

## Entrevista

### Conversaciones con un lingüista de corpus: Profesor Dr. Giovanni Parodi

por Ariel Novodvorski\* y Ana Fritz Herrera\*\*

Giovanni Parodi es actualmente Director de los Programas de Postgrado en Lingüística en la Pontificia Universidad Católica de Valparaíso, Chile, y también es Editor en Jefe de la *Revista Signos. Estudios de Lingüística*. Al mismo tiempo, es Director para Chile de la Cátedra UNESCO en Lectura y Escritura. Sus principales áreas de interés científico son la lingüística de corpus, la psicolingüística del discurso y la lingüística del texto. En la actualidad dirige proyectos de investigación en el discurso escrito académico y profesional con especial atención a los géneros multimodales especializados. El foco central de sus investigaciones está en profundizar en el conocimiento de los procesos textuales y psicolingüísticos implicados en la comprensión y producción de textos escritos en diversos géneros discursivos y en la construcción de conocimientos disciplinares y llevar estos hallazgos científicos a contextos educativos de enseñanza-aprendizaje en diversos niveles y dominios especializados con apoyo de tecnologías informáticas. Proyectos tales como El Grial y LECTES son programas asistidos por computadores tanto para la investigación básica como para el desarrollo de estrategias de lectura y escritura de alumnos de diversos niveles educacionales. Sus publicaciones incluyen más de 50 artículos en revistas en español e inglés, 19 libros (como autor único, coautor o editor) y diversos capítulos de libros en español e inglés. Sus últimos libros son: *Comprensión de Textos Escritos. La Teoría de la Comunicabilidad* (2014), *Academic and professional discourse genres in Spanish, Lingüística de Corpus: de la teoría a la empiria, Saber Leer, Alfabetización académica y profesional en el siglo XXI: leer y escribir en las disciplinas*. Por sus méritos académicos y logros científicos, en el año 2008 fue nombrado Miembro Correspondiente de la Academia Chilena de la Lengua. El sitio en Internet del profesor Parodi es <http://www.giovanniparodi.cl/>.

---

\* Doutor pela UFMG. Professor adjunto do ILEEL/UFU.

\*\* Mestranda do PPGEL/UFU.

### **1. Profesor Parodi, ¿cuáles han sido sus actuaciones más destacadas como lingüista y, más específicamente, como lingüista de corpus?**

R: Pensando como lingüista y como lingüista de corpus, y haciendo la conexión entre ambos, hay un asunto que me parece muy interesante y es pensar en el momento en que comenzamos a construir diferentes corpus, a construir bases de datos de textos escritos. Yo venía en ese entonces, y aún es una de mis áreas de investigación, desde la psicolingüística. Esto hace ya unos quince años atrás, y una de las primeras ideas que tuvimos fue tener un etiquetador morfosintáctico. Yo miraba con envidia (sana) los trabajos en inglés y cómo marcaban los textos. Para lograr este propósito, entonces, necesitábamos contar con grupos de documentos en formato *TXT*. Y en ese momento hubo una cosa que siempre comento, pues luego tendría gran impacto en un paso que daríamos años después. En ese momento contar con un gran corpus era un gran paso. Y decíamos, “Bueno, digitalicemos los textos y borremos de los textos todo lo que no sea sólo las palabras”, y yo siempre insistía, “Dejemos las palabras; así les decía a mis asistentes, porque lo que queremos son las palabras. Solo denme las palabras”. Y en ese momento lo que buscábamos obviamente era tener textos que fueran en formato plano para poder aplicarles algún tipo de etiquetador, y ese fue un desafío tremendo, aprendimos todos a digitalizar y vimos todo lo difícil que era llevar adelante esa tarea, y luego surgió la idea de: ¿y qué hacíamos con estos documentos digitales? ¿Dónde poníamos estos documentos? Y, entonces, tanto la idea de desarrollar un etiquetador morfosintáctico y la necesidad de un espacio de almacenaje y consulta, nos llevaron a un hito muy importante como fue la creación del sitio web [www.elgrial.cl](http://www.elgrial.cl).

