

# LA LINGÜÍSTICA COMPUTACIONAL Y SUS APLICACIONES

Milka VILLAYANDRE LLAMAZARES  
Universidad de León

## INTRODUCCIÓN

Aproximadamente entre la década de los años 50 y la de los 60 va a tener lugar una serie de acontecimientos dentro y fuera de la lingüística, cuyo resultado va a ser el nacimiento no sólo de la *lingüística aplicada* (Congreso de Nancy, 1964), sino también de otra serie de ciencias más o menos próximas y a menudo relacionadas entre sí, sobre todo por el hecho de girar alrededor del lenguaje como objeto de estudio a través de alguna de sus múltiples facetas. Se trata de la *lingüística computacional*, *lingüística matemática*, *lingüística algebraica*, *inteligencia artificial*, *cibernética*, *neurolingüística*, *psicolingüística*, *sociolingüística*, etc. Son todas ellas disciplinas científicas que surgen como consecuencia de unas mismas circunstancias y que comparten una serie de características comunes.

Entre esas circunstancias desencadenantes, hay que destacar, por una parte, las necesidades surgidas durante y a partir de la II Guerra Mundial. Se trata de causas o factores externos a la propia lingüística, de orden económico, político y social que, no obstante, van a acarrear el desarrollo de nuevos campos y teorías dentro de la lingüística, como respuesta a los problemas que se planteaban: i) nuevas exigencias de comunicación, gracias al avance en el campo de las telecomunicaciones (teoría de la comunicación); ii) importancia de las traducciones (empleo de ordenadores: traducción automática); iii) manipulación de grandes cantidades de información (teoría de la información, Shannon, 1949); iv) aprendizaje y enseñanza de ciertas lenguas, de amplia difusión o minoritarias (lingüística aplicada, psicolingüística).

Por otra parte, desde dentro de la propia lingüística, también había hechos que apuntaban a la necesidad de un cambio. Se trata de factores relacionados con nuevos enfoques y planteamientos en la lingüística, la distancia entre el trabajo teórico y la solución a problemas materiales y la necesidad de datos concretos que justifiquen las posturas teóricas.

Además, no hay que olvidar el papel central que va a desempeñar en la mayoría de estos campos el ordenador, que también aparece en estas fechas.

Pues bien, dentro de este contexto dominado por la creciente necesidad de adquirir, procesar y transmitir información es donde surgen los trabajos de lo que se conoce como *lingüística computacional*. En concreto, en este artículo nos centraremos en la faceta aplicada de la *lingüística computacional*. Para ello, en primer lugar, abordaremos su estatus en relación con la

*lingüística aplicada* (hay que tener en cuenta que, además de surgir en el mismo período de tiempo, los problemas relacionados con las necesidades de la *traducción automática*, uno de los campos pioneros de la lingüística computacional, en ese primer momento cercano a su nacimiento, se llegaron a identificar con la misma *lingüística aplicada*); en segundo lugar, esbozaremos una rápida delimitación del campo de la lingüística computacional frente a otros como el *procesamiento del lenguaje natural*, la *inteligencia artificial*, la *lingüística informática*, las *industrias de la lengua* o la *ingeniería lingüística*, por mencionar sólo algunos de los términos con los que se la relaciona; por último, a partir de la diferencia que se suele establecer dentro de la lingüística computacional entre 'objetivos científicos' y objetivos prácticos<sup>1</sup>, nos acercaremos a las aplicaciones de la lingüística computacional a través de la comparación de los modelos propuestos por diferentes autores.

## 1. LA LINGÜÍSTICA APLICADA

La *lingüística aplicada*<sup>2</sup> como disciplina independiente surge en Estados Unidos en los años 40-50 en centros académicos interesados por la enseñanza y el aprendizaje de idiomas (segundas lenguas). Si bien en un principio se podía equiparar prácticamente a la enseñanza de lenguas o a la lingüística matemática a través de la traducción automática<sup>3</sup>, el rápido desarrollo del campo, con la sucesiva inclusión de nuevas áreas de estudio en su seno, la ha obligado a constituirse como un dominio aparte, sobre todo porque pronto entró en relación con otras áreas dispares, como la lexicografía, la traducción, el tratamiento automatizado de textos, la enseñanza de la lengua materna, etc. Precisamente, en relación con estas áreas aparecieron nuevos campos interdisciplinares: psicolingüística, sociolingüística, antropología lingüística, etc. Y, aunque es imposible dar una fecha exacta de cuándo tuvieron lugar tales confluencias en torno a una única disciplina, lo cierto es que se generalizan poco después de la II Guerra Mundial<sup>4</sup>. No obstante, se suele mencionar como hito de las confluencias el

<sup>1</sup> Cf. GRISHMAN, R.: *Computational linguistics*, Cambridge University Press, Cambridge, 1986. Traducido al español por Antonio Moreno Sandoval como *Introducción a la lingüística computacional*, Visor, Madrid, 1991. Salvo indicación en contra, cito por la traducción al español.

<sup>2</sup> En adelante, se aludirá a la Lingüística Aplicada como LA.

<sup>3</sup> Hay que tener en cuenta que el "Coloquio de Nancy" (1964) estuvo promovido por el "Groupe de traduction automatique" de dicha universidad (Vid. SLAMA-CAZACU, T.: *Linguistique Appliquée: Une introduction*, La Scuola, Brescia, 1984:44).

<sup>4</sup> "En definitiva, más que del nacimiento repentino de una disciplina (o subdisciplina, en relación con la 'gran' lingüística), deberíamos hablar de *continuum* de confluencias en tradiciones diferentes, que representan una *dimensión aplicada* de la lingüística" (PAYRATÓ, LL.: *De profesión, lingüista. Panorama de la lingüística aplicada*, Ariel, Barcelona, 1998:29).

I Coloquio Internacional de Lingüística Aplicada, celebrado en Nancy, Francia, en 1964<sup>5</sup>.

La definición más elemental de LA es la de "subdisciplina o rama lingüística que consiste en aplicar teorías lingüísticas a un dominio práctico"<sup>6</sup>, en analogía con las ramas aplicadas de otras disciplinas como la psicología, las matemáticas, etc. Pero según T. Slama-Cazacu, no se puede definir la LA como una "simple 'application des théories linguistiques à un domaine pratique'"<sup>7</sup>. Además señala que no goza de un estatus autónomo, sino dependiente de la lingüística: "Si l'on considère la *linguistique* comme une science dont le contenu est très vaste et qui a pour objet l'étude de la langue en général, la LA pourrait apparaître comme l'*une* des branches de la *linguistique*"<sup>8</sup>. Por otra parte, como toda rama de la lingüística, comprende un aspecto aplicado y otro teórico<sup>9</sup>, es decir, elabora sus propios modelos teóricos que, a su vez, pueden revertir en la propia investigación teórica dentro de la propia lingüística: *Par conséquent, je n'estime pas que la LA soit 'la linguistique théorique (ou des théories de la linguistique) appliquée à un certain domaine pratique', mais qu'elle implique une élaboration théorique (ou, au moins, aussi une réélaboration des théories linguistiques existantes), en vue de trouver des solutions à des problèmes pratiques*<sup>10</sup>. En términos parecidos se expresa Cabré: *El progrés importantíssim de la teoria no ens faci magnificar-la i menystenir l'aplicació; que l'aplicació no es consideri autosuficient lluny de la recerca fonamental de la teoria. La lingüística aplicada és lingüística, i això significa que ha de treballar amb un sistema de regles i unitats que descriuen el comportament humà, tal com diu la teoria. Dels avenços de la teoria se n'hauria de beneficiar l'aplicació. Els resultats de la lingüística aplicada haurien de servir per reorientar, quan calgués, la recerca teòrica*<sup>11</sup>.

Es decir, según esta primera definición, la LA se entendería tal y como aparece reflejado en el gráfico siguiente<sup>12</sup>:

<sup>5</sup> Cf. T. SLAMA-CAZACU, *op. cit.*, p. 43.

<sup>6</sup> LL. PAYRATÓ, *op. cit.*, p. 18.

<sup>7</sup> T. SLAMA-CAZACU, *op. cit.*, p. 14.

