – María Moliner* (DMM), *Gran diccionario de uso del Español Actual* (GDUES), *Diccionario Salamanca de la lengua española* (DS), *Clave – diccionario del español actual* (CDEA).

Esto nos permitió efectuar el inventario de los siguientes tipos de sentidos señalados en los ejemplos de entradas de los diccionarios citados:

1. Ganar la guerra.
2. Nuestra intención es ganar el torneo.
3. Muere boxeador tras ganar pelea.
4. Correa confía ganar plebiscito y disolver el Congreso.
5. ¿Por cuántos goles Panamá le ganará a El Salvador este domingo?
6. El Tau Victoria, campeón tras ganar en la final al Unicaja por 100-98.
7. ¿Cómo ganar a las tragamonedas? Consejos y reglas para mejorar tus probabilidades de ganar.
8. Ganó una buena cantidad de dinero en la lotería.
9. Real Madrid vuelve a ganar con C. Ronaldo y Kaka.
10. Te he ganado en este juego.
11. España ganó ante un rival muy fuerte y agresivo.
12. América necesita empezar a ganar ante el DIM.
13. Al Madrid se le olvida que tenía que ganar ante el Bizkaia
14. ganar un premio literario,
15. ganar el primer premio,
16. Aseguró que el objetivo en Beijing es ganar la medalla de oro.
17. Schumacher quiere ganar el título.
18. ganar puntos
19. El ejército invasor ganó la ciudad con facilidad.
20. Al principio no ganaba mucho, pero pronto ascendió.
21. Ganó mucho dinero en esa operación financiera.
22. Gana cien mil pesetas al mes como vendedor
23. Gana algún dinero haciendo traducciones.
24. Gana dinero y premios haciendo encuestas.
25. Un millón de chilenos gana el sueldo mínimo, o menos que eso.
26. Con sus palabras, logró ganar la atención del auditorio.
27. Con ese gesto te has ganado el cariño de todos.
28. Se ganó una regañina con su comportamiento.
29. ¡Hay que ver cuánto ganas con ese peinado!
30. Hemos ganado desde que llegó el nuevo empleado.

4. Ganamos «ganar»

El análisis sintáctico, semántico y lexical de las construcciones obtenidas nos permitió proporcionar la aproximación siguiente de los sentidos de ganar, observados gracias a sus traducciones polacas, donde entre corchetes [...] se anotan las clases de objetos más generales y entre paréntesis en ángulo <...> las clases de objetos implicadas, más específicas, y todo esto siendo aún relativo:

4.1. X ANM – ganar – Y CONC <co1: encuentro; competición; confrontación> (– con – Z ANM) – wygrać coś (z kimś).

Ganar la guerra.
 Nuestra intención es ganar el torneo.
 Muere boxeador tras ganar pelea.
 Correa confía ganar plebiscito y disolver el Congreso.
 ¿Quién ganará la final de la Champions?
 Gisela Pulido vuelve a ganar el campeonato del Mundo de la PKRA Tarifa Rugby.
 Alemania, 12 años sin ganar un encuentro en el torneo.
 Los trucos de EEUU para ganar la votación sobre la guerra de Irak.
 Correa confía ganar plebiscito y disolver el Congreso.
 Jospin considera que ‘no es fácil ganar un referéndum con un poder tan impopular’
 Una buena manera de ganar polémica a costa de los demás, sin duda es usted listo.
 ♠ ganar como campeón – wygrać jako mistrz
 Simón se da el gusto de ganar como campeón.

4.2. X ANM – (Y ANM <co5: pronombre personal, traducido en polaco como CD>) – ganar por – Z CONC [adjetivo numeral]/cuántos – goles (– a – I CONC [adjetivo numeral]) (– goles) (– a – Y ANM) – wygrać (z kimś) ... golami do ... (goli).

¿Por cuántos goles Panama le ganará a El Salvador este domingo?
 Ganar por tres goles a dos.
 Cristiano: «Somos capaces de ganar por tres goles»
 ¿Por cuántos goles Panama le ganará a El Salvador este domingo?
 «Podemos ganar por dos o tres goles»: Cristiano.

4.3. X ANM – ganar – en – Y CONC – la final (– con – Z ANM) – wygrać finał (z kimś).

El Tau Victoria, campeón tras ganar en la final al Unicaja.
 Bojan dice que “hay que ganar en la final”.