Debo reconocer, sin falsa modestia, que en mi carrera profesional y académica existen varios hitos y momentos maravillosos. Uno de ellos que marca un antes y un después, ciertamente lo constituye el momento en que lanzamos El Grial. Fue un trabajo mancomunado con muchos colegas y grupos de investigación. Por ello, cuando miro hacia atrás y observo las cosas que hemos realizado y veo a toda la gente que se ha unido y con la que hemos trabajado y seguimos juntos en torno a El grial y sus proyecciones; entonces veo que construir El Grial y desarrollar el sitio web donde se alojan todos estos corpus, solo puedo decir que estoy seguro ha sido un desafío. Ha sido algo muy potente para desarrollar esta línea de investigación. Así, logramos tener un etiquetador, y luego construir la interface. No tan sólo un sitio donde colgáramos los textos y tuviéramos un etiquetador, sino que construir una interface que fuera realmente una plataforma que permitiera visualizar los corpora que están

allí y permitiera también hacer consultas, recuperar la información que está alojada ahí, y luego ir avanzando en establecer diferentes tipos de consultas. Bueno, eso ha sido un trabajo largo en varias etapas y varias fases en que han intervenido diferentes equipos de trabajo. Y, a veces, hay quienes dicen, pero bueno, no necesitamos tener un etiquetador morfosintáctico, podríamos hacer análisis basados en otros elementos. En fin, pero yo sigo pensando que el contar con un etiquetador morfosintáctico es necesario.

Lo que nosotros echábamos de menos era una herramienta para poder trabajar con el español, y una herramienta que, además, como hemos visto en las conversaciones en estos días de congresos aquí en Uberlândia, existe consenso en esta necesidad, que es compartida con los investigadores aquí en Brasil, lo que buscamos son herramientas disponibles gratuitamente. Por ello, desde la psicolingüística es muy valioso contar con investigación empírica desde la LC. Yo había sido un psicolingüista que estaba fuertemente concentrado en los procesos de producción y comprensión de textos y, no había visualizado hasta ese momento, como comentaba ayer en el marco del congreso, el estudio de los corpus, y gracias a Douglas Biber, debo reconocer, al descubrir el libro de Biber de 1988, visualizo ahí un campo de acción muy diferente. Creo que eso ha fortalecido la investigación en que hacemos psicolingüística, porque podemos investigar fenómenos que estén mucho mejor documentados, fenómenos que han sido indagadas descriptivamente en textos y luego podamos generar entonces instrumentos que tienen un basamento empírico muy importante que permiten hacer investigación psicolingüística. Este círculo virtuoso es muy robusta, en mi opinión, para una línea de investigaciones interdisciplinarias. Entonces, desde la psicolingüística, hemos ido a la lingüística de corpus y hemos vuelto a la psicolingüística.

**2. Como miembro de la Sociedad Chilena de Lingüística y de la Academia Chilena de la Lengua, cómo ve su participación desde el área de la lingüística de corpus en el ámbito de las organizaciones ya mencionadas. ¿Cómo lo ven sus pares? ¿Cuáles son sus aportes más específicos en esta área?**

R: Es difícil saber cómo lo ven los otros a uno, aunque algunos dicen que nosotros nos conocemos a partir de cómo los otros nos dicen que somos; entonces, cómo se construye el imaginario de uno mismo, a partir de lo que los otros nos dicen o uno mismo tiene su propio yo, digamos, ¿mi yo cuánto se alimenta del otro? Bueno, eso es casi filosófico.

Tal vez es más fácil comentar algunas acciones que hemos emprendido al interior de la Sociedad Chilena de Lingüística. Una cosa que hemos ido haciendo, a la luz de lo que recién conversábamos, es ir creando espacios para la Lingüística de Corpus (LC) y los estudio en esta línea de investigación, de modo que muy rápidamente fuimos proponiendo mesas de trabajo, o talleres, o seminarios, que fueran complementando los temas que se trataban clásicamente, la Sociedad Chilena de Lingüística, en general es una sociedad que ha estado, digamos, siempre ha tenido un área de psicolingüística, un área de gramática, un área de discurso, un área de análisis del discurso (más nueva por supuesto), pero obviamente no tenía un área de lingüística de corpus y yo creo que cuando uno mira el programa del congreso bianual de la Sociedad Chilena de Lingüística, cuya última reunión fue el año pasado en Concepción, uno observa un cambio en las temáticas en la forma de hacer investigación, yo creo que ahí aparece el rol de la LC, y yo creo que está gatillado también, y da mucho gusto, uno lo dice con orgullo, ver que muchas de las personas que están allí hoy en día trabajando son ex alumnos de nuestros programas de pregrado y de nuestros postgrados de magister y de doctorado. Entonces uno ve como una idea que parece tan pequeñita en un comienzo va creciendo y creciendo, y hoy en día esas líneas de investigación son fuertes.