<sup>8</sup> *Id.*, p. 17.

<sup>9</sup> *Id.*, p. 18.

<sup>10</sup> *Id.*, p. 95.

<sup>11</sup> CABRÉ, M.T. *et alii* (coord.): *La lingüística aplicada. Noves perspectives. Noves professions. Noves orientacions*, Publicacions Universitat de Barcelona, Fundació Caixa de Pensions, 1990:10.

<sup>12</sup> Aquí entendemos 'aplicada' y, por lo tanto, LA, en sentido estricto, según la definición tradicional de aplicación de conocimientos lingüísticos a la solución de problemas prácticos de naturaleza total o parcialmente lingüística (*Vid.* CABRÉ, M.T. y PAYRATÓ, LL.: "La lingüística aplicada avui" en CABRÉ, M.T. *et al.*, *op. cit.*, 1990, p. 14).

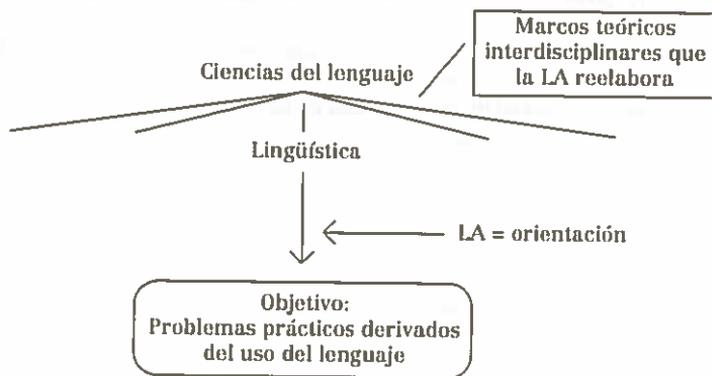
Lingüística (Teórica)



← LA = aplicación

Práctica

Pero, para definir la LA dentro del sistema de las ciencias, es más útil elegir como criterio no el de su relación con la lingüística teórica o con otras ramas de la lingüística, sino el de la *finalidad* de las diversas investigaciones lingüísticas: "La LA est le domaine caractérisé par une étude de la langue faite à une fin nettement appliquée, donc servant directement et spécifiquement à réaliser un certain objectif pratique"<sup>13</sup>. De esta forma, no se separa la LA de la lingüística y la LA no significa aplicar la lingüística: *En definitiva, pues, la lingüística aplicada puede concebirse como una orientación o dimensión de la investigación lingüística, propia de todos los campos de estudio incluidos en las ciencias del lenguaje, que, partiendo de marcos (teóricos) interdisciplinarios, persigue como objetivo la resolución de problemas (prácticos) derivados de la praxis lingüística, del uso lingüístico en que se concreta la capacidad humana del lenguaje*<sup>14</sup>. Es lo que reflejamos en este otro gráfico<sup>15</sup>:



<sup>13</sup> T. SLAMA-CAZACU, *op. cit.*, p. 19.

<sup>14</sup> LL. PAYRATÓ, *op. cit.*, p. 24.

<sup>15</sup> De acuerdo con esta segunda definición de LA como orientación, estamos ante el sentido amplio de 'aplicada', ya que las bases teóricas de la LA no son sólo las de la lingüística tradicional, sino también las de disciplinas externas, con las que la lingüística intersecciona a la hora de abordar problemas prácticos (*Vid. CABRÉ, M.T. y PAYRATÓ, LL., op. cit.*, p. 15).

Y es que, por una parte, el objeto de ambas disciplinas es el mismo<sup>16</sup>, la lengua, y su finalidad, convergente, un mayor conocimiento del fenómeno de la lengua<sup>17</sup>. Sin embargo, la especificidad de la LA reside en: i) su orientación específica hacia la *aplicación*, ya que su finalidad se relaciona con situaciones concretas de la vida<sup>18</sup>; ii) su necesidad de principios generales teóricos, pero siempre en relación con la realidad<sup>19</sup>; por este motivo, no es válida cualquier teoría. Según Slama-Cazacu, las exigencias de una teoría lingüística para la LA son dos: que considere la lengua como un fenómeno concreto —excepto en algunas áreas como la traducción automática o la lingüística computacional—, y que considere la lengua un fenómeno dinámico, en continuo cambio<sup>20</sup>; iii) además, los resultados prácticos deben estar marcados igualmente por un rasgo muy concreto, muy *específico*<sup>21</sup>.

Y, por otra parte, la LA se caracteriza por su *interdisciplinariedad*, ya que es un dominio científico que utiliza los conocimientos de la lingüística pero también de otras disciplinas. De ahí que se la suela considerar un dominio pluridisciplinar, desde el momento que la lengua como fenómeno humano exige un acercamiento interdisciplinar<sup>22</sup>.

Esta caracterización de la LA en torno a los conceptos de *orientación* o *finalidad práctica* e *interdisciplinariedad*, no es fortuita, sino que obedece a dos tendencias que conviven en la ciencia contemporánea: i) una tendencia hacia la especialización, como consecuencia del descubrimiento de nuevos aspectos<sup>23</sup>; de esta forma la LA es un campo abierto a nuevos temas, según las nuevas exigencias de la vida o los nuevos descubrimientos científicos<sup>24</sup>; ii) un acercamiento inter o pluridisciplinar, motivado por la necesidad de contar con medios complejos para el estudio de un fenómeno, aunque manteniendo la unidad dentro de la variedad: *En ce qui concerne la relation circulaire fondamentale-appliquée en linguistique, il faut donc souligner également (...) le caractère interdisciplinaire implicite du champ de problèmes de la LA et en définitive l'obligativité d'aborder ces problèmes de manière interdisciplinaire. Les objectifs de la LA sont nombreux, son champ est vaste, tout comme la réalité où l'on rencontre des faits de langue.*

---

<sup>16</sup> Para los elementos que definen el objeto de la LA, ver T. SLAMA-CAZACU, *op. cit.*, pp. 96-99. En concreto, «l'objet de la LA consiste donc à recueillir ou à réélaborer, à classer et systématiser, par des recherches scientifiques spéciales, les faits de langue et les principes de fonctionnement de la langue, nécessaires à des buts pratiques» (*Id.*, p. 99).

<sup>17</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, p. 22.

<sup>18</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, p. 24.

<sup>19</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, pp. 24-25.

<sup>20</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, pp. 132-135.

<sup>21</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, p. 25.

<sup>22</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, p. 28.

<sup>23</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, p. 26.

<sup>24</sup> *Vid.* T. SLAMA-CAZACU, *op. cit.*, p. 102.

*les réalisations concrètes de la langue dans la communication et, finalement, les sciences avec lesquelles elle a des liens implicites à savoir les branches d'activité humaine: neurologie et cybernétique, enseignement des langues étrangères et lexicologie, sociologie, sciences économiques et juridiques, sociolinguistique et domaine des traductions humaines ou automatiques...*<sup>25</sup>.

Así pues, la lingüística aplicada es sólo una dimensión particular de la lingüística, una orientación posible de cada una de las ramas de la lingüística<sup>26</sup>, en este caso, una orientación práctica y concreta encaminada a la solución de problemas reales, que se corresponde con un concepto de *aplicada en sentido estricto*, es decir, se aplican o utilizan "coneixements lingüístics amb vista a la solució d'uns problemes pràctics de naturalesa fonamentalment o parcialment lingüística"<sup>27</sup>, "i en correspondència la LA es pot concebre aleshores com el conjunt de les aplicacions de la lingüística, amb una denominació que ja s'utilitza força sovint"<sup>28</sup>.