4.4. X ANM – ganar – a/en – Y CONC <co2: juego>; <co3: deporte> (- con - Z ANM) – wygrać w coś (z kimś).

¿Cómo ganar a las tragamonedas?
 Argentina ganó en fútbol, hockey y básquet.
 Ganar en el fútbol.
 ¿Cómo ganar en el póker?
 Consejos de oro para ganar en la ruleta.
 ¿Cómo ganar en el Blackjack?
 Ganar en la bolsa.
 Consejos para ganar en los slots.
 ¿Cómo ganar en el bingo?

wygrać na:

Ganar a la máquina – wygrać na automacie do gier
 Ganar en los juegos olímpicos – wygrać na Olimpiadzie

wygrać z:

Ganar con el ordenador – wygrać z komputerem

Conclusión

Juegos olímpicos pertenecen a la clase <co3>. Los hemos incluido en este esquema por la aproximación al equivalente polaco *wygrać w coś/na czymś*.

Algunas construcciones en castellano determinan el uso en la traducción al polaco de otra preposición. En la lengua española se pueden observar tres distintas preposiciones: *a/en/con*. En polaco: *w/na*.

En cuanto a máquina y ordenador, hay que tratar estos elementos de la clase <juegos> no solamente como juegos, sino como cierto tipo de superficie, fondo que nos sirve para jugar. Así que podríamos excluir estos elementos de esta clase y para estas construcciones detallar la traducción con la preposición *na*.

En el caso del ordenador, el uso eventual de la preposición *z* (equivalente español *con*) significaría que personalizamos el mencionado objeto, tratándolo como “un contrario del mismo valor” que posee inteligencia extraordinaria. Hay que recordar que en polaco el uso de la preposición *con* es también posible:

Najstarszym zawodnikiem był 71-letni ... z Wągrowca (zajął 13 miejsce i jako jedyny wygrał z komputerem.)

Wyrok na Jacksona: człowiek znów wygrał z komputerem.

Dokładnie 14 lat temu, 17 lutego 1996 r., szachowy mistrz świata Garri Kasparow wygrał ostatni z sześciu pojedynków z maszyną i ostatecznie pokonał 4 do 2 komputer «Deep Blue».

4.5. X ANM – ganar (– Y CONC <cantidad de dinero/dinero>/<adjetivo numeral> – (de) CONC <co4: unidad monetaria>) – en/a – Z CONC <co1: encuentro; competición; confrontación>/<co2: juego> – wygrać (coś) w czymś.

Ganó una buena cantidad de dinero en la lotería.

Ganar cien euros en las tragaperras.

Conclusión

En este esquema tiene que aparecer de manera obligada <co1: encuentro; competición; confrontación> o <co3: juego>. En otro caso el sistema elegiría automáticamente el equivalente *zarabiać*, en construcciones con la palabra dinero.

Nótese que la clase <cantidad de dinero/dinero> contiene solamente estos dos elementos.

4.6. X ANM – ganar – con – Y ANM – wygrać z kimś.

Real Madrid vuelve a ganar con C. Ronaldo y Kaka.

La gente quiere ganar con Tiger en el torneo.

La gente quiere ganar con él en el campo.

Los Cavs siguen sin ganar con Antawn Jamison.

Alonso: «Schumacher tardó cinco años en ganar con Ferrari».

4.7. X ANM (– Y ANM <co5: pronombre personal, traducido en polaco como CD>) – ganar – (a/ante – Z ANM) (– a/en – I CONC <co1: encuentro; competición; confrontación>; <co2: juego>; <co3: deporte>) – pokonać kogoś (w czymś).

Te he ganado en este juego.

España ganó ante un rival muy fuerte y agresivo.

América necesita empezar a ganar ante el DIM.

Al Madrid se le olvida que tenía que ganar ante el Bizkaia.

Él me gana al ajedrez.

La U le costó ganar ante el «Chago».

El serbio Novak Djokovic reacciona tras ganar ante Marco Chiudinelli en Rotterdam.

España Saldrá a ganar ante Armenia.

Conclusión

GANAR A/ANTE ALGUIEN AL AJEDREZ, EN UN JUEGO, etc.

En esta construcción aparece obligatoriamente el complemento como un pronombre personal o como un objeto que pertenece a la clase animado introducido por la preposición.

Este complemento influye en la selección traductológica *pokonać* sin tomar en consideración la forma (de estas dos posibilidades) que pueda aparecer. En caso de que una frase no tenga este complemento, la construcción será automáticamente clasificada bajo el primer esquema (como equivalente a *wygrać w coś*).

También observamos que las dificultades que supone el pronombre personal en la tercera persona singular y plural le/les porque los dos no presentan diferencias de género y funcionan como CI.

Para este pronombre hemos elegido el género masculino. Cabe destacar que no es solamente problema de la traducción automática. El traductor del texto que incluye uno de los pronombres le o les sin hacer referencias al contexto más completo seguramente encontrará el mismo problema. La única manera de solucionar el mencionado problema es el análisis del contexto que nos puede aportar las informaciones sobre el género masculino/femenino.