Ahora bien, dentro de la Academia Chilena de la Lengua, existe la Comisión de Lexicografía, y en su interior se ha ido fortaleciendo el trabajo a partir de corpus y de herramientas tecnológicas. Así, progresivamente se ha impulsado la mirada desde los corpus para las revisiones que se hacen. Como seguramente ustedes saben, hoy en día el diccionario, y la gramática y la ortografía, se construyen desde las academias, de todas las academias de la lengua española. Ha habido un cambio aquí muy importante y estos cambios se han debido básicamente a la creación de la Asociación de Academias de la Lengua (ASALE) de modo que, la Nueva Gramática, la Ortografía y el Diccionario de la Lengua en las últimas versiones que tienen, no son ahora de la RAE sino que son de la Asociación de Academias de la Lengua, que son veintidós, donde la RAE es una más que se sienta en igualdad de términos a trabajar en este sentido. Esto ha sido muy potente. De hecho, el Diccionario de la Lengua Española (2014) que acaba de salir en su vigésima tercera edición, y tuvimos el placer de presentarlo en Valparaíso, la semana pasada, tan recientemente así, en la Pontificia Universidad Católica de Valparaíso, en su estreno en Chile, porque recién esta semana es su presentación oficial en el marco de la Feria Chilena del Libro, pero nosotros adelantamos la presentación oficial, y una cosa relevante que dijimos allí es que se llama Diccionario de la

Lengua Española, o sea, se llama DLE, y hay que trabajar para posicionar este nombre, ya no tenemos que llamarlo más DRAE (Diccionario de la Real Academia Española), y es muy importante, las palabras sí importan y como lingüistas tenemos que saberlo. Este diccionario es comunitario y, en el estricto sentido de la palabra, está elaborado por todas las academias; entonces, aprovecho de comentar esto porque creo que es un elemento muy importante para los latinoamericanos en general y sobretodo que el DLE tiene claramente hoy en día, y cada vez más asentado, la marca panhispánica.

Ahora bien, al interior de la Comisión de Lexicografía de la Academia Chilena de la Lengua, donde se preparan las revisiones de las nuevas palabras que probablemente se incorporen al diccionario y se revisan las existentes; allí se ha ido acentuando una mirada desde los corpus y su empleo en el trabajo que se realiza cada quince días. Y no tan solo ya basados en la experiencia de algunos académicos, que sigue siendo una fuente valiosa de información fundamental, sino también desde corpora digitales y disponibles en línea. De este modo, el trabajo en la Academia se apoya en nuevos recursos y en nuevos métodos más tecnológicos.

En un sentido más amplio, veo que en Chile se ha venido experimentado progresivamente un giro importante hacia los trabajos de corpus, lo que no quiere decir que tradicionalmente no hayan existido investigaciones que emplearan corpus, ya que siempre han estado presentes en la investigación lingüística. Me refiero al giro contemporáneo que la Lingüística de Corpus implica y también a la Lingüística Computacional. Esto se nota en las publicaciones que se incrementan en torno a una línea de mayor sustento empírico y con mayores desarrollos tecnológicos. También se aprecia, por ejemplo, en aumento sostenido de los proyectos con financiamiento externo a las universidades, tal como CONICYT y FONDECYT. Todo ello ha llevado a ampliar los textos en análisis y a superar estudios, por ejemplo, únicamente a partir de ejemplos ad hoc. Estas líneas de investigación son las que, entre otras, hemos venido impulsando fuertemente desde Valparaíso a través de diversos frentes de acción; en particular, desde la Escuela Lingüística de Valparaíso (ELV, [www.elv.cl](http://www.elv.cl)) y desde los Programas de Postgrado en Lingüística ([www.linguistica.cl](http://www.linguistica.cl)). En mi opinión, los enfoques desde la LC son temas muy relevantes que se van siendo cada vez más notorios y que, al mismo tiempo, va reforzando la idea de que los nuevos investigadores y estudiantes más jóvenes se ven muy motivados por estas líneas de investigación. Así, se va estableciendo con mucha fuerza que si tengo una hipótesis y quiero investigar un determinado

fenómeno lingüístico, ya no es estudiar un documento de dos páginas, sino que ver la posibilidad de construir un pequeño corpus o un gran corpus, o trabajar con un corpus disponible y hacer un estudio a partir de aquello; entonces yo creo que ha habido un cambio importante y creo que Chile está avanzando fuertemente en esto, al igual que otros países como Brasil y el mundo.

