

O PAPEL DA LEXICOGRAFÍA NAS NOVAS TECNOLOXÍAS APLICADAS Á TRADUCCIÓN¹

Enrique Jiménez Ríos
Universidade de Salamanca

Resumo: O obxectivo deste estudio é mostra-la importancia crucial que ten a lexicografía no desenvolvemento que están a experimenta-las novas tecnoloxías do eido da traducción. Un estudio pormenorizado das ferramentas electrónicas que empregan os tradutores amosa a importante posición que ocupa todo o relativo ó vocabulario, sobre todo se temos en conta a idea que se deriva do modelo xenerativo gramatical de *Principles and Parameters*, de que o que se entende por coñecemento sintáctico é simplemente o resultado da interacción da información específica das unidades léxicas e os principios de proxección do vocabulario na sintaxe. Por este motivo e dado que o vocabulario é o que determina a estrutura sintáctica, cando se elaboran programas de traducción automática e se crean dicionarios electrónicos, é fundamental que se apliquen os principios teóricos de lexicografía para mellora-los resultados actuais que proceden da utilización destes novas ferramentas lingüísticas.

Abstract: *The aim of this piece of work is to show the crucial role Lexicography must play in the development that new technologies in the field of translation are undergoing. A detailed study of the electronic tools used by translators, shows the important position occupied by everything related to vocabulary, above all if we take into account the idea arisen from the grammatical generative model of Principles and Parameters, of that what was considered as a syntactical knowledge is no more than the result of the interaction between the specific pieces of information from the lexical units and the principles of projection of vocabulary in Syntax. For this reason and, as being vocabulary the one which*

1. Traducción do castelán ó galego realizada por M^a Belén Souto García, licenciada en Traducción e Interpretación pola Universidade de Vigo.

determines the syntactical structure, the application of the theoretical principles of lexicography in the making of programmes of automatic translation and in the creation of electronic dictionaries, is a fundamental piece to improve the present results brought about by the use of these new linguistic tools.

1. Introducción

A aplicación das novas tecnoloxías ó estudio científico dunha lingua determinou a maneira de investigar. O feito de dispoñer en soporte magnético da información que se ten que analizar foi fundamental neste novo xeito de investigar. O apoio da informática ofrece moitas posibilidades no tratamento dos feitos lingüísticos e permite que o estudio poida facerse desde distintos puntos de vista, debido á súa capacidade de reutilización. Pero debido a este cambio de actitude non se debe deixar de conceder importancia á análise lingüística en si mesma, tal como se veu facendo no plano gramatical no marco dos distintos modelos teóricos que xurdiron na derradeira metade do século XX. As últimas versións destes modelos atribuíronlle un papel fundamental ó léxico. Ó mesmo tempo que se supuxo necesario abordar unha análise lingüística previa antes de facer uso do que soamente son ferramentas –certamente moi útiles– para a investigación, aínda que a súa utilización e explotación nunca son un fin en si mesmas, a investigación lingüística concedulle unha importancia preponderante ó estudio do léxico; deste xeito fíxose fincapé en que non só é necesaria a investigación gramatical, senón, sobre todo, a léxica.

O desexo de procesar automaticamente unha lingua natural foi, precisamente, unha das razóns que levou a formaliza-las linguas co obxectivo de desenvolver, entre outros recursos, técnicas de tradución automática (é dicir, tradución semiautomática con intervención humana, como fai Gómez Guinovart (1996:99)). No curso desa formalización foi cando se apreciou a importancia do léxico como elemento central da gramática, pola facultade que ten para determina-la estrutura oracional e organiza-los elementos que conforman unha frase en obrigatorios e opcionais (Gutiérrez Cuadrado 1994:643), o que confirmou que os problemas cos que se atopa un ordenador no momento de interpretar e analizar unha secuencia oracional se deben, como sinala Pinker (1999:246), a que

*“el lenguaje de la vida real está muy lejos de *El perro come helado* y que comprender una oración es mucho más que efectuar un análisis sintáctico”*

pola importancia do significado e pola necesidade –e dificultade– de introducir nos ordenadores unha gran cantidade de información procedente do “coñecemento do mundo”. Nestas páxinas centrareime en amosa-las razóns que levaron ó léxico a ocupar un papel preponderante na análise lingüística

–cambio que partiu do denominado *xiro lexicista*²– e o xeito en que se plasmou esa transformación no momento de codifica-lo léxico dunha lingua nos distintos tipos de dicionarios³ (en concreto, nos dicionarios destinados á produción, máis que a responder á función descodificadora para a que se deseñaron estas obras ata hai relativamente pouco tempo). Disto se deriva o papel determinante que ten que desempeña-la lexicografía nas novas tecnoloxías lingüísticas e, en particular, na tradución, pola necesidade de que os repertorios lexicográficos –sobre todo, os que conteñen os programas de tradución automática⁴– ofrezan unha información que, ou ben é nova para moitos dicionarios, ou, se xa aparecía, facíao dun modo deficiente⁵; ó mesmo tempo,

2. Diso ocupámonos nas seguintes páxinas, pero anticipamos aquí qué se entende por *xiro lexicista*. Trátase da importancia que ten o léxico sobre a sintaxe, fronte ó que se defendía nos posicionamentos iniciais da gramática xenerativa. Demonte (1990:115) explica do seguinte xeito o *xiro lexicista* que sufriu a investigación gramatical: “la investigación sintáctica de los últimos años ha experimentado lo que podríamos denominar un nuevo *giro lexicista*, mucho más bien definido, cierto es, y más prometedor que el que conocimos a finales de la década de los sesenta, en los años de la polémica sobre la semántica generativa. Lo que ha ocurrido, en realidad, lo que subyace a este cambio de mira en la investigación gramatical, es que se ha comprobado que las reglas y los principios que determinan la buena formación estructural de las oraciones son, en verdad, muy pocos y de una extrema generalidad. Al hilo de ese descubrimiento ha emergido la convicción de que buena parte de lo que denominábamos conocimiento sintáctico no es nada más que el resultado de la interacción entre información específica sobre unidades léxicas y principios de proyección del léxico en la sintaxis.”

Pola súa banda, Gutiérrez Cuadrado (1994:642) refírese tamén ó significado do predomínio do léxico sobre a sintaxe ó aborda-las conexións entre léxico e gramática. Señala que “se están desarrollando diferentes tendencias teóricas y ensayando nuevas aproximaciones al problema, incompletas y heterogéneas, pero todas coinciden en cuestiones fundamentales: que la gramática posee un lexicón compuesto por un conjunto no ordenado de entradas léxicas que especifican el conjunto de los argumentos inherentes, semánticamente seleccionados por un predicado y catalogados como papeles temáticos.”

3. Da importancia do léxico no traballo da tradución automática e, en concreto, na confección dos dicionarios que contén o sistema xa se ocupou, no marco do estruturalismo, Pottier (1961), co obxectivo de chama-la atención sobre a necesidade de considera-la existencia de campos ou redes semánticas no momento de deseña-los programas de tradución.

4. Nesta exposición móstrase a necesidade de facer unha distinción entre os programas de tradución automática desenvolvidos en organismos internacionais, como a Unión Europea, e os programas destinados á súa comercialización para usuarios de ordenadores persoais, se ben é certo que os avances da disciplina teñen que cristalizarse en ámbolos dous tipos de proxectos.

5. Pensemos, por exemplo, na falta de exemplos que avalan o uso das palabras en moitos dicionarios en soporte papel, situación que caracterizou os dicionarios da tradición lexicográfica española. Isto explica que Alvar Ezquerro (1996:12) sinala que “uno de los mayores males que ha afectado a nuestros diccionarios y que ahora puede comenzar a cambiar, repito, con lentitud, es la ausencia de ejemplos, tanto más significativa en los repertorios destinados a la enseñanza de la lengua [...] Ya nadie discute la conveniencia de las citas o ejemplos como modelo de empleo para que quien consulte la obra sepa cómo debe o puede emplear la voz tratada. Esto es, el diccionario no sólo debe cumplir con una función prescriptiva o normativa, sino también descriptiva, y, por supuesto, didáctica.”

imponse a necesidade de contar con bancos de datos de todo tipo –xerais, terminolóxicos ou fraseolóxicos– para que o traductor os empregue *on line*⁶.