Por la necesidad de crear la clase <co5> todos los pronombres han sido traducidos como pronombres de complemento directo porque lo exige el verbo *pokonać*, que en polaco introduce complemento directo. La lengua española en este sentido del verbo ganar se sirve del complemento indirecto (lo que muestra también una posibilidad alternativa de introducir complemento a través de las preposiciones a/ante).

4.8. X ANM – ganar – Y CONC <co6: premio> (– en – Z CONC <co1: encuentro; competición; confrontación>/<co3: deporte>) – zdobyć coś (w czymś)

ganar un premio literario,

ganar el primer premio,

ganar puntos.

Ha ganado un premio en el sorteo.

Aseguró que el objetivo en Beijing es ganar la medalla de oro.
Schumacher quiere ganar el título.

4.9. X ANM – ganar – Y CONC <co7: territorio u objeto de conquista> – zdobyć coś

El ejército invasor ganó la ciudad con facilidad.
El partido ganó terreno en casi todos los municipios.

4.10. X ANM – ganar (– Y CONC [adjetivo numeral]/<co8: adverbio de cantidad>) (– Z CONC <co4: unidad monetaria>) (– por/a/en – I CONC <co9: periodo de tiempo>) (– como – J ANM <co10: profesión>) – zarabiać jakąś kwotę pieniędzy (na jakiś okres pracy) (jako).

Al principio no ganaba mucho, pero pronto ascendió.
Gana cien mil pesetas al mes como vendedor.
Gana cinco mil pesetas al mes.
Laura gana casi doscientas mil al mes.

Ejemplos de la construcción con el adverbio de cantidad:

Al principio no ganaba mucho, pero pronto ascendió.
Ganaré poco con este trabajo.
◇ ganar dinero – zarabiać pieniądze

Conclusión

Conviene destacar que aunque creemos la clase profesión, ésta no es necesaria para la buena traducción del contexto porque ya han sido detallados todos los elementos imprescindibles que se hallan en el entorno del verbo traducido. La propuesta de establecer esta clase adicional de objetos después del adverbio aumenta (multiplica) solamente la posibilidad de encontrar un equivalente en la lengua polaca. A pesar de eso, hay que subrayar que los contextos complejos que representa todo lo incluido en la construcción sintáctico-semántica son muy poco probables en el uso.

En el caso del contexto con el adverbio de cantidad, el equivalente más frecuente en polaco será el verbo *zarabiać*. El uso del imperfecto nos aporta información sobre el desarrollo temporal, acción habitual, y entonces, si no aparece el adverbio de cantidad, podemos traducir el verbo ganar como *wygrywać*, en particular cuando al verbo acompaña la negación en construcción absoluta que no contiene otros elementos adicionales, como por ejemplo:

Al principio no ganaba mucho, pero pronto ascendió – zarabiać
Al principio no ganaba – wygrywać

4.11. X ANM – ganar (– Y CONC <co8: adverbio de cantidad>) – dinero – en – Z CONC <co11: actividad profesional> – zarabiać pieniądze na czymś.

Ganó mucho dinero en esa operación financiera.

4.12. X ANM – ganar (– dinero) – como – I ANM <co10: profesión; cargo; función> – zarabiać pieniądze jako.

Cuánto dinero es lo que debo de ganar como periodista, la revolución del mundo de los Blogs en los sueldos de los periodistas.

4.13. X ANM – ganar (– Y CONC <co8: adverbio de cantidad>) – dinero – haciendo – Z CONC <co11: actividad profesional> – zarabiać pieniądze wykonując coś.

Gana algún dinero haciendo traducciones.
Ganar dinero haciendo manualidades
Gana dinero y premios haciendo encuestas.
Ganar dinero haciendo apuestas.

4.14. X ANM – ganar – Y CONC <co12: bien material> – dostać coś.

Un millón de chilenos gana el sueldo mínimo, o menos que eso.
Yo gano el sueldo mínimo vital.
Las secretarías médicas ganaron un sueldo anual mediano de \$26.540 de mayo de 2004.
El primer sueldo que gané me lo gasté en regalos para mi familia.

Conclusión

El sueldo podemos *dostać* y *zarobić* y remuneración sólo *dostać*, por eso nos decidimos a traducir los dos contextos con el verbo polaco *dostać*.

4.15. X ANM – ganar – Y ABSTR <co13: opinión positiva> – zyskać coś.

Con sus palabras, logró ganar la atención del auditorio.
Con ese gesto te has ganado el cariño de todos.
Se he ganado nuestra confianza.