### **3. En sus palabras ¿Cómo definiría un Corpus y la Lingüística de Corpus? ¿Qué lugar le asignaría a la Lingüística de Corpus en el ámbito más amplio de la Lingüística?**

R: ¿Qué es un Corpus? Hay tantas definiciones de qué es un Corpus ya en la bibliografía que no parece necesario elaborar una nueva definición de qué es un Corpus, más bien conviene recordar qué han dicho o que se ha escrito al respecto. Una cosa interesante es que un corpus no es un texto ni un documento sino que un corpus se supone que es una colección grande o pequeña o mediana, porque corresponderá a ciertos parámetros que lo podrá hacer variar, de modo que el tamaño no es necesariamente una variable creo yo, aunque desde la Lingüística de Corpus siempre se ha dicho “más es mejor”, o sea está este principio donde sí incide el tamaño de un Corpus supuestamente, pero también está claro que hay corpus especializados y que por la naturaleza de esos textos puede ser reducido. Entonces, el tamaño, definitivamente, no es lo que define un corpus, aunque ojalá sea de la mayor extensión posible, nos da eso sí un tema más de representatividad posible. Pero el corpus tiene que ver con una selección de criterios que estimo son relevantes, como que podamos saber a qué responden, o sea, quiénes han producido estos textos, en qué contextos se han producido estos textos, qué nos llevó a recolectar estos textos. Hoy en día podríamos agregar la disciplinaridad como un tema importante, si es que son de alguna disciplina particular, si son orales o escritos, entonces hay una serie de parámetros que podríamos establecer para definir en este sentido qué es un corpus. Es una colección de documentos que debe estar alineada con algunos principios que tienen que ver con la recolección misma de este, de modo que no debiera tener una heterogeneidad en su carácter, sino que debiera tener algunos principios unificadores en este sentido y debiera, por ejemplo, responder a algunas preguntas como ya decíamos que tiene que hacerse el investigador que lo recolecta. Ojalá tuviéramos la mayor cantidad de información posible para poder saber qué es un corpus, o qué corpus es el que estamos juntando; también debiéramos tener algunos parámetros ecológicos, o sea, ojalá que los textos que conforman el corpus sean textos completos de la variedad que estamos recolectando, o

sea, si son conversaciones de algún tipo de registro oral, ojalá sean conversaciones completas, o sea que de acuerdo a los propósitos comunicativos que la guían, tengamos desde el principio al fin de esa conversación, donde los participantes, cuantos sean, sean los que están registrados. Si es un corpus de textos escritos, el ideal, sino la obligación, creo yo hoy, es que los textos sean completos o sea, sean textos donde está toda la información de principio a fin. Hay hoy en día una definición que hace rato da vuelta que es la que yo no creo mucho que es la de “elimine el paratexto”, y establecen una distinción entre el texto y el paratexto, y suponen entre otros que la portada, que las referencias bibliográficas, que toda la primera parte, por ejemplo, la editorial, el país, los autores de un documento escrito no son parte integral de un texto, sino que el texto comenzaría cuando el documento en sí comienza en la primera línea propiamente dicha. Yo creo que esas distinciones no corresponden, creo que un texto es todo lo que tenga como texto, y si lo que estoy tomando es un documento por ejemplo como un libro, irá todo lo que tenga que ver con la editorial, el lugar de impresión, con el número del tiraje, y eso es parte del texto propiamente tal, o sea, no son elementos “paratextuales”, si está una mini biografía en la solapa del autor, algunos consideran que eso no es parte integral del texto, pero a mí me parece que sí, si esta es una obra de un determinado autor eso es parte, entonces pensemos en una colección de textos donde los textos sean completos, tengan toda la “completitud” que corresponde y no lo sesguemos, sobre todo desde la teoría del género, que es fundamental pensar que los textos deben ser situados y completos en este sentido y que tengamos la mayor información respecto de qué los caracteriza. También hoy en día diríamos que un corpus debe ser digital. En general, un corpus no es una colección de textos en papel, un rasgo hoy, porque la LC está asociada al tratamiento de los corpus en términos computacionales, lo que no quiere decir que todo tratamiento de corpus deba -en mi opinión- ser obligatoriamente computacional. En mi opinión, una rica combinación puede darse entre investigación cualitativa y cuantitativa, pero sí es importante que los textos estén en formato digital, porque pensamos en que deben estar colgados en algún sitio, que sean accesibles a procesadores de algún tipo.

