

Desambigüación monolingüe y traducción automática

N. Bel Rafecas
EUROTRA-ESPAÑA
Barcelona

Resúmen

La aparición del clítico "SE" dispara en castellano, al igual que en otras lenguas romances una serie de fenómenos de difícil análisis si tenemos en cuenta que en todos sus valores la distribución es la misma y que la interpretación semántica no viene determinada exclusivamente por el valor del verbo sino por otros factores como son:

- correferencia entre sujeto y SE
- agentividad del sujeto gramatical
- que el sujeto gramatical presente el rasgo +definido o no
- transitividad/intransitividad del verbo
- sujeto cat=np /S' /INF

En el artículo se ofrece una vía de análisis que analiza una construcción como reflexiva, pasiva, media, impersonal o recíproca teniendo en cuenta todos estos factores además del valor léxico del verbo. Con este análisis aseguramos que el valor del verbo no será el único determinante del análisis, permitiendo el análisis correcto de verbos pronominales o reflexivos en construcciones pasivas o impersonales. Con esto se puede ofrecer una caracterización de la oración que permita una traducción más simple dentro del diseño del proyecto EUROTRA de traducción automática.

Desambiguación monolingüe y traducción automática

El diseño del proyecto EUROTRA de traducción automática obliga a tratar una lengua particular con el fin de desambiguar sus fenómenos teniendo en cuenta sólomente la información monolingüe a través de los diferentes niveles de representación lingüística en los que se estructura el sistema:

- EMS, nivel de representación morfológico
- ECS, nivel de representación sintáctico superficial (configuracional)
- ERS, nivel de representación relacional o sintaxis profunda
- IS, nivel de representación sintáctico profundo enriquecido con información semántica.

El hecho de que en el último nivel, el "semántico", se ofrezcan representaciones aptas para traducción a cualquier otra lengua, obliga a plantearse qué información va a ser indispensable, cómo representarla y dónde situarla.

La partícula SE en castellano ha sido un buen ejemplo de cómo enfrentarse a la desambiguación monolingüe teniendo que llegar al nivel de transferencia con unas especificaciones suficientes y necesarias para la traducción.

El primer paso fue identificar los diferentes valores semánticos que dispara la aparición de este clítico. Decidir que verbos tienen la posibilidad de aparecer con este clítico, que valor semántico pueden tomar y cómo marcarlo en la entrada léxica. Por último valorar la capacidad de llenar una posición argumental que tiene el clítico.

Tradicionalmente se establecen los siguientes valores de las construcciones con Se:

- . reflexividad y reciprocidad
 - . verbos pronominales intrínsecos:
 - absolutos: no tienen correlato no pronominal.
 - verbos pronominales que reducen un argumento gracias a la aparición del Se
 - incremento léxico: la aparición del SE causa una variación del significado que no es relevante.
- Juan bebe una cerveza /
Juan se bebe una cerveza
- . dativo ético
 - . pasividad
 - . impersonalidad

La distribución del clítico en estas construcciones es en todos los casos la misma:

- . se antepone al verbo flexionado o aparece como enclí-

tico con las formas nominales verbales.

. se antepone a cualquier otro clítico.

. concuerda en persona con el sujeto de la oración. En el caso de las impersonales y pasivas reflejas, la tradicional invariabilidad de la partícula puede entenderse como concordancia con la tercera persona del singular, forma canónica de la impersonalidad, o concordando con el SN, siempre no personal, que funciona como sujeto gramatical en las construcciones de pasiva refleja.

Eliminamos el análisis de las construcciones que no son significativamente relevantes: incremento léxico y dativo ético. La necesidad de reflejar el valor semántico de las oraciones responde a que éste no podrá ser estructuralmente representable en todos los casos. Para asegurar que la información es transferida al módulo de la lengua de llegada, de forma que ésta tenga posibilidad de representarla en superficie, se marca la oración con un rasgo, representando el valor que el análisis ha establecido.

De las construcciones anteriormente mencionadas se desprenden los siguientes valores semánticos:

. reflexividad y reciprocidad (correferencia entre el sujeto y el objeto -directo o indirecto- que encarna el clítico)

. valor medio "naturaleza" (se-moyen)
"Este libro se lee bien"
"Las heridas se curan solas"

. valor medio "anticausativo"
"El vaso se rompió"
"La puerta se cerró de golpe"
"El barco se hundió rápidamente"

. impersonalidad o sujeto humano de referencia arbitraria
"En España se vive bien"

. pasivización
"En España se leen libros poco recomendables"

La primera cuestión es cómo discriminar estos valores ya que la distribución del SE es la misma. Unas características que nos pueden ayudar son:

- correferencia entre Sujeto y Se, que nos separaría las reflexivas y recíprocas y las medias (siguiendo la definición de diátesis media) de las impersonales y pasivas reflejas.