◇ ganar tiempo – zyskać czas
◇ ganar peso – przybrać na wadze

4.16. X ANM – ganar con – Y CONC <co14: elemento físico que puede mejorar> – zyskać w czymś.

¡Hay que ver cuánto ganas con ese peinado!

4.17. X ANM – ganar – desde que – Y frase Σ – zyskać od kiedy...

Hemos ganado desde que llegó el nuevo empleado.

A continuación presentamos las clases de objetos que hemos destacado en las mencionadas construcciones. Cabe añadir que por las restricciones editoriales, las clases presentadas

incluyen solamente algunos elementos, objetos que sirven como ejemplo y suponen solo un extracto de unas clases mucho más amplias.

<i>co1</i>	
encuentro; competición; confrontación	spotkanie; zawody; starcie
batalla (f)	batalia
campeonato (m)	mistrzostwo
carrera (f)	zawody
certamen (m)	konkurs
combate (m)	bitwa
etc.	

<i>co2</i>	
juego	gra
acertijos (mpl)	zagadki
dominó (m)	domino
juego de escondite (m)	zabawa w chowanego
lotería (f)	loteryjka
máquina (f)	automat do gier
tragamonedas (fpl)	jednoręki bandyta
etc.	

<i>co3</i>	
deporte	sport
badminton (m)	badminton
bobsleigh (m)	bobsleje
deatlón (m)	dziesięciobój
fútbol (m)	piłka nożna
judo (m)	dzudo
relevos (mpl)	bieg sztafetowy
etc.	

<i>co4</i>	
unidad monetaria	jednostka monetarna
afghani (m)	afgani
balboa (m)	balboa
cent (m)	cent
colón (m)	colón
corona eslovaca (f)	korona słowacka

etc.	
------	--

<i>co5</i>	
pronombre personal CI, traducido en polaco como CD	zaimek osobowy w funkcji dopełnienia dalszego tu: tłumaczony jak dopełnienie bliższe
me	mnie
te	ciebie
le	jego
nos	nas
os	was
les	ich

<i>co6</i>	
premio	nagroda
medalla (f)	metal
puntos (mpl)	punkty
premio (m)	nagroda
título (m)	tytuł
etc.	

<i>co7</i>	
territorio u objeto de conquista	terytorium lub obiekt podboju
atalaya (f)	baszta
castillo (m)	zamek
ciudad (f)	miasto
país (m)	kraj
territorio (m)	terytorium
etc.	

<i>co8</i>	
adverbio de cantidad	przysłówek ilości
bastante	wystarczająco
demasiado	zbyt
enormemente	ogromnie
mucho	dużo
nada	nic
poco	mało
suficiente	dostatecznie

<i>co9</i>	
periodo de tiempo	okres czasu
abril (m)	kwiecień
diciembre (m)	grudzień
medio año (m)	pół roku
mes (m)	miesiąc
semana (f)	tydzień
verano (m)	lato
etc.	

<i>co10</i>	
profesión; cargo; función	zawód; stanowisko; funkcja
abogado (m)	adwokat
acomodador (m)	bileter
banquero (m)	bankier
carpintero (m)	stolarz
crítico de cine (m)	krytyk filmowy
enfermera (f)	pielęgniarka
etc.	

<i>co11</i>	
actividad profesional	działalność zawodowa
apuesta (f)	zakład
encuesta (f)	ankieta
manualidades (fpl)	robotki ręczne
operación financiera (f)	operacja finansowa
traducción (f)	tłumaczenie
etc.	

<i>co12</i>	
bien material	dobro materialne
asignación (f)	pensja
jornal (m)	dniówka
mensualidad (f)	płaca miesięczna
paga (f)	pensja
paga semanal (f)	tygodniówka
remuneración (f)	wynagrodzenie

salario (m)	płaca
salario mínimo (m)	płaca minimalna
sueldo (m)	wypłata
etc.	

<i>co13</i>	
opinión positiva	pozytywna opinia
admiración (f)	podziw
atención (f)	uwaga
cariño (m)	sympatia
confianza (f)	zaufanie
estimación (f)	szacunek
fama (f)	sława
honra (f)	honor
renombre (m)	rozgłos
respeto (m)	respekt
simpatía (f)	sympatia
etc.	

<i>co14</i>	
elemento físico que puede mejorar	element wyglądu fizycznego, który może wpłynąć na jego poprawę
maquillaje (m)	makijaż
peinado (m)	uczesanie
vestido (m)	sukienka
etc.	

Los resultados de estos análisis pueden ser representados en forma de una ficha descriptiva aprovechando solamente las características antes citadas, sin mencionar tales elementos como por ejemplo, sinónimos, antónimos o pertenencia a las clases de herencia semántica (Banyś, 2002a, 2002b).