En mi opinión, no es un principio obligatorio que todo análisis de corpus deba obligatoriamente tener un tratamiento estadístico y computacional, yo creo que la LC también tiene que estar abierta a tratamientos de esos corpus que puedan obedecer a principios más cualitativos, a investigaciones que tengan otros fines, no la cuantificación per se como un único objetivo. Un ejemplo, cuando estudiamos desde la teoría del género, los rasgos retórico

funcionales desde la perspectiva de Swales, ese estudio normalmente es posible hacerlo cualitativamente, o sea es un estudio que tenemos que hacerlo más bien manualmente, y que lo que busca es determinar organizaciones funcionales que puedan reconocerse en los patrones retóricos de esos textos, como clásicamente la introducción, la discusión, la parte metodológica, en fin, en un artículo de investigación. Es cierto que también puede estudiarse mediante técnicas computacionales automatizadas, aunque también es posible que se haga en un trabajo manual, de modo que yo lo veo reñido el trabajo cualitativo manual con el trabajo computacional estadístico, cuantificable, creo que se nutren ambos, por supuesto que hay una Lingüística de Corpus estadística per se, lingüística cuantitativa como lo llaman algunos también, y que sólo concibe una mirada muy cuantificable de los elementos textuales, pero yo creo que esa mirada se enriquece desde otras perspectivas también.

#### **4. ¿En qué lugar posicionaría la Lingüística de Corpus pensando en la gran área de la Lingüística?**

R: Si se mira, en un sentido diacrónico, no cabe duda de que el giro hacia la LC ha alcanzado un impacto muy significativo. En un periodo de tiempo relativamente breve (treinta a cuarenta años), los desarrollos en términos de sociedades nacionales e internacionales, revistas indexadas, grupos de investigación, series de libros, espacios al interior de otras asociaciones tradicionales, proyectos de investigación nacionales e internacionales y estudios multilinguas han alcanzado un impacto muy significativo. Al mismo tiempo, la LC ha ido ganando un espacio de relevancia dentro de las áreas de los estudios del lenguaje, tal como fácil se comprueba en fondos de financiamiento de proyectos, en congresos, en revistas, entre otros. No es una opción encapsulada, lo cual es muy bueno. Ahora, todo ello depende mucho de cómo entendamos lo que es la LC. En mi opinión, la LC es una metodología muy potente, para mí la LC no es una teoría. Es una forma de trabajar, no tiene necesariamente un conjunto de principios teóricos, de modo que yo puedo hacer LC desde la lingüística sistémica funcional, desde la teoría del género, o desde otras perspectivas lingüísticas. Por tanto, cuando lo decimos así, emerge claramente la idea que la LC no es un paradigma lingüístico en el sentido kuhniano, o sea que tenga un objetivo, un objeto y un método específico, y que tenga unos principios teóricos que la orientan; creo que es más bien una forma de aproximarse al tratamiento de la información que se recolecta, y eso hace que sea más poderosa aun, porque puede estar transida en todas las áreas de investigación lingüística, a lo mejor la lingüística

generativa puede estar reñida porque tiene una idea de la idealización, de no estudiar el uso, y la LC está comprometida con el estudio de uso y la variación, y yo creo que entonces casi todos los otros movimientos lingüísticos no tienen problema en asociarse con la LC. Así, pienso que la LC ha tenido un proceso creciente y yo le auguro un futuro promisorio en ese sentido.

### **5. Según su experiencia como profesor universitario ¿Cuál es el espacio que ocupa la LC en los programas de estudio a nivel nacional en Chile e internacional?**