Pero antes de trata-la repercusión que debe te-la lexicografía nas tecnoloxías aplicadas á traducción, voume ocupar brevemente das necesidades para procesar automaticamente unha lingua natural, pola importancia que teñen os avances en enxeñería lingüística para o desenvolvemento da intelixencia artificial no referente ó léxico, así como das limitacións que supón para a traducción automática o simple encadeamento de unidades léxicas nunha oración, sen que o deseño dos programas se apoie nun modelo teórico previo que conecte a gramática e o dicionario, o que ata hai relativamente pouco tempo non ocorría⁷.

2. Necesidades para o procesamento automático dunha lingua natural

2.1. Participación da enxeñería lingüística e da intelixencia artificial

A actividade de programar ordenadores para realizar tarefas intelixentes, tipicamente humanas, é o que se coñece como *intelixencia artificial*. Dentro destas tarefas está a de comprende-la linguaxe, para o que a mediados dos anos 50 xurdiron nos Estados Unidos as primeiras investigacións en lingüística computacional, encamiñadas a permiti-lo procesamento de feitos lingüísticos co obxectivo de lograr que os sistemas informáticos fosen capaces de interpretar e de xera-las linguas naturais⁸. Desde aquela, o cambio que foi experimentando unha lingua, a consecuencia dos avances en enxeñería lingüística, débese ás moitas posibilidades que brinda a informática para o desenvolvemento das novas tecnoloxías lingüísticas⁹: isto é o que se deduce, por exemplo, dos informes¹⁰ que sobre esta materia publica paulatinamente a

6. Cf. a información sobre recursos lingüísticos en internet para o uso por parte dos tradutores e, en particular, o acceso a bases de datos e repertorios lexicográficos que ofrece Gómez Clemente (1996:88-90). E cf. tamén as observacións que fan Santana e outros (1996:69-70) sobre as posibilidades dos actuais dicionarios *on line* consultados e examinados no seu artigo.

7. Gutiérrez Cuadrado (1994) expón a historia das relacións entre gramática e dicionario. Interézanos para o noso traballo a afirmación que fai (op. cit. p. 640) de que “así como ha evolucionado la teoría lingüística, igualmente debería evolucionar la gramática que está presente en los diccionarios.”

8. En concreto, a lingüística computacional ocúpase do estudo dos sistemas de computación utilizados para comprender e xerar unha lingua natural. Trátase de determinar cál debe se-la estrutura destes sistemas e cáles son os algoritmos de que deben dispoñer. Isto, que é o aparato técnico necesario, non pode deixar de lado o teórico.

9. Este aspecto comentouno Pascual (1996:146) cando chama a atención sobre a importancia que ten a informática no desenvolvemento da enxeñería lingüística, en xeral, e da comunicación intralingüística, en particular.

10. A finalidade destes informes é dar a coñece-los diversos usos da enxeñería lingüística, dentro dos proxectos desenvolvidos no sector da Enxeñería Lingüística do Programa de Aplicacións Telemáticas da Dirección Xeral XIII da Unión Europea (Soler

Unión Europea, en relación co respecto á diversidade lingüística dos Estados membros¹¹. Deste xeito, no documento titulado *Cómo aprovechar la fuerza del lenguaje* explícase a función da enxeñería lingüística:

“la ingeniería lingüística es la aplicación de los conocimientos sobre la lengua al desarrollo de sistemas informáticos para que puedan reconocer, comprender, interpretar y generar el lenguaje humano en todas sus formas. En la práctica, la ingeniería lingüística consiste en una serie de técnicas y recursos lingüísticos que se aplican, en el primer caso, por medio de programas informáticos y que, en el segundo, constituyen una fuente de conocimientos a los que se puede acceder por medio de estos mismos programas”¹²;

é dicir, que a lingua serve á informática para que posteriormente a informática sirva á lingua, nun proceso cíclico que vai do teórico ó técnico, co obxectivo de que o desenvolvemento das ferramentas técnicas reverta no coñecemento teórico. O que se pretende co uso dos medios técnicos é conseguir melloras na comprensión da linguaxe, dado que resulta fundamental para moitas aplicacións. Sen embargo, non sempre é necesario que a devandita comprensión sexa perfecta, como o reflicten, no eido da tradución automática, as traducións asistidas por ordenador que teñen que ser posteditadas ou revisadas: nestes casos pode ser de gran utilidade un achegamento parcial ó texto, que permita, en posteriores niveis de análise, adquiri-lo grao de comprensión necesario. Deste xeito,

1994). Poden citarse, por exemplo, a Comunicación da Comisión das Comunidades europeas, *La sociedad de la información multilingüe*, Bruxelas, 8 de novembro de 1995; a decisión do Consello das Comunidades europeas, do 21 de novembro de 1996 (DOCE, 28.11.1996), relativa á adopción dun programa plurianual para promover a diversidade lingüística da Comunidade na sociedade da información (anexo I); o documento *Lenguaje y Tecnología. De la torre de Babel a la aldea global*. Luxemburgo, Oficina de Publicacións das Comunidades europeas, 1997; o Plan de Traballo Trienal (1996-1998) da *Sociedad de la Información Multilingüe (SIML)*, do 5 de febreiro de 1997; *Language Engineering. Progress and Prospects'98*, da Dirección Xeral XIII, elaborado por Anite Systems (Luxemburgo) para o proxecto LINGLINK; o informe *Cómo aprovechar la fuerza del lenguaje*, preparado tamén por Anite Systems (Luxemburgo) para o proxecto LINGLINK (sen data e cunha tradución e adaptación ó español do Observatorio Español de Industrias de la Lengua do Instituto Cervantes); ou o *Multilingual Information Society (MLIS) Programme to Promote the Linguistic Diversity of the European Community in the Information Society*, publicado pola Comisión Europea o 3 de xuño de 1999.

Pola súa banda, Soler (1994:384) fala da necesidade de desenvolver aplicacións prácticas para o almacenamento, tradución, recuperación e manipulación da información monolingüe e multilingüe, tarefas que están presentes nos citados informes.

11. Soler (1994:383) que sostén que “a pesar de que las fuerzas económicas imponen cada vez más el uso del inglés como lengua de comunicación internacional, una solución monolingüe no es deseable en Europa donde la diversidad lingüística y cultural e identidad nacional son rasgos constituyentes de su carácter multilingüe y multicultural.”

12. *Cómo aprovechar la fuerza del lenguaje*, p. 5.

“el análisis superficial o parcial de los textos se emplea para conseguir una primera clasificación fiable de textos sin restricciones. Este análisis inicial puede utilizarse luego, por ejemplo, para centrarse en las partes más relevantes de un texto, con vistas a un análisis semántico más profundo que determine su contenido en un dominio delimitado. También puede usarse, junto con los conocimientos estadísticos y lingüísticos, para identificar automáticamente las características lingüísticas de las palabras desconocidas, que pueden luego incorporarse a los conocimientos del sistema”¹³;

isto é, que a análise paulatina do texto, polo procedemento de *ensaio-erro* ou pola interacción home-máquina, vai ofrecendo pistas sobre os cambios que se poden introducir no algoritmo do sistema, chegando, mesmo, a determina-las características idiosincrásicas do texto que se deben engadir ó programa; ou ben, dá lugar á aparición de distintos tipos de programas para distintos tipos de textos, como ben se reflicte nos que están destinados a *sublinguaxes* ou linguaxes específicas¹⁴.