Pero, por otra parte, sus bases teóricas son, no sólo las de la lingüística tradicional sino también las de disciplinas externas, con las que la lingüística intersecciona a la hora de abordar problemas prácticos<sup>29</sup>, que sería una *LA en sentido amplio*, "constituït per la intersecció entre el camp d'estudi (tradicional) de la llengua i d'altres zones properes"<sup>30</sup>. Este es el marco teórico donde la LA ha acabado trabajando: *Sea como fuere, el resultado final de esta controversia es que, hoy en día, la lingüística aplicada no puede entenderse ya como una simple rama o especialidad 'interna' (como la fonología, la lexicología, o la sintaxis, por ejemplo) susceptible de ser 'trasladada', ni tampoco como una (o algunas) de las intersecciones o interdisciplinas, sino que debe conceptualizarse como una dimensión, presente en todas las ramas de las llamadas ciencias del lenguaje<sup>31</sup> ...y tiene una finalidad práctica consistente en resolver los problemas provocados por los procesos comunicativos propios de las sociedades actuales<sup>32</sup>.*

## 2. LA LINGÜÍSTICA COMPUTACIONAL<sup>33</sup> COMO DISCIPLINA APLICADA

Como ya hemos mencionado, la lingüística computacional surgió en la década de los 50 en Estados Unidos, en ámbitos ligados sobre todo a

<sup>25</sup> *Id.*, p. 78.

<sup>26</sup> "Un treball és aplicat en sentit estricte quan s'encamina a la resolució d'un problema real que sorgeix en la pràctica lingüística" (M.T. CABRÉ y LL. PAYRATÓ, *op. cit.*, p. 19).

<sup>27</sup> *Id.*, p. 14.

<sup>28</sup> *Id.*, pp. 14-15.

<sup>29</sup> *Vid.* LL. PAYRATÓ, *op. cit.*, p. 15.

<sup>30</sup> M.T. CABRÉ y LL. PAYRATÓ, *op. cit.*, p. 15.

<sup>31</sup> LL. PAYRATÓ, *op. cit.*, p. 27.

<sup>32</sup> *Id.*, p. 33.

<sup>33</sup> En adelante, LC.

la traducción automática, identificada a veces con la propia lingüística aplicada. A partir de definiciones amplias como: "la lingüística computacional es el estudio de los sistemas de computación utilizados para la comprensión y la generación de lenguas naturales"<sup>34</sup> o "bajo la denominación de *Lingüística computacional* es posible agrupar un conjunto relativamente heterogéneo de teorías, métodos, herramientas, aplicaciones y productos que tienen en común la consideración de la lengua como un objeto susceptible de ser tratado mediante procedimientos informáticos"<sup>35</sup>, se suele situar la lingüística computacional, bien como rama de la lingüística<sup>36</sup>, bien como rama de la informática (inteligencia artificial)<sup>37</sup>, o a medio camino entre las dos<sup>38</sup>. Aunque en la actualidad lo más frecuente es su inclusión como área de la lingüística aplicada: *La LM [lingüística mecanizada] hay que situarla en una de las varias direccio-*

<sup>34</sup> R. GRISHMAN, *op. cit.*, p. 15. En el original: «Computational linguistics is the study of computer systems for understanding and generating natural language» (R. GRISHMAN, *op. cit.*, p. 4).

<sup>35</sup> MOURE, T. y LISTERRI, J.: "Lenguaje y nuevas tecnologías: el campo de la lingüística computacional", en M. FERNÁNDEZ PÉREZ (coord.), *Avances en Lingüística aplicada*, Servicio de Publicaciones e Intercambio Científico, Universidade de Santiago de Compostela, 1996, p. 147.

<sup>36</sup> Es el caso de CRYSTAL, D.: *A Dictionary of Linguistics and Phonetics*, Basil Blackwell, Oxford, 1980, 3ª ed. corregida y aumentada, 1991. La definición que ofrece de LC es la siguiente: «Computational linguistics: A branch of LINGUISTICS in which computational techniques and concepts are applied to the elucidation of linguistic and PHONETIC problems. Several research areas have developed, including SPEECH synthesis, speech recognition, automatic translation, the making of concordances, the testing of GRAMMARS, and the many areas where statistical count and analyses are required (e.g. in literary textual studies)» (*Id.*, pp. 70-71).

<sup>37</sup> Por ejemplo, HALVORSEN, P.: "Computer applications of linguistic theory" en FREDERICK J. NEWMAYER (ed.): *Linguistics: The Cambridge Survey. Vol. II: Linguistic Theory: Extensions and Implications*, Cambridge University Press, Cambridge, 1988, 198-219, quien señala que: "Computational linguistics is best viewed as a branch of artificial intelligence (AI). As all fields within AI, it is concerned with the investigation and modeling of a cognitive capacity. In the case of computational linguistics it is the language capacity that is in focus. (...) The goal is rather to identify and characterize the classes of processes and the types of knowledge which are implied by the ability to communicate and assimilate information using natural language regardless of their psychological status. One of the contributions of computational linguistics is a set of techniques which make it possible for linguistic knowledge to guide and constrain the linguistic processing performed in a natural language system" (*Id.*, pp. 202-203).

<sup>38</sup> "... a separate subdiscipline of computer science and linguistics, namely computational linguistics. Computational linguistics tries to mediate between competence theory and the particular kind of linguistic performance attributable to machines by turning linguistic theory into algorithms which allow the simulation of linguistic behavior while obeying the linguistic constraints and generalizations embodied in linguistic theory and competence grammars" (*Id.*, p. 201).

nes de la Lingüística Aplicada y en un contexto ideológico que podemos calificar de neopositivista tal como se ha desarrollado en los últimos tiempos<sup>39</sup>.

M. Fernández, siguiendo a T. Slama-Cazacu<sup>40</sup>, opta por la elección de un criterio interno y técnico, el de la finalidad de las investigaciones<sup>41</sup>, "modo de investigación"<sup>42</sup> o "aproximación"<sup>43</sup>, para distinguir entre *lingüística teórica*, que persigue lograr el conocimiento, la teoría, y la *lingüística aplicada*, que busca alcanzar soluciones a los problemas materiales que se plantean en las lenguas<sup>44</sup>. Pues según este mismo criterio, la citada autora incluye la *lingüística computacional* dentro del ámbito de la *lingüística aplicada*: *La Lingüística computacional se encuadra en el grupo de las disciplinas aplicadas porque proyecta determinados conocimientos sobre el funcionamiento de las lenguas a la resolución de problemas concretos*<sup>45</sup>, junto a áreas como *didáctica de lenguas*, *traductología*, *lingüística clínica*, *planificación lingüística*, etc.<sup>46</sup>, de acuerdo con los problemas concretos a los que se enfrentan.

Cada una de estas subáreas o subdisciplinas tiene su objeto, metodología y finalidades particulares, con el punto en común de su orientación hacia la búsqueda y hallazgo de vías de solución a los problemas materiales<sup>47</sup>. En concreto, la lingüística computacional se puede caracterizar según Moure y Llisterri<sup>48</sup> por: i) un *objeto de estudio* consistente en "elaborar teorías y procedimientos para lograr el procesamiento o tratamiento automático de las lenguas"; una *metodología* "híbrida, a medio camino entre informática y lingüística"; y el *objetivo* de "obtener productos tecnológicos relacionados con las industrias de la lengua". Esto quiere decir que la lingüística

---

<sup>39</sup> GARRIDO MORAGA, A.M.: "La lingüística y los ordenadores. Consideraciones sobre lingüística mecanizada", *Analecta Malacitana*, VII, 2, Universidad de Málaga, 1984:213. O también "La Lingüística Computacional es una disciplina aplicada. Entre sus usos principales figuran, entre otros, la traducción automática, los interfaces hombre-máquina, la recuperación y extracción de información y los correctores sintácticos y estilísticos" (MORENO SANDOVAL, A.: *Lingüística computacional. Introducción a los modelos simbólicos, estadísticos y biológicos*, Síntesis, Madrid, 1998:30).

<sup>40</sup> *Op. cit.*

<sup>41</sup> *Vid.* FERNÁNDEZ PÉREZ, M.: "El campo de la lingüística aplicada. Introducción", en FERNÁNDEZ PÉREZ, M. (coord.), *op. cit.*, pp. 12-13.

<sup>42</sup> ELIASSON, S.: "The Interrelations between theoretical and Applied Linguistics", en TOMIC, O.M. y SHUY, R.W., *The relation of theoretical and applied linguistics*, Plenum Press, New York/London, 1987, 21-49.

<sup>43</sup> BUGARSKI, R.: "Applied Linguistics as Linguistics Applied", en TOMIC, O.M. y SHUY, R.W., *op. cit.*, 3-19.