Además, presentaremos más abajo algunos comentarios sobre el análisis hecho y sus implicaciones teóricas. Sobre todo hacemos hincapié en:

- la asignación posible de un sentido concreto que aparece en el discurso a un sentido definido en el diccionario,
- el número de los sentidos de una palabra en un diccionario monolingüe y bilingüe,
- la validez de la división de los sentidos de las palabras de la lengua de origen según su traducción en la lengua meta,
- la creación de las clases de objetos.

Analizando la primera cuestión que hemos destacado, es completamente evidente que

hay que distinguirla de la segunda y, si no lo hacemos, que es lo que observamos en algunas ocasiones en A. Kilgarriff (1993), no se podrá asignar un sentido nuevo u ocasional a uno de los sentidos presentados en el diccionario, porque será precisamente este sentido el nuevo u ocasional.

Fijémonos en esta cuestión que Y. Wilks (1972), por ejemplo, en su semántica de preferencias otorgó un lugar privilegiado al estudio de las divergencias en la interpretación de las ocurrencias en relación a los sentidos determinados que pueden ser considerados como una norma lexical o una preferencia en el diccionario, y si hay un sentido de ocurrencia que diverge de los sentidos encontrados en el diccionario o se intenta asignar a uno de ellos, por ejemplo, los procedimientos propuestos por Y. Wilks (1972) o, si esto no es posible, se construye un sentido distinto, los procedimientos propuestos por Y. Wilks (1978).

La segunda y la tercera cuestión están integralmente ligadas en el enfoque presentado. Vamos a aclarar nuestra posición, analizando uno de los ejemplos estudiados por Y. Wilks (1972: 12). Y. Wilks (1972), discutiendo la cuestión de asignación del sentido de una ocurrencia dada en un discurso a uno de los sentidos presentados en un diccionario y tratándola como *very difficult*, estudia el caso de *stake EN* en las frases del tipo:

I have a stake in this country.
My stake on the last race was a pound.

y pregunta si las dos ocurrencias de *stake* representan el mismo sentido o no.

Según Y. Wilks (1972), eso depende del hecho si el diccionario semántico contiene un sentido general –fluyente– para *stake* (Stake as any kind of investment in any enterprise) junto al sentido *Stake as a post*, y, si fuese así, la respuesta sería afirmativa, mientras que, si no, entonces existen tres sentidos –*post*, *share*, *amount risked*– para *stake* y, entonces, la respuesta sería negativa, ya que, ante todo, «word sense disambiguation is relative to the dictionary of sense choices available and can have no absolute quality about it» (Wilks, 1972: 12).

Todavía cabe resaltar la pregunta, de cuántos sentidos para, por ejemplo, *stake* hay que introducir en un diccionario por ejemplo, EN – ES o EN – PL, para garantizar una buena WSD con vistas a una buena TAO.

La respuesta en el caso de la traducción es obvia: tantos sentidos para que se pueda traducir correctamente *stake* en ES o en PL y, en efecto, sería difícil traducir las frases antes citadas de la misma manera:

Mam udziały w tym kraju./?x Mam stawkę w tym kraju.
?x Mój udział w ostatnim wyścigu wynosił funta./Moja stawka w ostatnim wyścigu wynosiła funt.
Tengo una participación en este país./?x Tengo una apuesta en este país.
?Mi participación en la última carrera fue un pound./Mi apuesta en la última carrera fue un pound.

El problema en realidad es más complejo, porque incluso el sentido concreto de *stake* tiene al menos dos equivalentes en ES y PL: *estaca/hoguera* y *styp/stos* y sería imposible traducir *stake*:

In Medieval Europe many women were accused of being witches were burnt at the stake
[Cambridge Advanced Learner's Dictionary – CALD] como estaca o slup.

Por todo ello, en particular para la TAO, pero también para la traducción humana, un diccionario bilingüe debería contener al menos estos cuatro sentidos y la traducción – perspectiva de otra lengua– impone este tipo de repartición semántica, como se ha visto anteriormente a través del ejemplo con el verbo *ganar*.

¿Cuáles son –si existen– las causas para confeccionar un diccionario monolingüe?

Es obvio que todo dependerá del objetivo del diccionario y del público pretendido: se puede hacer de la misma manera que lo hace por ejemplo, Collins English Dictionary (CED) o CALD, con las distinciones muy detalladas, sobre todo CED, o Longman Dictionary of Contemporary English (LDOCE) o Wordwise Longman Dictionary, con sus 4 sentidos para el LDOCE.