- pérdida de agentividad/anticausativización, que separaría las medias de las reflexivas y recíprocas. Hay, sin embargo, que hacer una salvedad: cuando en una oración media se mantiene el mismo sujeto humano que en la oración en su uso

transitivo, es difícil distinguir entre reflexividad y anti-causatividad:

"Juan se descuelga" / "El cuadro se descuelga"

La pérdida de agentividad en las medias tiene un reflejo sintáctico, que es la pérdida de una posición argumental o "intransitivización". Como se ha comentado antes, cuando se pierde el argumento "tema" y se mantiene el mismo "agente" o sujeto que en el uso transitivo, se desdibuja la distinción entre reflexividad y diátesis media; entenderemos, a partir de ahora, como construcciones de valor medio sólo aquellas en las que se produce un cambio

OBJ => SUBJ

y la consiguiente intransitivización del verbo.

Sin embargo esta característica también puede servir para definir la impersonalidad y la pasiva refleja. La pérdida de agentividad en todas estas construcciones hace difícil distinguir entre ambos valores para frases como:

"Las puertas se cierran a las nueve en punto"

En ocasiones la tematización, anteposición del sujeto, hace preferente una interpretación a otra, y, en ocasiones, la aparición de la preposición A ante el objeto desambigua sintácticamente la construcción:

"El paisaje mediterráneo se caracteriza por la abundancia de pinos"

"Se caracteriza al paisaje mediterráneo por la abundancia de pinos"

pero es agramatical:

"Al paisaje mediterráneo se caracteriza por la abundancia de pinos"

que debería ser:

"Al paisaje mediterráneo se LO caracteriza ..."

donde aparece claramente el efecto de la dislocación.

Las aproximaciones teóricas a este problema suelen señalar la necesidad de representar en el léxico la posibilidad de algunos verbos de tomar valor medio con la aparición de la partícula SE. Sin embargo, la duplicación de la entrada léxica estableciendo un esquema argumental transitivo con interpretación neutra y, por otro lado, una entrada del mismo verbo que acepte un Se-verbo con valor medio tiene un peligro: admite un solo análisis del verbo así marcado cuando aparece con el clítico, aunque es posible

tener pares como:

"La economía europea se hundirá a menos que ..."

con valor medio/anticausativo, frente a:

"De continuar esta política, se hundirán mercados poco estables"

donde sería preferible una lectura pasiva. O bien:

"Estos documentos se leen con facilidad"

"Se lee poco en España"

"Los proyectos se desarrollaron en centros de investigación privados"

"Se desarrolló un nuevo proyecto de traducción automática"

El registrar en el léxico el posible valor medio es inevitable ya que no todos los verbos pueden tomar este valor que parece quedar restringido a aquellos verbos que tiene objeto implicado (Moreno 1985). Proponemos entonces:

1) marcar, ya desde el nivel configuracional, los verbos con un rasgo [pron=] que tomará los siguientes valores:

- . abso (absoluto) cuando sea un verbo pronominal absoluto
- . si, cuando se trate de un verbo pronominal con correlato no pronominal.
- . reflex, cuando se trate de un verbo tradicionalmente reflexivo o recíproco.
- . nil, cuando no pueda aparecer con valor medio.

Con esta clasificación nos ahorraremos entrar los verbos pronominales como entradas a parte.

2) en el nivel configuracional se analizará cualquier construcción con SE como

s

x cli GV y z ..

y para los verbos marcados como "abso" habrá una regla que añadirá el rasgo "expletivo" al clítico, pudiendo así ser borrado en el paso al nivel siguiente.

3) en el nivel relacional, el clítico que no acompañe a un verbo marcado como "abso" tomará la posición argumental de sujeto. De esta forma no necesitaremos multiplicar el análisis de las oraciones dependiendo de su función y podremos mantener los esquemas argumentales de los verbos codificados como verbos transitivos normales o intransitivos.

4) En el paso al nivel de interfaz (IS) tenemos tres reglas

que distinguirán las construcciones de la siguiente manera:

. Cuando la oración no lleve complementos la regla de traducción entre ERS e IS le dará valor IMPERSONAL. La razón es fundamentalmente que mientras las oraciones que llevan complementos pueden traducirse por pasiva, aquellas que no los llevan han, forzosamente, de traducirse por impersonales.

. Cuando la oración con SE lleve en posición de objeto directo o indirecto) un nombre marcado como "humano", la regla lo analizará como REFLEXIVO por las razones que hemos comentado anteriormente. Vemos así que tenemos dos valores reflexivos: uno que se le adjudica al verbo desde el léxico otro que se le da a la frase. De esta forma la traducción a otra lengua podrá tener información tanto léxica como contextual. Esta regla también ordenará los argumentos de forma que el SE quede en su lugar de Objeto directo o indirecto.

. Otra regla de traducción separará las oraciones con SE que no hayan casado con las reglas anteriores dejándolas en un valor para el raso "diátesis". En el nivel siguientes reglas contexto/acción distinguirán entre:

- diátesis pasiva: para oraciones con verbo marcado [pron=nil] o con verbo [pron=yes] que tengan un argumento-2 (es decir el sujeto en la representación superficial) no definido;

- diátesis media: para oraciones con verbo marcado [pron=yes] cuyo argumento-2 sea definido.

Estas últimas reglas distinguirán entre oraciones como:

"Los precios se mantuvieron estables"

"Se mantuvo una postura crítica ..."

Con este tratamiento no se ha de esperar una total ambigüedad, pero también es cierto que los límites entre estos fenómenos son muy difusos.