Esta ampliación e mellora do uso dunha lingua como ferramenta eficaz no eido da enxeñería lingüística, débese á gran cantidade de coñecementos que se teñen hoxe sobre as linguas, gracias ás investigacións de que foron obxecto. Isto explica que a historia do procesamento dunha lingua natural se enlace coa das investigacións sobre a linguaxe desenvolvidas en disciplinas afíns, tales como a lingüística formal¹⁵, a psicoloxía cognitiva ou a lóxica e, no ámbito da informática, a compilación de datos e, sobre todo, a intelixencia artificial. Un dos resultados máis espectaculares que se deriva diso é precisamente a creación dos primeiros programas de traducción automática, baseados na correspondencia entre as unidades léxicas dos idiomas participantes, coas consecuentes carencias e limitacións, a partir das que se desenvolveu e impulsou a investigación lingüística aplicada neste terreo, pois, como apuntan Hutchins e Somers (1995:26),

13. *Cómo aprovechar la fuerza del lenguaje*, p. 8.

14. Esta maneira de proceder deu como resultado a confección de programas de traducción automática creados cunha linguaxe controlada para un determinado tipo de textos (vid. Sabah (1991:39)). Abonda con citar, por exemplo, o proxecto canadense TAUM-MÉTEO para a traducción do inglés ó francés de partes meteorolóxicos.

15. Jordán (1994:252) explica que as investigacións no tratamento dunha lingua natural se inscriben dentro da intelixencia artificial e da informática, pero están separadas da lingüística tradicional, se ben a continuación sinala que disciplinas como a lingüística e a psicoloxía cognitiva se tocan: “la lingüística trata de modelos estructurales, formales y generales, de lenguajes naturales. Los lingüistas tratan con modelos que captan las regularidades del lenguaje y tratan de generalizaciones lo más apropiadas posible lingüísticamente. No se preocupan de la eficacia computacional (gramáticas chomskianas, por ejemplo).”

Pero a isto pódese obxectar que o feito de que as gramáticas chomskianas non xurdisen por motivos computacionais non significa que a lingüística computacional non se servise delas como punto de partida para face-lo seu traballo.

“los principales obstáculos a la hora de traducir por ordenador no han sido ni son informáticos, sino lingüísticos. Son problemas tales como la ambigüedad léxica, la complejidad sintáctica, las diferencias de vocabulario entre lenguas, las construcciones elípticas o incorrectas gramaticalmente; en definitiva, todo lo que se requiere para extraer el “significado” de las oraciones y textos a partir del análisis de signos escritos y para obtener oraciones y textos en otro conjunto de símbolos lingüísticos que tengan un significado equivalente”.

Co fin de mellora-los resultados da traducción automática, o deseño dos primeiros programas informáticos, destinados a procesar unha lingua natural, obrigou a adapta-la linguaxe –entendiendo adaptación como formalización dunha lingua– ás posibilidades de procesamento que ofrecen os ordenadores, pois, como sinala Pinker (1999:212)

“la comprensión de una oración es una de estas sencillas tareas complicadas. Para poder trabajar con ordenadores, aún es necesario aprender su idioma, ya que ellos no son lo bastante listos para aprender los nuestros. De hecho, se tiende a atribuir a los ordenadores más capacidades de comprensión de las que en realidad poseen”;

o que determinou a inserción das investigacións de lingüística teórica nas que estaba comezando a desenvolverse no terreo do procesamento dunha lingua. Iso favoreceu a aparición, por unha banda, de analizadores sintácticos (*parsers*) co fin de supera-la localidade da traducción *palabra por palabra* das primeiras propostas de traducción automática e, por outra, a elaboración de modelos semánticos baseados na representación do significado con independencia das linguas que interviñesen, xa que

“los modelos semánticos se emplean para representar el significado de la lengua en términos de conceptos y de las relaciones existentes entre ellos. Un modelo semántico puede usarse, por ejemplo, para relacionar una solicitud de información con su significado básico, independientemente de la terminología utilizada o de la lengua en la que se haya realizado la consulta. Así, se puede acceder a la información multilingüe sin necesidad de conocer la terminología o la estructura empleadas para indizar¹⁶ la información. La combinación de análisis y generación con un modelo semántico permite la traducción de textos. En su estado actual, los programas de traducción exigen vocabularios y conceptos limitados para que puedan utilizarse los recursos de ingeniería lingüística adecuados. Si se requiere la generación de textos de gran calidad, puede recurrirse a plantillas para la estructura de documentos y frases frecuentes con partes variables. Es posible partir de la representación semántica de un texto como base para la generación del len-

16. Quérese dicir, como se explica no glosario que acompaña o informe, para “construir un medio conciso de referencia a la información en una base de datos que, en el caso de la información textual, puede basarse en palabras clave o conceptos.”

guaje. Una interpretación de datos básicos o el significado subyacente de una oración o frase pueden representarse como una cadena superficial de caracteres, tanto en una lengua determinada como en función de las indicaciones de carácter estilístico de un sistema de planificación de textos¹⁷.

A necesidade de *analizar e xerar* unha lingua, como eixes do procesamento dunha lingua natural, levou a contar coa participación da lingüística teórica, capaz de somete-los feitos lingüísticos a datos básicos, formalizables e “traducibles” á linguaxe dos ordenadores. No proceso de análise e xeración, os problemas de ambigüidade léxica¹⁸ outórganlle, unha vez máis, un papel importante ó léxico, pola necesidade de restrinxi-lo uso dunha peza léxica atendendo ó seu comportamento sintagmático e contextual; de feito, como sinala Jordán (1994:257) no proceso de xeración dunha lingua natural trátase de que

“la elección del léxico no ocurra de una forma aislada sino como parte del problema de la elección lexicogramatical. [...] Se han desarrollado generadores basados en la teoría significado-texto, donde el léxico desempeña el papel central, que influye en el proceso de generación”.

2.2. Participación da lingüística teórica e da análise formal dunha lingua

Entre as aplicacións principais da lingüística computacional está, certamente, a tradución automática, pero tamén se conta coa recuperación da información e a interacción home-máquina. Sen embargo, moitos dos investigadores que se ocupan da lingüística computacional perseguiron unicamente proba-la validez das gramáticas propostas pola lingüística teórica; isto é, a súa finalidade era utiliza-lo ordenador para verifica-lo funcionamento dun conxunto determinado de regras¹⁹. Gracias a esta última aplicación, é posible

17. *Cómo aprovechar la fuerza del lenguaje*, p. 8. Esta mesma idea é a que propón Jordán (1994:253) cando di que “en el T[ratamiento]L[lengua]N[atural] general se realiza la traducción de una aserción a una representación interna sin ambigüedad con inferencia basada en un sistema de conocimiento del mundo real, potencialmente sin límite.”

18. A ambigüidade é o problema principal para o procesamento dunha lingua natural. Como apunta Carbonell (1994:248) “el lenguaje natural es localmente ambiguo y la resolución de ambigüedades es necesaria para un procesamiento eficaz.” É o que vemos en exemplos como o seguinte, con ambigüidade léxica: *She cannot bear children*, que ten un significado susceptible de varias interpretacións “que non pode ter fillos” ou “que non soporta coidar ou estar con nenos”. A este tipo de ambigüidade pódesele engadi-la estrutural e a referencial.