<sup>44</sup> *Vid.* M. FERNÁNDEZ, *op. cit.*, p. 21.

<sup>45</sup> T. MOURE y J. LLISTERRI, *op. cit.*, p. 210.

<sup>46</sup> *Cf.* M. FERNÁNDEZ, *op. cit.*, pp. 22-27.

<sup>47</sup> *Id.*, p. 21.

<sup>48</sup> *Op. cit.*, p. 209.

computacional combina "objetivos científicos" o teóricos con las aplicaciones u objetivos prácticos<sup>49</sup>. Por una parte, integra conocimientos teóricos de otras disciplinas, pero también elabora sus propios formalismos y conceptos teóricos según las necesidades concretas a las que tenga que atender<sup>50</sup> que, en una relación de reciprocidad, revierten en beneficio de la propia lingüística, ya que "la Lingüística computacional está de hecho pesando en los desarrollos teóricos de la Gramática, de la Lexicología e incluso de la Pragmática, en donde las formalizaciones vuelven a estar de moda"<sup>51</sup>. En cuanto a las aplicaciones, Grishman<sup>52</sup> enumera la traducción automática, la recuperación de información y las interfaces hombre-máquina. Esto sin mencionar todo lo relacionado con las industrias de la lengua, etc., aspectos de los que daremos cuenta en el apartado correspondiente. Por otra parte, no hay que olvidar que ambas facetas integran conocimientos teóricos, técnicas y métodos de otras disciplinas<sup>53</sup> como la inteligencia artificial, la informática, la lógica, las matemáticas, la lingüística, la psicolingüística, la psicología, etc., cumpliendo así con el requisito de la interdisciplinariedad.

Así pues, creemos que no cabe duda sobre la inclusión de la lingüística computacional dentro de la lingüística aplicada, puesto que responde a las características que se han dado como definitorias de aquélla: i) por una parte, «la Lingüística computacional se encuadra en el grupo de las disciplinas aplicadas porque proyecta determinados conocimientos sobre el funcionamiento de las lenguas a la resolución de problemas concretos»<sup>54</sup>, es decir, es capaz de desarrollar teorías y aplicaciones; ii) y, por otra parte, lo hace desde una perspectiva interdisciplinar, acudiendo a conceptos, técnicas, etc. de otras disciplinas. En palabras de Moure y Llisterra: *Con ello se está consolidando un campo de trabajo caracterizado principalmente por la interdisciplinariedad y por ofrecer la posibilidad de convertir las teorías en realidades, materializadas en productos que, en última instancia, tienen como objeto ayudar a las personas en aquellas tareas en las que el lenguaje juega un papel preponderante*<sup>55</sup>.

Igual que la lingüística aplicada, adolece de los mismos problemas que se plantean respecto a su estatus, ya que: *Frente a una gruesa tradición gramatical, la mayoría de las disciplinas aplicadas en Lingüística no se han reconocido como tales hasta época muy reciente. En ocasiones porque su cometido no quiso verse como un objeto de estudio relevante (...) y, en*

<sup>49</sup> Cf. R. GRISHMAN, *op. cit.*, pp. 15-16.

<sup>50</sup> Cf. T. MOURE y J. LLISTERRI, *op. cit.*, p. 152.

<sup>51</sup> M. FERNÁNDEZ, *op. cit.*, p. 28.

<sup>52</sup> Cf. R. GRISHMAN, *op. cit.*, pp. 15-16.

<sup>53</sup> Como dicen T. MOURE y J. LLISTERRI, *op. cit.*, p. 152: «El conocimiento sobre el lenguaje se integra pues, junto con elementos procedentes de otras disciplinas, tanto en la teoría de la lingüística computacional como en la concepción de sistemas y herramientas y en la creación de recursos lingüísticos».

<sup>54</sup> *Id.*, p. 210.

<sup>55</sup> *Id.*, p. 153.

ocasiones, si la práctica correspondiente era una necesidad dictada por la realidad, (...) porque se confiaban a especialistas diferentes del propio lingüista. La Lingüística quedaba a salvo de cualquier problema práctico. En el caso que nos ocupa la disociación viene dictada porque su juventud es un imperativo material: hace pocos años que existen soportes técnicos suficientes como para sustentar una lingüística computacional de perfil específico<sup>56</sup>. Sin embargo, su juventud y su consideración como disciplina autónoma no implican que esté desvinculada de la Lingüística, "entendida ésta como etiqueta general, que abarca proyecciones teóricas y aplicadas. Al fin, sólo se trata de subrayar que la Lingüística computacional lleva a cabo una investigación científica y no, simplemente, aplica útiles informáticos a datos procedentes del lenguaje"<sup>57</sup>.

### 3. DELIMITACIÓN DEL CAMPO DE LA LINGÜÍSTICA COMPUTACIONAL

Ya se ha visto en el apartado anterior que no existe ni mucho menos unanimidad a la hora de definir la LC, indefinición motivada entre otros factores por la juventud de la disciplina, por la cantidad de ciencias con las que intersecciona y por el vertiginoso avance de la base tecnológica que la sustenta. También puede que sea un simple capricho, un hecho que obedece al uso idiosincrásico que hacemos de las palabras<sup>58</sup>. Sea como fuere, el caso es que con frecuencia, junto a la denominación de *lingüística computacional*, nos encontramos con términos como *lingüística informática*, *procesamiento del lenguaje natural (PLN)*, *inteligencia artificial (IA)*, *industrias de la lengua*, *ingeniería lingüística*, etc. entre otros. Por ello, consideramos necesario delimitar, aunque sea de forma un tanto somera y sucinta, los límites entre los mismos antes de tratar las aplicaciones de la LC, ya que la clasificación que hagamos de éstas dependerá en cierta medida de las diferenciaciones previas en el marco de la LC. Así, siguiendo a A. Moreno Sandoval<sup>59</sup>, nos encontramos con que, en torno a los conceptos fundamentales de lenguas naturales y ordenadores, se agrupan disciplinas como las que acabamos de enumerar. Todas ellas —y otras no mencionadas— compartirían el mismo objeto de estudio, el lenguaje, aunque diferirían en la perspectiva que adoptan a la hora de abordarlo.

---

<sup>56</sup> *Id.*, p. 212.

<sup>57</sup> *Id.*, p. 210.

<sup>58</sup> Como decía Lewis Carol: «Cuando yo uso una palabra», dijo Humpty-Dumpty en un tono completamente desdeñoso, «esa palabra significa lo que yo quiero que signifique, —ni más, ni menos». Cita aportada por la alumna Simone Ortner en un trabajo de clase.

<sup>59</sup> Cf. A. MORENO SANDOVAL, *op. cit.*, pp. 13-16.

### 3.1.- Concepción restringida de Lingüística computacional. Su relación con la Inteligencia Artificial (IA)

Si, como propone A. Moreno Sandoval<sup>60</sup>, comparamos la definición que da Grishman<sup>61</sup> de *lingüística computacional* y la que ofrece Allen<sup>62</sup> de *procesamiento del lenguaje natural*, llegaremos a la conclusión de que ambas son lo mismo, es decir, tratan del "desarrollo de programas de ordenador que simulen la capacidad lingüística humana"<sup>63</sup>. Y esto, lo pueden hacer desde una vertiente teórica o desde una práctica<sup>64</sup>, como ya se ha señalado.

Desde el momento que entra en juego la simulación, nos situamos en el terreno de la *inteligencia artificial*, subdisciplina de la Informática encargada de «codificar en un programa informático facultades cognitivas»<sup>65</sup>. En este sentido, el tratamiento o procesamiento del lenguaje natural aparece como uno de los cinco grandes bloques en que la IA estructura el comportamiento humano inteligente y que constituyen sus líneas de investigación básicas. Así, por ejemplo, lo considera Zaccagnini<sup>66</sup>. Pero el PLN no es un bloque cualquiera dentro de la IA, ya que el dominio del lenguaje, como capacidad cognitiva fundamental, siempre ha sido una de sus metas principales, de cara a obtener la comunicación hombre-máquina, o persona-ordenador. Es decir, el objetivo de la IA respecto al procesamiento del lenguaje es más amplio o global, ya que lo entiende como una parte integrante de un sistema inteligente.