La elaboración de un diccionario monolingüe debe ejecutarse, al menos en principio, conforme a las distinciones de sentidos encontradas eventualmente en el seno de la lengua y no únicamente conforme a las diferentes traducciones a otra lengua, para evitar *the sense defined by language2 problems*. También huelga recordar a este respecto lo defendido en el título de la obra I. Dagan, A. Itai y U. Schwall (1991) de que *two languages are more informative than one*.

Por otra parte, la cuestión siguiente totalmente práctica que se impone es la de saber cómo cualificar los contextos para determinar los parámetros pertinentes para la distinción entre por ejemplo, el sentido de participación/apuesta o bien estaca/hoguera.

Se pueden tomar en cuenta los parámetros siguientes:

- o lo que permite al lexicógrafo-lumper, el que tiene solamente un sentido global en el diccionario a modo de Wierzbicka de relevar el sentido concreto de una palabra utilizada en el contexto dado al locutor;

- o como lo que permite al lexicógrafo-splitter, de observar la necesidad de distinguir los diferentes sentidos de la palabra y meterlos en el diccionario para «facilitar» la comprensión de la palabra al locutor;

- o como lo que permite al lexicógrafo, que tiene como objetivo la TAO por ejemplo, de distinguir los diferentes sentidos conforme a los diferentes conjuntos de parámetros y a las diferentes traducciones en la lengua meta, asegurando de esta forma una WSD adecuada.

Tal posición es la que hemos adoptado nosotros. Por todo ello, como hemos visto, se ha utilizado con este objetivo el enfoque orientado al objeto con la elaboración de las clases de objetos. Las clases de objetos, desde el punto de vista presentado aquí, forman solamente una variante conceptual de categorías semejantes a las que se encuentran por ejemplo, en un tesoro porque, por una parte, son construidas a partir de datos lingüísticos y no «ontológicos» y, por otra, son mucho más específicas y detalladas que otras.

Las clases de objetos concebidas de esta manera, pueden ayudar naturalmente a clasificar todos los elementos en un texto, por ejemplo, los elementos del contexto que sobrepasan el entorno inmediato de una palabra para traducir; y de este modo se puede determinar por ejemplo, la extensión de 10 a 50 palabras para clasificarlas en una clase de objetos a la izquierda o a la derecha de la palabra en cuestión y aumentar así las posibilidades de una buena WSD (Yarovsky, 1992, 1993; Gale, Church y Yarovsky, 1992).

Por eso, el debate del tipo splitter (más sentido), o lumpers (menos sentido) es interesante solamente en el caso de la elaboración de los diccionarios monolingües y está todavía muy relativizada al objetivo del diccionario.

Para la TAO la cuestión no es tan significativa y a la pregunta de cuántos sentidos tiene una palabra de la lengua origen se puede responder de la siguiente manera: tantos como es necesario para garantizar una buena WSD en la lengua origen y, en consecuencia, la buena traducción en la lengua meta, condicionando así su número al número de las diferentes traducciones en la lengua meta.

NP1	verbo	prep.	NP2	prep.	NP3	NP1	NP1	verbo	prep.	NP2	NP2	prep.	NP3	NP3
CO	verbo	prep.	CO	prep.	CO	caso	CO	verbo	prep.	caso	CO	prep.	caso	CO
[ANM]	ganař	[DET]	[CONC] <CO1>	con	[ANM]	[NOM]	[ANM]	wygrać	,	[ACC]	[CONC] <CO1>	,	[ABL]	[ANM]
[ANM]	ganař	(a)	([ANM]/ [ANM] <CO5>)	por	[CONC] <adjetivo numeral> [CONC] <goles>	[NOM]	[ANM]	wygrać	(z)	([ABL])	([ANM]/ [ANM] <CO5>)	,	[ABL]	[CONC] <adjetivo numeral> [CONC] <golami>
[ANM]	ganař	en	[CONC] <la final>	(con)	[ANM]	[NOM]	[ANM]	wygrać	,	[ACC]	[CONC] <final>	(z)	[ABL]	[ANM]
[ANM]	ganař	a/en	[CONC] <CO2> <CO3>	(con)	[ANM]	[NOM]	[ANM]	wygrać	w	[ACC]	[CONC] <CO2> <CO3>	(z)	[ABL]	[ANM]
[ANM]	ganař	a/en	[CONC] <CO2 : máquina> <CO3 : juegos olímpicos>	(con)	[ANM]	[NOM]	[ANM]	wygrać	na	[ABL]	[CONC] <CO2 : máquina> <CO3 : juegos olímpicos>	(z)	[ABL]	[ANM]
[ANM]	ganař	con	[CONC] <CO2 : ordenador>	,	-	[NOM]	[ANM]	wygrać	,	[ABL]	[CONC] <CO2 : ordenador>	,	-	-
[ANM]	ganař	[DET]	[CONC] <cantidad de dinero/ dinero>/ <adjetivo numeral> de [CONC] <CO4>	en/a	[CONC] <CO1> <CO2>	[NOM]	[ANM]	wygrać	,	[ACC]	[CONC] <cantidad de dinero/ dinero>/ <adjetivo numeral> de [CONC] <CO4>	w	[ABL]	[CONC] <CO1> <CO2>
[ANM]	ganař	con	[ANM]	,	-	[NOM]	[ANM]	wygrać	,	[ABL]	[ANM]	,	-	-
[ANM]	ganař	(a)/(ane)	([ANM]/ [ANM] <CO5>)	a/en	[CONC] <CO1> <CO2> <CO3>	[NOM]	[ANM]	pokonać	,	[ACC]	([ANM]/ [ANM] <CO5>)	w	[ABL]	[CONC] <CO1> <CO2> <CO3>
[ANM]	ganař	[DET]	[CONC] <CO6>	en	[CONC] <CO1> <CO3>	[NOM]	[ANM]	zadobyć	,	[ACC]	[CONC] <CO6>	w	[ABL]	[CONC] <CO1> <CO3>
[ANM]	ganař	[DET]	[CONC] <CO7>	,	-	[NOM]	[ANM]	zadobyć	,	[ACC]	[CONC] <CO7>	,	-	-