19. Existe o *Transformational Grammar Tester* de Friedman, como comprobador de gramática transformacional. Este sistema, como sinala Grimshaw (1991:16), xeraba oracións de acordo cunha gramática transformacional dada para que os lingüístas puidesen verificar se as súas gramáticas producían só oracións gramaticais. Sen embargo, gran parte do marco formal da teoría lingüística (a natureza das regras de movemento, as restricións sobre as transformacións, a forma das regras de interpretación se-

conecta-la lingüística computacional coa lingüística teórica, da que ineludiblemente hai que partir, pois canto máis explícita sexa a descrición da lingua que ofrezca a lingüística teórica, máis fácil resultará a súa adecuación computacional²⁰; máis aínda, non se tería producido ningún avance en lingüística computacional se antes non se tivese feito o mesmo coa lingüística teórica, se ben é certo que a lingüística teórica e a lingüística computacional teñen obxectivos e puntos de vista diferentes, como sinala Grishman (1991:17) ó afirmar que

“a pesar de que el objetivo de ambas es la comprensión de los procesos lingüísticos, la lingüística teórica y la lingüística computacional tienen métodos y puntos de vista bastante diferentes. Los que se dedican a esta última se han ocupado del desarrollo de procedimientos para el tratamiento de un subconjunto de las lenguas naturales. Están, en general, dispuestos a aceptar soluciones aproximadas que cubran la mayor parte de las oraciones de interés y a conformarse con un sistema aunque éste no sea capaz de tratar algunos aductos peculiares. La exigencia de construir sistemas completos que funcionen les ha llevado a buscar un entendimiento del proceso total de la generación y comprensión del lenguaje”;

se ben, a pesar do carácter xeral da lingüística computacional fronte o carácter restrictivo da lingüística teórica, este autor reconece inmediatamente despois (op.cit., páx.18) que, a pesar destas diferencias,

“la lingüística teórica puede dar a la lingüística computacional un sustento a menudo ignorado. Las cuestiones de agramaticalidad son importantes porque la experiencia ha demostrado que una restricción gramatical que en un caso determina si una oración es o no aceptable, puede ser necesaria en otros casos para decidir cuáles de los varios análisis posibles de una oración son correctos. Las relaciones que se establecen entre conjuntos de oraciones, tema primordial en la gramática generativa, especialmente en el modelo de Harris, son esenciales para los procedimientos de análisis del lenguaje, puesto que permiten reducir una amplia variedad de oraciones a un número relativamente pequeño de estructuras. Además, las reglas formales de interpretación semántica estudiadas por Montague y sus discípulos, y desarrolladas por otros lingüistas, están comenzando a dar frutos significativos a la lingüística computacional”²¹.

mántica) estase a cuestionar, e os esforzos actuais da lingüística teórica non se centran na construción de gramáticas concretas, que poidan ser probadas computacionalmente. O resultado disto foi un escaso uso dos ordenadores como comprobadores das teorías lingüísticas.

20. As tarefas previas requiridas para o procesamento dunha lingua natural son a análise morfolóxica, a sintáctica, a semántica e a pragmática; isto é, unha detallada e exhaustiva análise lingüística (vid. Carbonell 1994:250).

21. A proposta de Montague combina teoría sintáctica con lóxica formal. O principio básico de dita proposta é o da *composicionalidade semántica*; en conformidade con ela, o significado dunha expresión complexa –como as oracións– é o resultado dos significados de expresións básicas –como as palabras.

Isto explica que, na actualidade, o desenvolvemento experimentado pola investigación lingüística nos últimos corenta anos, centrada en deseñar e explicar cómo é a gramática que ten interiorizada un falante, que lle permite xerar un número infinito de oracións a partir de elementos finitos, poida servir como punto de partida para a formalización computacional dunha lingua natural, cun conxunto finito de elementos, susceptible de ser recollido nunha base de datos e cun sistema de regras capaces de combina-los devanditos elementos²², a pesar das diferencias que poidan apuntarse entre os obxectivos de ámbalas "lingüísticas"²³.

Durante un primeiro momento pensouse que esa gramática interiorizada ou gramática universal estaba formada por un sistema de regras acompañado dun procedemento para elixir entre elas²⁴, aínda que hoxe dita gramática se concibe como un modelo no que existe un sistema de *principios* e *parámetros*: é dicir, un conxunto moi restrinxido de condicións sobre a boa formación dunha oración con valor universal e uns parámetros ou valores susceptibles de ser tomados por eses principios, despois de que se entra en contacto cun medio lingüístico concreto. Dentro do sistema de principios hai varios subsistemas,

22. Como sinala Pinker (1999:88) "la gramática constituye un ejemplo de sistema combinatorio discreto en el que un número finito de elementos discretos (palabras, en este caso) son objeto de selección, combinación y permutación para crear estructuras más extensas (frases, en este caso) que presentan propiedades muy distintas de las de sus elementos constitutivos".

23. Hutchins e Somers (1995:127-128) sinalan que "con frecuencia la investigación en el campo de la TA ha recibido críticas por soslayar los avances efectuados en el ámbito de la lingüística teórica [...] Hasta hace poco tiempo, las teorías lingüísticas no ofrecían una explicación adecuada del uso del lenguaje en toda su dimensión. Aunque se han realizado, por ejemplo, análisis convincentes de los cuantificadores o de la coordinación, no se han descrito todos los aspectos de su empleo real, tal y como requiere la TA. Recientemente, teorías como la gramática léxico-funcional, o la gramática de estructura sintagmática generalizada y sus derivados, han comenzado a cubrir de forma explícita una amplia gama de aspectos referidos tanto a lenguas particulares como a conjuntos de lenguas diferentes. En el pasado, una buena parte de la teoría lingüística se basaba en fenómenos observados del inglés, la lengua que mejor se ha estudiado tras la revolución generativista. Por desgracia, y hablando en términos generales, esto sigue sucediendo hoy en día. La desatención hacia otras lenguas ha sido otro de los motivos por los que la teoría lingüística ha tenido una influencia menor sobre la TA de lo que algunos observadores habrían esperado. En otras palabras, la teoría lingüística rara vez ha estudiado problemas de lingüística comparada, es decir el modo en que las lenguas se distinguen al emplear medios diferentes para expresar significados e intenciones semejantes. Tales problemas son, obviamente, de importancia fundamental para la TA."

24. Descubrimiento de pouca utilidade para a intelixencia artificial como apunta Sabah (1991:45), que sinalara, ó narra-la historia do nacemento da intelixencia artificial en 1956, que "de manera paralela aunque independiente, Chomsky presenta la teoría de las gramáticas formales que han influido mucho sobre la investigación en lingüística [...] Aunque haya sufrido diversas modificaciones, tal teoría permanece sin aplicar en la inteligencia artificial".

módulos ou teorías que se manifestan en catro niveis de representación. No nivel da Estructura-P, as oracións aparecen na súa forma máis básica ou pura (probablemente reflexo directo do léxico) e suponse que na determinación desa estrutura básica interveñen dous subsistemas do coñecemento lingüístico: o de carácter formal, que é o *módulo ou teoría do X-con barra*²⁵ e o de carácter semántico, que é o *módulo ou teoría temática*. A importancia e conexión entre estes dous subsistemas explícaa Demonte (1991:135) cando sinala que

“las condiciones formales y semánticas (o temáticas) que asociábamos a las estructuras-P dan parcialmente razón de la buena formación de las oraciones. [...] La suposición que subyace al establecimiento de estas condiciones es la de que en el trasfondo de la correcta formación sintáctica se encuentran dos tipos de conocimiento: el relativo a los principios que regulan las relaciones entre constituyentes [determinado por el módulo o teoría de la X-barra] y el que se refiere a las relaciones entre predicados y argumentos [determinado por el módulo o teoría temática]”;

se ben a relación entre ambos módulos é de dependencia, xa que o primeiro módulo está condicionado polo segundo. Isto explica que se puxese en cuestión a hipótese de que a gramática debe incluír dúas especificacións paralelas e independentes: por unha banda, a solución semántica e, por outra, a selección categorial. Por esta razón, co fin de evita-la redundancia na representación destas dúas informacións relacionadas e dependente a segunda da primeira, considérase, como fai Demonte (1991:90), que

“las configuraciones sintácticas serán siempre un reflejo de las propiedades del léxico entrelazadas con restricciones formales mínimas impuestas por los, principios relativos a la forma general de los sintagmas (teoría X) [...], sin que medie entre ambos niveles –el léxico y la estructura P– ninguna información categorial específica”.