Igual que ocurría con la LC, la IA tiene una doble vertiente *teórica* y *práctica*, relacionada con las dos perspectivas de abordar el campo presentes ya desde sus inicios en los años 50, según la razón por la que se pretenda que las máquinas realicen tareas inteligentes: (a) la primera, la *teórica*, se correspondería con lo que A.M. Ramsay llama "el punto de vista de la *ciencia cognitiva*"<sup>66</sup>, orientada hacia la investigación psicológica, en una perspectiva de la IA como "una herramienta para investigar sobre la naturaleza de la inteligencia"<sup>67</sup> de cara a averiguar cómo ejecuta el hombre esos procesos,

<sup>60</sup> *Ib.*

<sup>61</sup> Véase nota 33.

<sup>62</sup> «The goal of this research is to create computational models of language in enough detail that you could write computer programs to perform various tasks involving natural language» (ALLEN, J.: *Natural Language Understanding*, The Benjamin/Cummings Publishing Company, Redwood City, CA, 2ª ed., 1995:1).

<sup>63</sup> A. MORENO SANDOVAL, *op. cit.*, p. 14.

<sup>64</sup> Cf. R. GRISHMAN, *op. cit.*, pp. 15-16.

<sup>65</sup> A. MORENO SANDOVAL, *op. cit.*, p. 14.

<sup>66</sup> ZACCAGNINI, J.L.: «Introducción al campo de la Inteligencia Artificial», en P. ADARRAGA y J.L. ZACCAGNINI (eds.), *Psicología e Inteligencia Artificial*, Trotta, Madrid, 1994:25 y ss.

<sup>67</sup> RAMSAY, A.M.: "Artificial Intelligence" en K. MALMKAER (ed.), *The Linguistics Encyclopedia*, Routledge, London, 1991:28-29.

<sup>68</sup> FERNÁNDEZ, G. y SAEZ VACAS, F.: *Fundamentos de Informática. Lógica, Automatas, Algoritmos y Lenguajes*, Anaya Multimedia, Madrid, 1995:220.



### 3.2.- Concepción amplia de la Lingüística Computacional. Su relación con la Lingüística Informática

Pero también podemos entender la LC como parte de la Lingüística Informática, entendiendo por ésta "la disciplina que abarca todo uso de ordenadores con relación al lenguaje y las lenguas. Aquí, se incluirían no sólo los sistemas que simulan el lenguaje humano, sino todo tipo de programas y herramientas informáticas que ayudan en el estudio de las lenguas y de la Lingüística"<sup>74</sup>. Estamos ante los trabajos pioneros en LC, en los que el ordenador aparece como una herramienta eficaz para abordar tareas mecánicas relacionadas con contar, clasificar, buscar y archivar información, por la rapidez y exactitud que ello supone. También se incluiría en esta concepción amplia todo lo relativo a la elaboración de índices y concordancias de textos. En palabras de Gazdar y Mellish: *As those who count sheep know well, counting is a very boring task. Even the very earliest computers counted fast and accurately, and they did not get bored. (...) Some of the earliest work that came to be known as CL did exactly this kind of counting. (...) Other work then considered to be CL involved the use of computers to derive indexes and concordances from computer-readable texts*<sup>75</sup>.

Es lo que Moure y Llisterri denominan 'herramientas informáticas para la investigación lingüística', herramientas que "constituyen elementos auxiliares, aunque actualmente del todo imprescindibles, para llevar a cabo una investigación, sea para ordenar, clasificar y verificar la coherencia de los datos, sea para procesarlos y extraer información nueva que permita establecer generalizaciones o determinar tendencias estadísticas"<sup>76</sup>. También Bott da cuenta de este sentido amplio de la LC: *El problema que hemos examinado constituye un ejemplo de la aplicación de las computadoras a trabajos de interés puramente lingüístico o literario [concordancias e índices de palabras]. El término 'lingüística computacional' recubre tanto esta clase de trabajos como la aplicación de la lingüística a problemas pertenecientes básicamente a la computación; es decir, problemas cuyo objetivo consiste en elaborar un programa de computadora que ejecute una tarea de manera más eficiente, más rápida y barata de lo que la ejecutarían los seres humanos*<sup>77</sup>.

No obstante, hay que señalar que hoy en día se rechaza la denominación de LC para esta segunda orientación tan amplia vinculada a la lingüística

<sup>74</sup> A. MORENO SANDOVAL, *op. cit.*, pp. 14-15.

<sup>75</sup> GAZDAR, G. y MELLISH, C.: «Computational Linguistics», en J. LYONS *et alii* (eds.), *New horizons in linguistics 2. An introduction to contemporary linguistics researchs*, Penguin Books, London, 1987: 225-226.

<sup>76</sup> *Op. cit.*, p. 148.

<sup>77</sup> BOTT, M.F.: "Lingüística computacional" en J. LYONS: *Nuevos horizontes de la lingüística*, Alianza Editorial, Madrid, 1975. Cito por la edición española de 1975: 234-235.

informática, en beneficio de la LC en sentido estricto o LC propiamente dicha, identificada con el PLN y, en consecuencia, dependiente de la IA: *Computational linguistics: 1. (formerly, and still occasionally) A very broad label covering virtually any activity involving computers and natural language, such as machine translation of natural-language texts, computer searching of texts or the preparation of concordances for literary works by computer. Now usually called 'literary and linguistic computing'. 2. (more usually today) A synonym for natural-language processing*<sup>78</sup>.

Y es que, como aseveran de forma drástica Gazdar y Mellish, aunque en el momento fundacional de la LC las actividades que se englobaban bajo la denominación de lingüística informática eran el único tipo de LC posible, "nowadays such work no longer counts as computational linguistics, or even as an academic activity, since humble word-processing programs often come equipped with sophisticated indexing utilities"<sup>79</sup>.

### 3.3.- Ingeniería lingüística e industrias de la lengua

Por último, nos queda por considerar el ámbito de las industrias de la lengua y de la ingeniería lingüística, surgido a raíz de su mención en el contexto europeo de los programas marco de investigación. Por *ingeniería lingüística* entiende Moreno Sandoval «toda aquella aplicación potencialmente comercial que implique el uso de nuevas tecnologías y lenguas»<sup>80</sup>. O más específicamente: "la aplicación de los conocimientos sobre la lengua al desarrollo de sistemas informáticos que puedan reconocer, comprender, interpretar y generar lenguaje humano en todas sus formas"<sup>81</sup>. Las *industrias de la lengua* constituyen un ámbito más amplio que abarca "una serie de actividades comerciales en las que el tratamiento del lenguaje por personas o por máquinas o por una combinación de unas y otras, forma una parte fundamental del producto o servicio"<sup>82</sup>. Su existencia exige el desarrollo previo de la ingeniería lingüística, que es la encargada de proporcionar las *técnicas* y los *recursos* necesarios para crear los *productos* y *herramientas* que utilizan el lenguaje de alguna manera<sup>83</sup>. En definitiva, lo que caracteriza tanto a la ingeniería lingüística como a las industrias de la

<sup>78</sup> TRASK, R.L.: *A Dictionary of Grammatical Terms in Linguistics*, Routledge, London-New York, 1993:53.

<sup>79</sup> *Op. cit.*, p. 226.

<sup>80</sup> A. MORENO SANDOVAL, *op. cit.*, p. 16.

<sup>81</sup> *Ingeniería lingüística. Cómo aprovechar la fuerza del lenguaje*, Observatorio Español de Industrias de la Lengua-DG XIII/E, Luxemburgo, p. 5. Publicación electrónica en: <http://www2.echo.lu/langeng/es/broch/harness.html>.

<sup>82</sup> *Lenguaje y tecnología. De la torre de Babel a la aldea global*, Oficina de Publicaciones Oficiales de las Comunidades Europeas, Luxemburgo, 1997:12.