[ANM]	ganać	[DET]	[CONC] (<adjetivo numeral-/ <CO8>)/ ([CONC] <CO4>)	por/aten	[CONC] <CO9>	[NOM]	[ANM]	zarabiać	,	[ACC]	[CONC] (<adjetivo numeral-/ <CO8>)/ ([CONC] <CO4>)	na	[ACC]	[CONC] <CO9>
[ANM]	ganać	[DET]	[CONC] (<adjetivo numeral-/ <CO8>)/ ([CONC] <CO4>)	como	[CONC] <CO10>	[NOM]	[ANM]	zarabiać	,	[ACC]	[CONC] (<adjetivo numeral-/ <CO8>)/ ([CONC] <CO4>)	jako	[NOM]-	[CONC] <CO10>
[ANM]	ganać	[DET]	[CONC] <CO8> [CONC] <dinero>	en	[CONC] <CO11>	[NOM]	[ANM]	zarabiać	,	[GEN]	[CONC] <CO8> [CONC] <pieniędzy>	na	[ACC]	[CONC] <CO11>
[ANM]	ganać	[DET]	[CONC] <dinero>	en	[CONC] <CO11>	[NOM]	[ANM]	zarabiać	,	[ACC]	[CONC] <pieniędzy>	na	[ACC]	[CONC] <CO11>
[ANM]	ganać	[DET]	([CONC] <dinero>)	como	[CONC] <CO10>	[NOM]	[ANM]	zarabiać	,	[ACC]	([CONC] <pieniędzy>)	jako	[NOM]	[CONC] <CO10>
[ANM]	ganać	[DET]	[CONC] <CO8> [CONC] <dinero>	haciendo	[CONC] <CO11>	[NOM]	[ANM]	zarabiać	,	[GEN]	[CONC] <CO8> [CONC] <pieniędzy>	wykonując	[ACC]	[CONC] <CO11>
[ANM]	ganać	[DET]	dinero	haciendo	[CONC] <CO11>	[NOM]	[ANM]	zarabiać	,	[ACC]	pieniędzy	wykonując	[ACC]	[CONC] <CO11>
[ANM]	ganać	[DET]	[CONC] <CO12>	,	-	[NOM]	[ANM]	dostać	,	[ACC]	[CONC] <CO12>	,	-	-
[ANM]	ganać	[DET]	[ABSTR] <CO13>	,	-	[NOM]	[ANM]	zyskać	,	[ACC]	[ABSTR] <CO13>	,	-	-
[ANM]	ganać	con	[CONC] <CO14>	,	-	[NOM]	[ANM]	zyskać	w	[ABL]	[CONC] <CO14>	,	-	-
[ANM]	ganać	desde que	Σ	,	-	[NOM]	[ANM]	zyskać	od kiedy	-	Σ	,	-	-

REFERENCES

- Banyś, W. (2002). Bases de données lexicales électroniques – une approche orientée objets: Partie I et II. *Neophilologica*, 15, 7-29, 206-249.
- Banyś, W. (2005). Désambiguïsation des sens des mots et représentation lexicale du monde. *Neophilologica*, 17, 57-76.
- Bar-Hillel, Y. (1960). Automatic Translation of Languages. *Advances in Computer*, 1, 91-163.
- Dagan, I., Itai, A., Schwall, U. (1991). Two Languages are More Informative than One.