O módulo temático ou teoría temática ten, entón, a función de determinar cales son as relacións semánticas que se establecen entre os constituíntes dunha oración e é o encargado de explica-la natureza destas relacións, os actores das mesmas e os principios que as regulan²⁶. O que pretende a devandita teoría é poñer de manifesto que o valor semántico dunha expresión referencial situada no interior dun sintagma ou oración, non só está determinado polo contido intrínseco de cada un dos elementos constituíntes, é dicir,

25. A propósito da participación desta teoría no procesamento da linguaxe natural, Hutchins e Somers (1995:68) afirman que “la idea básica de la teoría X puede encontrarse en varias teorías lingüísticas actuales, así como en el ámbito de la lingüística computacional, aunque las aplicaciones directas en la TA son poco habituales y relativamente recientes”.

26. Hai un antecedente disto na gramática de casos de Fillmore (1968) que basea a análise da frase nas relacións entre o verbo e os obxectos.

polo valor de cada peza léxica recollida por medio de definicións nos dicionarios, senón tamén polo significado que recibe da súa relación cos restantes membros do sintagma; ese valor é o que se coñece como *papel temático*. Diso despréndese a existencia de certos elementos asignadores de papeis temáticos, que determinan o valor que toman os argumentos seleccionados polos devanditos asignadores. Así, na determinación dunha estrutura oracional resulta clave a estrutura argumental que desenvolva unha peza léxica, entendida esta como a lista de papeis temáticos correspondentes ós argumentos que deben necesariamente proxectarse con ela cando se realice sintacticamente. Deste xeito, concédese unha importancia capital ó léxico, pola cantidade de información que hai depositada nel e por se-lo responsable último da configuración estrutural dunha oración²⁷. Isto fai que o encadeamento das unidades léxicas nunha oración non estea rexido unicamente por principios de linearidade, senón tamén e, sobre todo, de xerarquía, o que, aplicado ó ámbito da tradución automática, revela as limitacións do método de tradución *palabra por palabra*, propios dos programas de primeira xeración.

3. O léxico como motor do encadeamento das unidades léxicas nunha oración

O que se deduce da existencia dunha orde estrutural en cada secuencia oracional, que se rexe pola información contida no léxico, é que o mecanismo seguido para xerar computacionalmente unha oración dista moito de ser un simple sistema de encadeamento, entendido este como un sistema combinatorio discreto capaz de crear un número ilimitado de combinacións diferentes a partir dun conxunto finito de elementos, segundo un procedemento similar ó que explica Pinker (1999:95):

“un sistema de encadenamiento de palabras comprende montones de listas de palabras (o frases prefabricadas) y un conjunto de instrucciones para pasar de una lista a otra. Un procesador construye una oración a base de seleccionar una palabra de una lista, luego otra de otra lista y así sucesivamente. Y para reconocer una frase, lo que hace es localizar de una en una cada palabra en su correspondiente lista”.

27. Outra cosa é considerar se se pode prescindir ou non dunha análise sintáctica e quedar só cunha análise semántica, como fai Grishman (1991:22): “algunos lingüistas computacionales han optado por omitir la regularización sintáctica por completo y dejar que el componente semántico opere directamente sobre la totalidad de las estructuras oracionales. En tales sistemas, la regularización sintáctica se subsume, de hecho, en el proceso de análisis semántico. Esto requiere, sin embargo, reglas semánticas más complejas, por lo que la inclusión de un componente sintáctico de regularización sigue siendo la norma en los sistemas de procesamiento de lenguas naturales.” E máis adiante continua expoñendo que “si las clases semánticas son subcategorizaciones de las sintácticas, entonces resulta claro que las transformaciones podrían formularse en términos de clases semánticas. [...] El argumento más fuerte de los que defienden un análisis sintáctico guiado semánticamente es la capacidad que tienen los hablantes para interpretar oraciones a partir de indicios semánticos.”

Co fin de que este procedemento se aproxime á realidade do uso da linguaxe, tratouse de que as transicións dunha lista de palabras a outra reflectan posibilidades reais de combinación entre palabras. Para iso, elaboráronse bases de datos que amosan a frecuencia de aparición dunhas palabras con outras, o que obrigou a someter extensas mostras textuais a análise informática, algo que tamén se aplica na confección de bases de datos, corpus de referencia e dicionarios de uso²⁸.

A crítica a esta maneira de proceder está en que os sistemas de encadeamento de palabras non son válidos como explicación do funcionamento da linguaxe, como demostrou Chomsky no que se pode considera-lo inicio da gramática moderna. Iso débese, en primeiro lugar, a que unha oración non ten nada que ver cunha cadea de palabras asociadas unhas con outras en virtude de probabilidades de transición (exemplo disto é *Verdes ideas incoloras dormen furiosamente*, famosa frase á que recorreu Chomsky para demostrar que as oracións sen sentido poden ser gramaticalmente correctas e que as secuencias de palabras altamente improbables tamén o poden ser); en segundo lugar, aprender unha lingua non é aprender a poñer palabras en orde, senón categorías de palabras, no que se apoia o coñecemento implícito dunha estrutura; e, en terceiro lugar, hai secuencias nas que o sistema de encadeamento obriga a lembra-lo que apareceu antes, algo que resulta problemático nun sistema de encadeamento desta natureza.

Por iso, se é importante a adecuada descrición e representación da estrutura oracional dunha lingua, tarefa que é cometido da gramática, non menos interese ten o léxico, que desempeña o papel de motor da xeración da devandita estrutura, pois, como sinala Wasow (1989:198-199) no epílogo ó libro de Sells (1989):

“es interesante observar que las teorías sintácticas actuales parecen converger en la idea de que la estructura oracional es predecible por lo general a partir del significado de las palabras, pues esto viene a coincidir con la

28. Certamente a análise textual como medio para extrae-lo léxico permite confeccionar bases de datos reais, moi útiles para a posterior confección de todo tipo de dicionarios –de uso, escolares, de segunda lingua, de colocacións, etc-. Un bo resultado da utilidade do corpus témola no dicionario *Collins-Cobuild*, dirixido por Sinclair (1987). Tamén se procedeu co exame da dispoñibilidade léxica dos falantes como resposta á audición dunha determinada secuencia de palabras, diso dan boa mostra as investigacións que se desenvolveron, por exemplo, no ámbito hispánico centradas en determinar cal é o léxico dispoñible dos falantes españois e hispanoamericanos. Vid., por exemplo, J. López Chávez e L. Rodríguez Fonseca, “Un modelo para la programación de la enseñanza del vocabulario”. *Actas del I Congreso Internacional sobre la enseñanza del español*. Madrid, CEMIP, 1992; J. López Chávez e C. Strassburger Frías, “Un modelo para el cálculo del índice de disponibilidad léxica individual”; H. López Morales (ed.): *La enseñanza del español como lengua materna*. *Actas del II Seminario Internacional sobre Aportes de la lingüística a la enseñanza del español como lengua materna*. Río Piedras, Editorial de la Universidad de Puerto Rico, 1991, pp. 91-112.

impresión ingenua que tienen muchos legos en lingüística. El profano tiende a identificar las lenguas con colecciones de palabras y a suponer, por ejemplo, que aprender una lengua nueva equivale a aprender su vocabulario. No tendría nada de particular, entonces, que los lingüistas hayan terminado por aproximarse a esta visión popular. Yo afirmo, por el contrario, que esto sería bien chocante, ya que la sabiduría convencional, examinada más de cerca, resulta irremisiblemente simplista. [...] Lo que ha ocurrido en torno a este cambio es que los estudiosos de la sintaxis han identificado los modos como las lenguas se apartan de la concepción ingenua. Han aislado ciertos tipos de información gramatical no predecible a partir de la semántica léxica y han elaborado teorías que permitan expresar esa información de manera concisa. Lo sorprendente (para los lingüistas) es lo poco que ha tenido que estipularse más allá del significado léxico”.