<sup>83</sup> LISTERRI, J. y GARRIDO ALMIÑANA, J.M.: «La ingeniería lingüística en España», en *El español en el mundo. Anuario del Instituto Cervantes*, Centro Virtual Cervantes, Instituto Cervantes (España)-Arco/Libros, Madrid, 1998:299. Publicación electrónica en <http://cvc.cervantes.es/obref/anuario/parte2/parte2.htm>.

lengua es su marcada orientación comercial. Por este motivo, sólo una parte de la LC tiene cabida en ellas, aquélla de clara orientación práctica y dirigida hacia la comercialización de un producto.

### 3.4.- Definición final de LC

Para terminar este apartado, recogemos la definición que propone Moreno Sandoval de LC, ya que da cuenta de todos los aspectos aquí mencionados. Según dicho autor, la LC se podría definir como sigue: *En resumen, y por oposición a las otras disciplinas, la LC trata de la construcción de sistemas informáticos que procesen estructura lingüística y cuyo objetivo sea la simulación parcial de la capacidad lingüística de los hablantes de una lengua, independientemente de su carácter comercial o de investigación básica*<sup>84</sup>.

## 4. LAS APLICACIONES DE LA LC

En este último apartado, la distinción establecida por Grishman<sup>85</sup> entre objetivos 'científicos' y aplicaciones nos va a servir como punto de partida para comparar una serie de modelos en lo que a las aplicaciones de LC se refiere. Por tanto, nos vamos a situar ante la vertiente práctica o aplicada de la LC. Los autores que hemos elegido para efectuar dicha comparación son los siguientes: Butler<sup>86</sup>, Grishman<sup>87</sup>, Allen<sup>88</sup>, Moreno Sandoval<sup>89</sup>, Moure y Llisterri<sup>90</sup> y Gómez Guinovart<sup>91</sup>. Esta selección incluye tanto manuales clásicos de LC, como pueden ser los de Grishman o el de Gazdar y Mellish, como libros y artículos de autores destacados dentro del panorama actual español, con la finalidad de cubrir todas las posibles líneas de investigación dentro de la LC que hemos esbozado en el apartado anterior.

### 4.1.- Grishman

Para Grishman la LC, entendida como "el estudio de los sistemas de computación utilizados para la comprensión y la generación de lenguas naturales"<sup>92</sup>, tiene por objetivo general dar cuenta del poder del lenguaje: *A través de la comprensión de los procesos lingüísticos en términos de procesamiento podemos lograr que los sistemas informáticos sean capaces de generar e interpretar las lenguas naturales. Esto haría posible que los ordenadores realizasen tareas lingüísticas (como la traducción), procesa-*

<sup>84</sup> A. MORENO SANDOVAL, *op. cit.*, p. 16.

<sup>85</sup> R. GRISHMAN, *op. cit.*, pp. 15-16.

<sup>86</sup> BUTLER, C.: *Computers in Linguistics*, Basil Blackwell, Oxford, 1985.

<sup>87</sup> *Op. cit.*

<sup>88</sup> *Op. cit.*

<sup>89</sup> *Op. cit.*

<sup>90</sup> *Op. cit.*

<sup>91</sup> GÓMEZ GUINOVART, J. et alii (eds.): *Panorama de la investigación en Lingüística Informática*, ResLA, volumen monográfico, Logroño, 1999.

<sup>92</sup> *Op. cit.*, p. 15.

sen datos en forma de texto (libros, revistas, periódicos) y facilitarán al público en general el acceso a los datos almacenados en los ordenadores<sup>92</sup>. Es decir, la principal motivación de la LC es de tipo práctico, y así, "aunque los objetivos de la investigación sobre lingüística computacional sean muy variados, una primera motivación ha sido el desarrollo de *sistemas prácticos específicos* que tengan que ver con las lenguas naturales"<sup>94</sup>. Grishman considera que son tres las aplicaciones principales de la LC<sup>95</sup>: la *traducción automática*, la *recuperación de la información* y las *interfaces hombre-máquina*, también llamadas por Grishman 'aplicaciones orientadas a la ingeniería'<sup>96</sup>, estableciendo así la conexión con la doble vertiente distinguida en la IA<sup>97</sup>.

Sin embargo, esto no quiere decir que la LC se desentienda de la faceta teórica, ya que "la mayor parte de los que trabajan en el campo de la lingüística computacional también tienen *objetivos 'científicos'* de investigación independientes de cualquier aplicación"<sup>98</sup>. Entre esos objetivos, Grishman destaca el de probar las gramáticas propuestas por la lingüística teórica, así como aquéllos relacionados con áreas de investigación nuevas como la ciencia cognitiva y la representación del conocimiento, interesados en la investigación de los modelos psicológicos subyacentes al procesamiento lingüístico. No obstante, el propio Grishman reconoce que "los objetivos prácticos y los puramente científicos van generalmente juntos"<sup>99</sup>.

#### 4.2.- Allen

Para Allen, la disciplina que nos ocupa se centra en describir *research about how language comprehension and production work. The goal of this research is to create computational models of language in enough detail that you could write computer programs to perform various tasks involving natural language. (...) Computational models are useful both for scientific purposes —for exploring the nature of linguistic communication— and for practical purposes —for enabling effective human-machine communication*<sup>100</sup>.

Así pues, Allen menciona dos motivaciones a la hora de desarrollar modelos computacionales del lenguaje, fiel reflejo de la doble concepción de la IA, ya presente en Grishman:

- a) La primera de ella —'scientific motivation'— estaría vinculada a los objetivos 'científicos' de Grishman y, como en este autor, conduce hacia los dominios interdisciplinares de la ciencia cognitiva, ya que se trata de un objetivo común a varias disciplinas: *The scientific motivation is to*

<sup>92</sup> *Id.*, p. 11.

<sup>94</sup> *Id.*, p. 15.

<sup>95</sup> *Id.*, pp. 15-16.

<sup>96</sup> *Id.*, p. 16.

<sup>97</sup> Ver epígrafe 3.1.

<sup>98</sup> *Id.*, p. 16.

<sup>99</sup> *Id.*, p. 17.

<sup>100</sup> *Op. cit.*, p. 1.

obtain a better understanding of how language works. It recognizes that any one of the other traditional disciplines does not have the tools to completely address the problem of how language comprehension and production work. Even if you combine all the approaches, a comprehensive theory would be too complex to be studied using traditional methods. But we may be able to realize such complex theories as computer programs and then test them by observing how well they perform. By seeing where they fail, we can incrementally improve them. Computational model may provide very specific predictions about human behavior that can then be explored by the psycholinguist. By continuing in this process, we may eventually acquire a deep understanding of how human language processing occurs. To realize such a dream will take the combined efforts of linguists, psycholinguists, philosophers, and computer scientists. This common goal has motivated a new area of interdisciplinary research often called cognitive science<sup>101</sup>.

- b) La segunda motivación —'practical or technological motivation'— enlaza directamente con los objetivos prácticos de Grishman, destacando la importancia de las interfaces: *The practical, or technological, motivation is that natural language processing capabilities would revolutionize the way computers are used. Since most of human language is recorded in linguistic form, computers that could understand natural language could access all this information. In addition, natural language interfaces to computers would allow complex systems to be accessible to everyone. Such systems would be considerably more flexible and intelligent than is possible with current computer technology. For technological purposes it does not matter if the model used reflects the way humans process language. It only matters that it works*<sup>102</sup>.

Igual que Grishman, adopta un punto de vista intermedio entre ambos objetivos o motivaciones, ya que: *on the one hand, this reflects a belief that natural language is so complex that an ad hoc approach without a well-specified underlying theory will not be successful. Thus the technological goal cannot be realized without using sophisticated underlying theories on the level of those being developed by linguists, psycholinguists, and philosophers. On the other hand, the present state of knowledge about natural language processing is so preliminary that attempting to build a cognitively correct model is not feasible. Rather, we are still attempting to construct any model that appears to work*<sup>103</sup>.

Más adelante, no obstante, hace una distinción más sutil que las de Grishman cuando aborda las aplicaciones de lo que él denomina 'natural language understanding'<sup>104</sup>. Éstas se dividirían en dos clases principales:

<sup>101</sup> *Id.*, p. 2.

<sup>102</sup> *Id.*, pp. 2-3.

<sup>103</sup> *Id.*, p. 3.