- Proceedings of the 29th Annual Meeting of ACL*, 130-137.
- Dagan, I., Itai, A. (1994). Word Sense Disambiguation Using a Second Language Monolingual Corpus. *Computational Linguistics*, 20(4), 563-596.
- Desclés, J.-P., Banyś, W. (1997). Dialogue à propos des invariants du langage (dans une perspective cognitive). *Études Cognitives*, 2, 11-36.
- Diccionario de la Lengua Española. Real Academia Española*. (1997). Madrid: Espasa Calpe.
- Diccionario Salamanca de la lengua española*. (2006). Madrid: Santillana Educación.
- Fillmore, Ch.J., Atkins, B.T.S. (1992). Towards a Frame-Based Lexicon: the Semantic of RISK and its Neighbours. *Frames, Fields and Contrasts*, 75-102.
- Gale, W., Church, K., Yarowsky, D. (1992). One Sense per Discourse. *Proceedings of the DARPA Speech and Natural Language Workshop*, 223-237.
- Gran diccionario de uso del Español Actual*. (2001). Madrid: SGEL, S.A.
- Gross, G. (1996). *Les expressions figées en français: noms composés et autres locutions*. Paris: Ophrys.
- Gross, G. (1997). Les classes d'objets et la désambiguïsation des synonymes. *Cahiers de Lexicologie*, 70, 27-40.
- Gross, G., Class, A. (1997). Synonymie, polysémie et classes d'objets. *Meta*, 42(1), 147-155.
- Gross, G. (1999a). Élaboration d'un dictionnaire électronique. *Bulletin de la Société Linguistique de Paris*, XCIV(1), 113-138.
- Gross, G. (1999b). La notion d'emploi dans le traitement automatique. *La pensée et la langue*, 24-35.
- Hanks, P. (1994). Linguistic Norms and Pragmatic Exploitations or, why Lexicographers Need Prototype Theory, and vice versa. *Papers in Computational Lexicography, COMPLEX '94*, 89-113.
- Ide, N., Véronis, J. (1998). Word Sense Disambiguation: The State of the Art. *Computational Linguistics*, 24(1), 1-40.
- Kilgariff, A. (1993). Dictionary Word Sense Distinctions: An Enquiry into their Nature. *Computers and the Humanities*, 2(1-2), 365-387.
- Kilgariff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, 31(2), 91-113.
- McRoy, S. (1992). Using Multiple Knowledge Sources for Word Sense Disambiguation. *Computational Linguistics*, 18(1), 1-30.
- Diccionario de uso del Español de M. Moliner*. (1994). Madrid: Editorial Gredos.
- Ng, H., Lee, H. (1996). Integrating Multiple Knowledge Sources to Disambiguate Word Sense: An Exemplar-Based Approach. *Proceedings of the 34th Meeting of the Association for Computational Linguistics, ACL-96*, 40-47.
- Nuevo diccionario esencial de la lengua española*. (2000). Dorland: Santillana USA Publishing Company.
- Słownik języka polskiego PWN, 1-3, wersja 1.0*. (2007). Warszawa: Wydawnictwo Naukowe PWN.
- Wierzbicka, A. (1989). *Semantics, Culture and Cognition*. Oxford: Oxford University Press.
- Wilks, Y. (1972). *Grammar, Meaning and the Machine Analysis of Language*. London: Routledge.

- Wilks, Y. (1978). Making Preferences More Active. *Artificial Intelligence*, 11(3), 197-223.
- Wilks, Y. (1997). Senses and Texts. *Computers and the Humanities*, 31, 77-90.
- Wilks, Y.A., Slator, B.M., Guthrie, L.M. (1996). *Electric Words: Dictionaries, Computers and Meanings*. London: MIT Press.
- Wilks, Y., Stevenson, M. (1997). Combining Independent Knowledge Sources for Word Sense Disambiguation. *Proceedings of the Third Conference on Recent Advances in Natural Language Processing Conference (RANLP-97)*, 1-7.
- Wilks, Y., Stevenson, M. (1998a). The Grammar of Sense: Using Part-of-Speech Tags as a First Step in Semantic Disambiguation. *Journal of Natural Language Engineering*, 4(2), 135-144.
- Wilks, Y., Stevenson, M. (1998b). Word Sense Disambiguation Using Optimised Combinations of Knowledge Sources. *The 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics (COLING-ACL-98)*, 1398-1402.
- Yarowsky, D. (1992). Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora. *COLING - 92*, 454-460.
- Yarowsky, D. (1993). One Sense per Collocation. *Proceedings ARPA Human Language Technology Workshop*, 266-271.