4. O papel da lexicografía e a codificación do léxico nos dicionarios para a traducción

Ata hai pouco tempo a lexicografía encargouse de confeccionar dicionarios coa pretensión de que os seus usuarios puidesen mellora-la súa expresión, aprende-lo significado dunha palabra e, no que respecta ós dicionarios bilingües, traducir dunha lingua a outra. Deste xeito, o dicionario considerouse algo independente e autónomo con respecto á gramática coa que non se relacionaba²⁹. Hoxe a situación cambiou notablemente, dado que os dicionarios se consideran parte do sistema xeral da lingua, o que ten a súa orixe nas modificacións feitas ó modelo da gramática xenerativa-transformacional que situaron o léxico nunha posición principal de control sobre o resto dos compoñentes do modelo gramatical. O resultado foi que o labor lexicográfico se entenda hoxe coma un traballo solidario no que interveñen tódalas ramas da lingüística. A razón estriba en que as entradas léxicas que conteñen os dicionarios teñen que achegar grandes cantidades de información, que teñen que ver coa fonética, a morfoloxía e a etimoloxía, a sintaxe, a semántica e, incluso, a pragmática; e isto é máis necesario aínda nos programas de traducción automática, porque os dicionarios electrónicos concibidos para esta función non contan coa complicitade do consultor humano do dicionario en soporte papel que pode supli-las carencias que poida presenta-la obra, como expón Gutiérrez Cuadrado (1994:638)³⁰. De maneira que, como sinalan

29. Como apunta Gutiérrez Cuadrado (1994:640) a discusión aberta sobre a relación entre a gramática e o dicionario é un episodio recente da lingüística e arrinca de varios feitos, como son as necesidades do ensino de segundas linguas, a aparición e consolidación da teoría lexicográfica ou metalexigrafía, o desenvolvemento das teorías gramaticais e o desenvolvemento da traducción automática.

30. É o problema que presentan definicións pouco precisas como, por exemplo, as de adxectivos de relación do tipo ‘pertencente ou relativo a’ que non conteñen máis cá explicación da relación que o adxectivo ten coa súa base substantiva da que deriva (Jiménez Ríos 1998-1999).

Hutchins e Somers (1995:44 e 93), este tipo de dictionarios ten que recoller moita información e ten que facelo de forma exhaustiva e explícita:

“el lexicon o diccionario de un programa de TA debe suministrar la información necesaria para el procesamiento semántico y sintáctico: la categoría gramatical (nombre, verbo, etc.), los rasgos de subcategorización (esto es, a qué “subcategoría” pertenece una palabra, por ejemplo, verbo transitivo o intransitivo, sustantivo masculino o femenino), además de información semántica (nombre animado, verbo que exige un sujeto humano). [...] Los lexicones que se emplean en los sistemas de TA deben incluir también la información que resulta obvia: toda la información que precisen los programas debe hallarse disponible. Los lexicones propios de un sistema de TA incluyen información sobre las partes del discurso, si bien de forma mucho más explícita: [...] los significados –si están cifrados– deben adoptar una forma que permita al ordenador utilizarlos, como, por ejemplo, mediante rasgos semánticos o restricciones de selección; las definiciones que se ofrecen en los diccionarios manuales no resultan prácticas para los programas de TA”.

Isto correspóndese ben co xeito de proceder dos modelos de análise lingüística, que destacan o papel fundamental do léxico e resaltan a importancia do *verbo* como palabra clave para a construción dunha oración (vid. Lewis (1992:92)): o verbo leva implícita unha información que determina os elementos cos que ten e pode combinarse, pois, como apunta Sells (1989:196),

“la estructura clausal es ampliamente predecible a partir de la semántica de predicados. Es decir, si se sabe qué significa un verbo (o un adjetivo predicativo o un nombre), pueden predecirse muchas cosas sobre todo lo que aparecerá en la cláusula que aquél encabeza. Las reglas gramaticales sólo se necesitan para establecer ciertas generalizaciones de gran alcance lingüístico sobre el modo de unir las piezas oracionales y resolver excepciones aparentes a los patrones habituales. La mayor parte de lo que se estipulaba en las gramáticas de las viejas teorías ahora se considera en función de la semántica léxica”.

O anterior outorga un papel fundamental ó léxico e cristalízase no tratamento que a lexicografía dá ás palabras. Este xeito de proceder é o resultado da distinción, xa clásica, entre *sintaxe gramatical* e *sintaxe léxica*, enunciada polo gramático venezolano Andrés Bello, a mediados do século XIX, e que o seu contemporáneo, o filólogo colombiano Rufino José Cuervo levou á práctica na preparación do seu diccionario de construción e réxime³¹. A primeira

31. *Diccionario de construcción y régimen de la lengua castellana* (Bogotá, Instituto Caro y Cuervo, reedición y continuación, 1994). Un dos continuadores da obra do filólogo colombiano, Porto Dapena (1980:152), afirma que “el DCR tiene por objeto el estudio de aquellas palabras que ofrezcan alguna peculiaridad sintáctica, sea por los contextos de categoría en que son susceptibles de aparecer, sea por los cambios de categoría que pueden experimentar, sea, en fin, por los distintos matices de relación que son capaces de establecer entre palabras u oraciones.”

ocúpase das estruturas oracionais e a segunda dos tipos de complementación das unidades léxicas individuais³². Entón, se a traducción automática ten que conceder importancia ás unidades léxicas, é lóxico que a maioría dos sistemas se baseen na *gramática de dependencias* –con modificacións procedentes doutras teorías gramaticais–, dado que o interese está na dependencia dunhas unidades doutras no eixe sintagmático³³ e é incompreensible o significado dunha palabra sen a descrición das súas posibilidades combinatorias. A isto úneselle a importancia que ten a frecuencia de uso dunha voz e as posibilidades de que dúas formas coaparezan nun mesmo texto, o que fai do recurso ós *corpus lingüísticos*³⁴ outro dos piares fundamentais da traducción automática, pois, como sinala Pascual (1996:149-150),

“para saber qué palabras se emplean en una lengua y con qué sentido, cuáles son propias de un determinado lugar, cuáles se consideran formales, anticuadas o rabiosamente neológicas, cuáles se reputan como cultas, es decir, para conocer el uso real que los hablantes hacen de una lengua, no podemos suponer que una sola persona, aunque se trate de un gramático muy afamado, y ni siquiera un grupo de ellas, por muy cultos que sean sus integrantes, pueden dar cuenta por medio de la introspección de la realidad de la lengua. Estamos más cerca de esa realidad de los usos lingüísticos, si somos capaces de allegar una muestra representativa –aunque, naturalmente, parcial– de ellos, por medio de una serie de textos, a cuyos datos tendremos acceso cómodo si somos capaces de someterlos a un tratamiento informático. Es lo que se conoce como un corpus textual”;

o que leva a facer que o ordenador estea preparado de tal xeito que o programa deseñado permita, como sinala Pinker (1999:235-236), proceder do seguinte xeito:

“la entrada léxica más frecuente de un verbo parece inducir al analizador a buscar los participantes que le son apropiados. [...] Las palabras también pueden colaborar en este proceso indicándole al analizador con qué otras

32. A partir de aquí as investigacións desenvolvidas no ámbito da lexicografía hispánica desde mediados do século XX ocúpase das relacións do léxico e a sintaxe e da necesidade de introducir e distinguir adecuadamente o que é información para o uso dunha palabra do que é o seu significado: véxase Casares (1950), Haensch (1982) e Ahumada Lara (1989). Son temas interesantes o dos verbos e os seus correspondentes substantivos abstractos, estudiado por Lázaro Carreter (1971), ou o da delimitación do contorno, como precisa Seco (1987).

33. Isto explica a utilidade da *lexicografía de valencias*, centrada na *valencia* das palabras que Martin Mingorance (1994:29) explica do seguinte xeito: “la *valencia*, término metafórico muy ajustado para describir la naturaleza combinatoria, es una propiedad idiosincrásica de las unidades léxicas de sentido pleno que permite la producción e interpretación del discurso.”