<sup>104</sup> *Ib.*

'text-based applications' y 'dialogue-based applications'<sup>105</sup>. Las primeras implican "the processing of written text, such as books, newspapers, reports, manuals, e-mail messages, and so on. These are all reading-based tasks"<sup>106</sup>. Más o menos estamos ante la recuperación de información y la traducción automática de Grishman. Las segundas se centran en la comunicación hombre-máquina: "Most naturally this involves spoken language, but it also includes interaction using keyboards"<sup>107</sup>.

Ésta sería una síntesis del modelo de Allen<sup>108</sup>:

COMPUTATIONAL LINGUISTICS	
'TECHNOLOGICAL GOALS'	'SCIENTIFIC GOALS'
APLICACIONES	
'TEXT-BASED APPLICATIONS'	'DIALOGUE-BASED APPLICATIONS'
<ul style="list-style-type: none"> <li>* finding appropriate documents on certain topics from a database of texts</li> <li>* extracting information from messages or articles on certain topics</li> <li>* translating documents from one language to another</li> <li>* summarizing texts for certain purposes</li> </ul>	<ul style="list-style-type: none"> <li>* question-answering systems, where natural language is used to query a database</li> <li>* automated customer service over the telephone</li> <li>* tutoring systems, where the machine interacts with a student</li> <li>* spoken language control of a machine</li> <li>* general cooperative problem-solving systems</li> </ul>

<sup>105</sup> *Id.*, p. 4.

<sup>106</sup> Hay que aclarar que todas estas tareas pueden ser llevadas a cabo sin que por ello estén implicadas técnicas relacionadas con el procesamiento del lenguaje natural. Sólo estaremos ante aplicaciones de este campo, es decir, ante sistemas de comprensión del LN, cuando se constate la siguiente característica en el sistema de comprensión: «it must compute some representation of the information that can be used for later inference» (*Ib.*). Es decir, a Allen sólo le interesan estas aplicaciones en la medida en que nos movamos en el ámbito del PLN, por oposición al de la lingüística informática, donde no existe ningún tipo de comprensión.

<sup>107</sup> *Id.*, p. 5. Aquí se puede decir lo mismo que en el apartado anterior. Un sistema de reconocimiento del habla no implica necesariamente que se produzca la comprensión. Se limita a reconocer palabras, no a comprenderlas. Como dice Allen: «To be an understanding system, the speech recognizer would need to feed its input to a natural language understanding system, producing what is often called a spoken language understanding system» (*Id.*, p. 6).

<sup>108</sup> *Cf.*, pp. 1-5.

#### 4.3.- A. Moreno Sandoval

Para este autor, la disciplina que nos ocupa "trata básicamente de dos cosas: *lenguas naturales y ordenadores*"<sup>109</sup>. Ambos objetivos son comparados por otras disciplinas más o menos afines aunque desde diferentes perspectivas. Por esta razón, lo primero que hace es delimitar el campo de estudio<sup>110</sup>, distinguiendo entre LC-PLN, IA, Lingüística Informática e industrias de la lengua-ingeniería lingüística. Por lo tanto, Moreno Sandoval no hace una distinción entre objetivos, como autores anteriores (Grishman, Allen), sino que delimita el campo. Sin embargo, en lo que a las aplicaciones de la LC se refiere, Moreno Sandoval clasifica las aplicaciones prácticas en cuatro bloques<sup>111</sup>:

	APLICACIONES
I) Sistemas que tratan de emular la capacidad humana de procesar lenguas naturales	-Traducción Automática -Recuperación y extracción de información -Interfaces hombre-máquina
II) Sistemas que ayudan en las tareas lingüísticas	-Herramientas de análisis textual (frecuencias de aparición, concordancias, estadísticas) -Herramientas para manejo de corpus (etiquetadores categoriales, analizadores sintácticos) -Bases de datos lexicográficas y bases de datos terminológicas
III) Programas de ayuda a la escritura y composición textual	-Correctores ortográficos -Correctores sintácticos y de estilo
IV) Enseñanza asistida por ordenador	-Programas educativos para la enseñanza de lenguas extranjeras -Recursos multimedia

Como podemos observar, el bloque I de aplicaciones se corresponde con los objetivos prácticos de Grishman y Allen, mientras que el resto de bloques se ubicarían dentro de la Lingüística Informática (bloque II) y de las industrias de la lengua (III y IV), tal y como veremos en las propuestas de los siguientes autores.

#### 4.4.- Butler

Se trata de un fiel representante de la tradición de la lingüística informática, por lo tanto, del sentido amplio de LC o del uso de los ordenadores como mera herramienta. Sin ningún tipo de definición ni distinción entre posibles objetivos, Butler enumera toda una serie de aplicaciones de la lingüística informática<sup>112</sup>:

<sup>109</sup> *Op. cit.*, p. 13.

<sup>110</sup> Véase apartado 3.

<sup>111</sup> *Id.*, pp. 27-29.

<sup>112</sup> *Op. cit.*, pp. 12-35. Tomamos el cuadro de A. MORENO SANDOVAL, *op. cit.*, p. 15.

LINGÜÍSTICA INFORMÁTICA	
Áreas de conocimiento	Aplicaciones
Investigación literaria y estilística	Análisis cuantitativos de textos literarios (listas de palabras, índices, concordancias); análisis métricos, sintácticos y semánticos sobre textos etiquetados manualmente; asignación de autoría y localización cronológica
Lexicografía	Corpus electrónicos y bases de datos lexicográficas; lematización automática, ordenación alfabética; edición e impresión
Edición de textos	Edición crítica de textos literarios: colación automática de variantes
Enseñanza y aprendizaje de lenguas	Estudios sobre corpus (p.e., frecuencias) para establecer prioridades en el diseño de métodos pedagógicos; presentación de los materiales pedagógicos en formato electrónico (aprendizaje asistido por ordenador)
Simulación del procesamiento del lenguaje humano	Sistemas de análisis y generación de lenguas naturales; sistemas basados en la escritura y en el habla; traducción automática; sistemas de síntesis y reconocimiento de habla
Lingüística descriptiva basada en corpus	Corpus en formato electrónico, anotados morfológica y sintácticamente
Análisis computacional de datos no textuales	Programas estadísticos aplicados a la sociolingüística, enseñanza de lenguas, dialectología

Vemos cómo esta visión amplia de la disciplina que ofrece Butler recoge, tanto las 'aplicaciones' de Grishman o Allen ('simulación del procesamiento del lenguaje humano'), como otras ligadas estrechamente a esta vertiente de la disciplina ('investigación literaria y estilística', 'edición de textos', 'análisis computacional de datos no textuales'), junto con lo que Moreno Sandoval y otros autores considerarían parte de las industrias de la lengua ('enseñanza y aprendizaje de lenguas', 'lexicografía').

#### 4.5.- Gómez Guinovart

Este autor, en la introducción del volumen monográfico de ResLA denomina al campo objeto de estudio *lingüística informática* o *computacional*, y lo define como aquél "encaminado a incorporar en los ordenadores la habilidad en el manejo del lenguaje natural humano y a facilitar el tratamiento informatizado de las lenguas y de su estudio"<sup>113</sup>. Igual que Butler, adopta una visión amplia de la disciplina, aunque señalando la existencia de tres líneas de investigación dentro de la misma, cada una con sus respectivas aplicaciones:

<sup>113</sup> *Op. cit.*, p. 7.

- a) "Aplicación de los ordenadores a la investigación lingüística, es decir, al estudio científico del lenguaje y de las lenguas"<sup>114</sup>. Es lo que se suele conocer como *lingüística informática* (sentido estricto) o *informática aplicada a la lingüística*, términos procedentes del inglés *linguistic computing*. Como dice el propio autor: *Estos términos pueden aplicarse en sentido amplio a todas las subdisciplinas de la lingüística que emplean herramientas informáticas, aunque en general su empleo se reserva para las áreas de investigación en donde estos instrumentos poseen una mayor incidencia, como la lingüística de corpus, la lingüística estadística, la estilometría, la lingüística histórica computacional, la informática aplicada a la sociolingüística y la lexicografía asistida por ordenador*<sup>115</sup>.