34. Sobre as características dos corpus, o seu deseño, tipoloxía e marcación, vid., por exemplo, M. Alvar Ezquerro e J. A. Villena Ponsoda (eds.), *Estudios para un corpus del español*. Málaga, Universidad, 1994.

palabras suelen aparecer dentro de un mismo sintagma. [...] Un analizador pertrechado con un buen aparato estadístico puede optar con cierta garantía por una determinada estructura cuando ésta es la que se usa con mayor frecuencia. El analizador humano es bastante sensible a las probabilidades de asociación entre pares de palabras; así, muchas vías muertas resultan seductoras porque contienen pares de palabras asociadas, como *bebía vino*. Al margen de que el cerebro aproveche o no las propiedades estadísticas del lenguaje, no cabe duda de que los ordenadores sí lo hacen. En los laboratorios de AT&T e IBM se emplean ordenadores para tabular millones de palabras extraídas de textos de prensa, en la esperanza de que si se suministra a los analizadores las frecuencias de uso de cada palabras y las frecuencias de asociación entre palabras, éstos podrán resolver ambigüedades convenientemente”³⁵.

Habrá, polo tanto, que conferirlle a cada entrada léxica toda a información idiosincrásica necesaria, así como a súa frecuencia de aparición, o que ten que reflectirse na maneira en que se recolla a devandita unidade no diccionario electrónico, pois, como tamén advirte Pacual (1996:149),

“lo que no puede hacer una máquina es tomar una decisión con respecto a la organización que los diccionarios de uso dan en la actualidad a las palabras”;

se ben, na actualidade existe un tipo particular de diccionarios –os destinados ó ensino e a aprendizaxe dunha lingua– nos que se está empezando a insistir na necesidade de amosa-la capacidade combinatoria das palabras.

5. Melloras na confección de diccionarios electrónicos de traducción automática

As melloras na creación de diccionarios electrónicos destinados á traducción automática pasan por poñe-la énfase nos seguintes tres aspectos, relacionados entre si (que, mesmo, poden ser de utilidade para os propios usuarios dos programas de traducción, capaces de completa-los diccionarios do sistema ó tempo que avanzan no seu traballo). A facilidade para que estes aspectos se poñan en práctica débese a que os novos soportes de diccionarios non presentan limitacións canto ó espacio:

a) A detallada e completa información da microestrutura. O léxico dunha lingua, do que teñen competencia os falantes desa lingua, é, como xa se apuntou, unha especie de diccionario no que as entradas especifican as propiedades idiosincrásicas dos seus elementos. Fronte ó que sucede nun diccionario tradicional –xeralmente en soporte papel, aínda que hoxe tamén en formato CD-ROM– no que aparecen formas bases ou canónicas como entradas, os diccionarios destinados a servir para a traducción automática caracterí-

35. O resaltado en cursiva é meu.

zanse por presentar tódalas formas do paradigma da base ou forma canónica (claro que non faltan programas con analizadores morfolóxicos que permiten xera-la forma que se requira). En cada entrada ou cabeceira –que será unha para cada acepción ou sentido en que poida aparecer esa voz, o que, no caso dos dicionarios bilingües, evita os problemas da ambigüidade de transferencia ou traducción– detallarase, ademais dos seus trazos morfolóxicos, sintácticos e semánticos, o comportamento sintagmático da devandita forma léxica (vid. Lewis (1992:80) e Meijs (1992:142)).

b) O tratamento das unidades léxicas formadas por máis dunha palabra, polas dificultades que as devanditas unidades presentan para a traducción. A carón da unidade palabra, hai outros elementos que teñen que ser recollidos nos dicionarios, como son os compostos e as unidades pluriverbais, entre as que se poden establecer dous grupos: a) as *estructuras fixas*, caracterizadas por teren un significado non transparente ou opaco (locucións, modismos, frases feitas, etc.; en definitiva, todo o que pertence ó que Coseriu (1981:297) denominou “discurso repetido”); e b) as *colocacións*, que posúen un significado transparente a partir dos elementos constituíntes. Para o tratamento deste último tipo de unidades pluriverbais faise necesario tamén recorrer ós corpus, concordancias e bases de datos co fin de ver cáles son as palabras que adoitan aparecer xuntas³⁶.

c) A reutilización previa de traducións xa feitas para elaborar a partir delas unha *base temática de datos textuais* que permita facer un uso de ocorrencias, por unha banda reais e, pola outra, contextualizadas. Algúns dos problemas insalvables que presenta a traducción automática –ambigüidade, homografía, etc.– desembocaron no recurso á traducción por exemplos e á memoria de traducción (vid. Hutchins e Somers (1995:184-186); Gómez Guinovart (1996:101)), o que, gracias ó progresivo incremento da capacidade dos ordenadores, se presenta como unha solución útil.

6. Como conclusión

Do que quedou exposto ó longo destas páxinas pódense extraer dúas conclusións: por unha banda, a necesidade de combinar teoría e práctica no traballo da traducción automática (é dicir, imponse facer unha lingüística e lexicografía aplicadas á traducción, en xeral, e á traducción automática, en particular); por outra banda, a urxencia irrenunciabile de combinar gramática e léxico ou sintaxe e lexicografía no deseño dos dicionarios electrónicos para os programas de traducción automática, aproveitando o camiño xa percorrido polos dic-

36. Un caso problemático é, por exemplo, o que presentan os complementos dun nome susceptibles de aparecer como adxectivos de relación ou como sintagmas preposicionais do tipo “de+substantivo”. Refírome a casos como *batalla campal*, *incendio forestal*, *vea xugular*, etc. que rexeitan a estrutura preposicional como substituta do adxectivo.

cionarios pioneiros non electrónicos (especialmente, os de ensino e aprendizaxe de segundas linguas) que inciden nesa conexión para profundizar na sintaxe da palabra e que son un bo exemplo da aplicación lingüística teórica na actividade lexicográfica. Se isto ten que ver co xeito de face-la investigación no ámbito da traducción automática, hai outro feito secundario, pero non menos importante, e é que os avances que a disciplina poida experimentar con iso teñen que verse plasmados en todo tipo de programas destinados á traducción automática –tanto naqueles creados polos grandes organismos internacionais e as grandes empresas multinacionais, coma nos destinados ó gran público para o seu uso en ordenadores persoais, que son, en última instancia, os que serven para medi-las melloras que se vaian producindo³⁷. Con iso deixará de ser un reclamo a crítica que facía Soler (1994:384) cando sinalaba que

“el usuario medio de productos lingüísticos, por su parte, tiene elevadas expectativas sobre las funcionalidades y la calidad de estos productos, expectativas que generalmente no se corresponden con lo que dichos productos pueden ofrecer”;

quizais porque non estiveron nin están ben definidas as funcións para as que foron deseñados, ou porque pretenden cumprir ó mesmo tempo con moitas funcións. Por iso, a determinación de obxectivos e a utilización de programas para uns fins específicos contribuirán a que se poidan detecta-los avances experimentados pola disciplina³⁸.

37. Cf. a análise dos tradutores para o seu uso en ordenadores persoais que fan Amores Carredano (1989), Lewis (1992:101), Gómez Guinovart (1993, 1994) e Díez (1995).

38. Ata hoxe o uso deste tipo de programas só serve para obter traducións en borrador, o que non deixa de ter utilidade, sobre todo, para a traducción científica e técnica, como apunta Gómez Guinovart (1996:101) ó sinalar que “en resumo, se ben é certo que a calidade dos resultados da TSIH [traducción semiautomática con intervención humana] non sería aceptable pola maioría dos compradores actuais de traducións, o rendemento destes produtos informáticos é moito mais satisfactorio cando se trata de producir borradores de traducción urxentes e a baixo prezo”, traducións que permiten adquirir un coñecemento superficial do texto de orixe.

REFERENCIAS BIBLIOGRAFICAS

- AHUMADA LARA, I. 1989. *Aspectos de lexicografía teórica*, Granada: Universidad, 1989.
- ALVAR EZQUERRA, M. 1996. "Los diccionarios del español en nuestros días", *Cuadernos Cervantes*, 11, pp. 9-14, 1996
- AMORES CARREDANO, J. G. 1989. "El sistema de traducción asistida por ordenador MicroCat. Sus ventajas y limitaciones", *Revista Española de Lingüística Aplicada*, 5, pp. 63-72, 1989.
- BÁEZ SANJOSÉ, V. 1994. "Algunos criterios lingüísticos seguidos en la elaboración de un diccionario multilingüe del español estándar", *Actas del Congreso de la Lengua española*, Madrid: Instituto Cervantes, pp. 293-306.
- BERGENHOLTZ, H. e TARP, S. (ed.) 1995. *Manual of Specialised Lexicography*, Amsterdam: John Benjamins, 1995.
- CABRÉ, M. T. 1993. *La terminología. Teoría, métodos y aplicaciones*, Barcelona: Antártida/Empúries, 1993.
- CARBONELL, J. 1994. "El procesamiento del lenguaje natural, tecnología en transición". *Actas del Congreso de la Lengua española*, Madrid: Instituto Cervantes, pp. 247-250, 1994.
- CASARES, J. 1950. *Introducción a la lexicografía moderna*, Madrid: CSIC, 3ª edición, 1992.
- COSERIU, E. 1981. *Lecciones de lingüística general*, Madrid: Gredos, 1981.
- DEMONTE, V. 1990. "Transitividad, intransitividad y papeles temáticos", en Demonte, V. e Garza Cuarón, B. (eds.), *Estudios de lingüística de España y México*, México: UNAM e El Colegio de México, pp. 115-150, 1990.
- 1991. *Teoría sintáctica: De las estructuras a la rección*, Madrid: Síntesis, 1991.
- DÍEZ, J. 1995. "Traductores para Windows", *PCWord*, 115, pp. 330-340, 1995.
- DOOD, W. S. 1994. "La lexicoinformática", en Hernández, H. (coord.), *Aspectos de Lexicografía contemporánea*, Barcelona: Bibliograf, pp. 121-137, 1994.
- FILLMORE, CH. 1968. "The Case for Case", en Bach, E. e Harms, R. (eds.), *Universals in Linguistic Theory*, Nueva York: Holt, pp. 1-90, 1968.
- FRIEDMAN, J. 1971. *A Computer Model of Transformational Grammar*, Nueva York: American Elsevier, 1971.
- GÓMEZ CLEMENTE, X. M. 1996. "Recursos para traductores na rede Internet", *Viceversa, Revista Galega de Traducción*, 2, pp. 83-98, 1996.
- GÓMEZ GUINOVART, J. 1993. "El inglés más cerca. Cinco programas de TA inglés-castellano", *Binary*, 51, pp. 51-67, 1993.
- 1994. "Programas de traducción para Windows", *BYTE España*, 1, pp. 132-145, 1994.
- 1996. "Traducción automática e traducción asistida por ordenador: aspectos terminológicos e tipoloxía", *Viceversa, Revista Galega de Traducción*, 2, pp. 99-103, 1996.
- GRISHMANN, R. 1991. *Introducción a la lingüística computacional*, traducción de A. Moreno Sandoval, Madrid: Visor, 1991.

- GROSS, M. 1972. "Dictionnaires électroniques et traduction automatique". *Langages*, 116, pp. 48-58, 1972.
- 1991. "La elaboración de los diccionarios electrónicos del francés", en Vidal Beneyto, J. (coord.), *Las industrias de la lengua*, Madrid: Sánchez Ruipérez, pp. 147-184, 1991.
- GUTIÉRREZ CUADRADO, J. 1994. "Gramática y diccionario", *Actas del Congreso de la Lengua española*, Madrid: Instituto Cervantes, pp. 637-656, 1994.
- 1999. "Notas a propósito de la ejemplificación y la sinonimia en los diccionarios para extranjeros", en Vila, M. N. et alii (eds.), *Así son los diccionarios*, Lleida: Universidad de Lleida, pp. 77-95, 1999.
- HAENSCH, G. et alii 1982. *La lexicografía. De la lingüística teórica a la lexicografía práctica*, Madrid: Gredos, 1982.
- HUTCHINS, W. J. e SOMERS, H. L. 1995. *Introducción a la traducción automática*, traducción coordinada por J. Abaitua, Madrid: Visor, 1995.
- JIMÉNEZ RÍOS, E. 1998-1999. "El tratamiento de los adjetivos de relación en el diccionario de la Academia", *Revista de Lexicografía*, 5, pp. 89-114, 1999.
- JORDÁN, A. G. 1994. "Lenguas y tecnologías de la información", *Actas del Congreso de la Lengua española*, Madrid: Instituto Cervantes, pp. 251-258, 1994.
- KATZ, J. e FODOR, J. 1963. "The Structure of a Semantic Theory", *Language* 39, pp. 170-210, 1963.
- LÁZARO CARRETER, F. 1971. "Transformaciones nominales y diccionario", *Estudios de Lingüística*, Barcenola: Crítica, 1980, pp. 73-81, 1971.
- LEWIS, D. 1992. "Computers and Translation", en Butler, C. (ed.), *Computers and Written Texts*, Oxford: Blackwell, pp. 75-113, 1992.
- MARTÍN MINGORANCE, L. 1994. "La lexicografía de valencias", en Hernández, H. (coord.), *Aspectos de Lexicografía contemporánea*, Barcelona: Biblograf, pp. 29-37, 1994.
- MEIJS, W. 1992. "Computers and Dictionaries", en Butler, C. (ed.), *Computers and Written Texts*, Oxford: Blackwell, pp. 141-165, 1992.
- NEWMARK, P. 1992. *Manual de Traducción*, Madrid: Cátedra, 1992.
- PASCUAL, J. A. 1996. "Escándalo o precaución: sobre el futuro de nuestra lengua", en Marqués de Tamarón (ed.), *El peso de la lengua española*, Valladolid: Fundación Duques de Soria, pp. 135-171, 1996.
- PINKER, S. 1994. *El instinto del lenguaje*, traducción de José Miguel Igoa, Madrid: Alianza Universidad, 1994.
- POTTIER, B. 1961. "Les travaux lexicologiques préparatoires a la traduction automatique", en *Cahiers de Lexicologie*, 3, pp. 200-205, 1961.
- SABAH, G. 1991. "El tratamiento automático del lenguaje escrito", en Vidal Beneyto, J. (coord.), *Las industrias de la lengua*, Madrid: Sánchez Ruipérez, pp. 39-54, 1991.
- SANTANA, O. et alii 1996. "Diccionarios en soportes informáticos", en *Cuadernos Cervantes*, 11, pp. 68-77, 1996

- SECO, M. 1987. "El "contorno" en la definición", en *Estudios de lexicografía española*, Madrid: Paraninfo, pp. 35-45, 1987.
- SELLS, P. 1989. *Teorías sintácticas actuales*, versión española de Ramón Cerdá, Barcelona: Teide, 1989.
- SINCLAIR, J. (ed.) 1987. *Looking Up. An Account of the COBUILD Project in Lexical Computing*, London-Glasgow: Collins ELT, 1987.
- SOLER, J. 1994. "Lengua y tecnología en la política de investigación y desarrollo de la Comunidad europea", *Actas del Congreso de la Lengua española*, Madrid: Instituto Cervantes, pp. 383-396, 1994.
- SUBIRATS RÜGGERBERG, C. 1994. "Sistema de diccionarios electrónicos del español", *Actas del Congreso de la Lengua española*, Madrid: Instituto Cervantes, pp. 316-330, 1994.
- WASOW, T. 1989. "Epílogo", en Sells, P. 1989. *Teorías Sintácticas Actuales*, Barcelona: Teide, pp. 187-199, 1989.
- ZGUSTA, L. 1971. *Manual of Lexicography*, La Haya: Mouton, 1971.