Recoge aquí más o menos las mismas aplicaciones que Butler, excepto aquéllas relacionadas con la 'simulación del procesamiento del lenguaje humano' y la 'enseñanza y aprendizaje de lenguas'.

- b) "*Lingüística computacional* por antonomasia, en la cual pueden aislarse al menos tres objetivos: la elaboración de modelos lingüísticos en términos formales e implementables computacionalmente, la aplicación de estos modelos a cualquier nivel de descripción lingüística, y la comprobación automatizada de la consistencia de una teoría lingüística o de sus predicciones"<sup>116</sup>. Aproximadamente lo que Grishman y Allen incluían como 'objetivos científicos'.
- c) "La elaboración más tecnológica de la lingüística informática (entendida ahora nuevamente en sentido amplio) se centra en el diseño y elaboración de sistemas informáticos capaces de trabajar con enunciados orales y escritos en lenguas naturales; y se concreta en el desarrollo de aplicaciones lingüísticas de la informática cuyos productos más populares pueden agruparse en seis categorías"<sup>117</sup>, dentro de lo que se puede llamar *procesamiento del lenguaje natural, industrias de la lengua, ingeniería lingüística o tecnologías de la lengua*. En resumen:

---

<sup>114</sup> *Ib.*

<sup>115</sup> *Ib.*

<sup>116</sup> *Ib.*

<sup>117</sup> *Id.*, pp. 7-8.

LINGÜÍSTICA INFORMÁTICA O COMPUTACIONAL (SENTIDO AMPLIO)	
Líneas de investigación	Aplicaciones
a) <i>Lingüística informática</i> (sentido estricto) o <i>informática aplicada a la lingüística</i>	Herramientas útiles en: -lingüística de corpus -lingüística estadística -estilometría -lingüística histórica computacional -informática aplicada a la sociolingüística -lexicografía asistida por ordenador
b) <i>Lingüística computacional</i> (PLN)	'Objetivos científicos' de Grishman y Allen
c) <i>Industrias de la lengua, ingeniería lingüística o tecnologías de la lengua</i>	-herramientas de ayuda a la escritura (diccionarios electrónicos, sistemas de verificación automática de la ortografía, la sintaxis y el estilo) -herramientas de ayuda a la traducción (programas de traducción automática, bases de datos terminológicos, memorias de traducción) -aplicaciones de las tecnologías del habla (programas de dictado, sistemas de conversión de texto a voz) -sistemas de gestión documental (sistemas de extracción de información, sistemas de recuperación de la información textual) -aplicaciones didácticas para la enseñanza de lenguas (métodos de aprendizaje de idiomas asistidos por ordenador, programas de creación de ejercicios) -sistemas de diálogo (interfaces de consulta en lenguaje natural a bases de datos, sistemas automáticos de diálogo por línea telefónica)

Lo que más llama la atención de esta clasificación es el apartado c), donde aparecen mezcladas aplicaciones propias de las industrias de la lengua con las aplicaciones clásicas (traducción automática, recuperación de información e interfaces persona-ordenador) que Grishman y Allen proponen para la LC-PLN. Tal vez la explicación la podamos encontrar en el último modelo que vamos a contemplar.

#### 4.6.- Teresa Moure y Joaquim Llisterri. Conclusión

Consideran estos autores que: *bajo la denominación de Lingüística computacional es posible agrupar un conjunto relativamente heterogéneo de teorías, métodos, herramientas, aplicaciones y productos que tienen en común la consideración de la lengua como un objeto susceptible de ser tratado mediante procedimientos informáticos*<sup>117</sup>, adoptando así una perspec-

<sup>117</sup> *Op. cit.*, p. 147.

tiva muy amplia. Para delimitarla con más precisión distinguen entre cuatro corrientes principales: *lingüística informática, procesamiento del lenguaje natural, tecnologías del habla e industrias de la lengua*, que podemos sintetizar en el siguiente cuadro-resumen<sup>119</sup>:

Corrientes	Definición	Aplicaciones
LINGÜÍSTICA INFORMÁTICA	= herramientas informáticas para la investigación lingüística o aplicación de la informática a distintos campos lingüísticos	-Lingüística sincrónica y diacrónica -Filología -Literatura -Estilística -Métrica -Lexicografía -Enseñanza de lenguas -Traducción -Etc.
PROCESAMIENTO DEL LENGUAJE NATURAL (PLN)	= reproducir con medios informáticos la conducta lingüística	-Traducción automática -Comprensión automática de textos -Generación automática de textos -Diccionarios automáticos -Resumen del contenido de un texto -Extraer información de sistemas -Correctores ortográficos, gramaticales y de estilo
TECNOLOGÍAS DEL HABLA	= tratamiento de la lengua hablada	-Conversión de textos escritos en su equivalente oral (síntesis) -Transformación de habla en texto escrito (reconocimiento) -Traducción automática de conversaciones -Identificación o verificación de hablantes en servicios telefónicos -Sistemas de diálogo oral entre personas y máquinas (interfaces) -Enseñanza de lenguas asistida por ordenador -Ayudas a personas discapacitadas
INDUSTRIAS DE LA LENGUA E INGENIERÍA LINGÜÍSTICA	= aplicaciones del PLN y del habla en el desarrollo de productos comerciales, destinados a usuarios finales	-Sistemas de conversión texto-habla -Dictado automático -Traducción automática o asistida -Correctores ortográficos y gramaticales -Ayudas para el tratamiento multilingüe de textos -Programas de enseñanza de lengua asistida por ordenador -Sistemas de recuperación de información a partir del análisis de textos

Se puede observar cómo Moure y Llisterri diferencian claramente las distintas corrientes —y, en consecuencia, las aplicaciones— que convergen en la LC, desde el simple uso instrumental del ordenador como un elemento auxiliar (*lingüística informática*), hasta aplicaciones de orientación comercial o productos de mercado (*industrias de la lengua*), pasando por el núcleo central de la disciplina (*procesamiento del lenguaje natural*), donde establecen la distinción entre sistemas que trabajan con textos escritos y aquéllos que lo hacen con textos orales (*tecnologías del habla*).

No obstante, estos mismos autores señalan la existencia de puntos de convergencia en los intereses de los distintos campos señalados<sup>120</sup>: i) integración entre el procesamiento del lenguaje natural y las tecnologías del habla en proyectos europeos, metodologías empleadas (modelado estadístico), sistemas de traducción oral automática, publicaciones, revistas, congresos, etc.; ii) fronteras difusas entre el procesamiento del lenguaje natural y del habla y las industrias de la lengua; iii) necesidad de los llamados *recursos lingüísticos*, es decir, “recopilaciones de textos escritos, de grabaciones de lengua hablada, diccionarios, terminologías especializadas o gramáticas, todos ellos en un formato digital que posibilite su acceso y tratamiento informáticos”<sup>121</sup>.

Así pues, las diferentes aplicaciones de cada uno de los apartados señalados vienen a confluír a la hora de la verdad. Por eso, Moure y Llisterri optan por insistir en la necesidad de cubrir la faceta teórica: *En definitiva, los problemas que estas aplicaciones encuentran para la construcción de sus productos son una fuente empírica de datos lingüísticos. Si la teoría no toma en cuenta esos datos, simplemente estará describiendo una ficción. Tanto para incrementar nuestro conocimiento como para concebir productos más perfeccionados, que sirvan a las necesidades del mundo moderno, parece imprescindible orientar la investigación en lingüística computacional hacia objetivos teóricos y fundamentales*<sup>122</sup>.

Por lo tanto, la visión que ofrecen Moure y Llisterri, además de ser la más clara desde el punto de vista de las divisiones internas que establecen en el marco de la LC, así como de las diferentes aplicaciones que se corresponden con cada una de ellas, no deja de insistir en la necesidad de trabajar también los aspectos teóricos que sustentan la disciplina, sobre todo en el contexto actual en el que las líneas de trabajo apuntadas tienden a difuminarse. Por todo ello creemos que puede servir como conclusión de este trabajo.

---

<sup>119</sup> Elaborado a partir de la información que aparece en las páginas 148-150.

<sup>120</sup> *Id.*, p. 150.

<sup>121</sup> *Id.*, p. 151.

<sup>122</sup> *Id.*, p. 214.