R: En este momento, a pesar de lo que decía más arriba, soy de la idea de que en Chile aún el espacio de la LC es reducido. Siendo muy honesto, creo que todavía está en ciernes este tipo de trabajo y, sobre todo en Chile, hemos tenido un proceso largo de reacomodación de currículos universitarios, a través de la discusión de qué tipo de profesionales formamos en las carreras de pregrado, donde en nuestros ámbitos están la pedagogía en castellano, la pedagogía en inglés, la traducción, los periodistas, entre otros. La cuestión de fondo es si estamos formando profesionales o si estamos formando investigadores, y en Chile existe una cierta tensión entre la idea del investigador al alero de las licenciaturas y una relación pensada con el profesional que tiene que hacer en el mundo laboral. Entonces muchas veces hay una formación más pedagogizante, más didactizante en algunos casos y menos investigadora de la lingüística per se; ciertamente se puede hacer investigación acción, investigación en el aula, investigación didáctica, o de lingüística aplicada, pero mi visión en este momento es que todavía en Chile, visualizo los estudio de corpus a nivel de formación de grado están lentos y tienen una incorporación ingenua o inicial, donde estamos todavía tratando de generar un espacio en ese ámbito. El más fuerte es en el ámbito de las licenciaturas donde está más asentada la idea de formar investigadores y, por ende, hacer investigación. Paralelamente, en algunos programas de magister y doctorado la LC ha tenido otro derrotero y sí ha encontrado adeptos muy entusiastas. En estos nichos creo que sí ha calado dentro de las universidades, pero –al mismo tiempo- creo que todavía está en franco desarrollo.

A pesar de que en Chile existe un número importante de universidades en que se desarrollan estudios lingüísticos, hay muchas universidades en que creo que lamentablemente la LC no se visualiza. No obstante ello, a mí me parece que esas son grandes oportunidades y quiero verlo más bien como una cosa positiva, una gran oportunidad en la cual tenemos una tarea. A mí me encanta encontrar desafíos, encontrar tareas que hacer, porque entonces si

estuviera todo hecho, tendría que casi jubilarme, pero como veo que hay mucho que hacer, encuentro motivaciones, encuentro nuevas tareas, me doy nuevos objetivos, y por ejemplo creo que este es un gran tema, situar más fuertemente y pensar como uno puede incluir en la investigación de grado. Si bien en maestrías y doctorados tiene mucha más potencia, creo que en la formación de grado hay que sentar las bases, hay que motivar a los alumnos inicialmente; mostrarles las potencialidades que hay en estos estudios, y cuando lo hemos hecho, tenemos grandes logros. Yo, con mucho orgullo, tengo en los últimos años publicaciones con alumnos de grado, y puede parecer curioso que en estos últimos años no tenga publicaciones con alumnos de maestría y doctorado, pero sin embargo tengo publicaciones con alumnos de grado. Y lo valoro grandemente

### **5.1 ¿Y a nivel internacional?**

R: Bueno, separemos Latinoamérica del mundo, creo que en la lingüística y tal como lo hemos discutido en este congreso aquí en Brasil, hay que mirar a la LC, desde diferentes ángulos, un ángulo de donde la podemos mirar creo yo es desde la lingüística, sólo desde la lingüística porque ciertamente la LC tiene otros desarrollos desde la lingüística computacional, desde el procesamiento natural del lenguaje, y eso no corresponde ciertamente tal vez a lingüistas, pero desde la lingüística solamente, creo que dentro de Latinoamérica su impacto es lento todavía, creo que nuevamente ahí tenemos mucho que hacer. Si uno toma una asociación importante en Latinoamérica, todavía es poco el impacto.

Los eventos donde se hace LC en Latinoamérica creo que son pequeños en términos cuantitativos, justamente en cuanto a número de personas, a cuántos somos los investigadores que estamos ahí y presentamos trabajos. En cuanto al trabajo con corpus y de recolectar grandes corpus y la investigación con apoyo tecnológico, veo que aún esta línea es escasa. Existen sí grupos de investigación; identifico claramente gente en Brasil, Argentina, México y Chile, pero menos en otros países latinoamericanos, nuevamente debo decir que eso me parece un desafío, creo que tenemos un compromiso con otros países latinoamericanos, creo que España ha avanzado bastante, hace diez años atrás no existía mucho en esta línea de investigación, hoy en día se ve bastante, en algunas reuniones en las que he estado he visto, hay un desafío grande con esto, veo que está cambiando y ojalá que cambie más.

En síntesis, en términos generales, tal como se decía más arriba, la investigación en LC en el nivel internacional también es variada. En algunos países de Europa, ha sido un

proceso muy dinámico y muestra impactos fuertes. Lamentablemente en otros espacios internacionales se detecta escaso o casi nulo desarrollo de la LC. En lengua española, ha habido en los años recientes un importante avance, aunque en algunos países se nota mucho mayor interés que en otros. En parte, como decíamos más arriba, esto más que ser negativo, debería ser una motivación para crear e impulsar nuevos espacios de interacción.

## **6. ¿Cómo proyecta el futuro de la Lingüística de Corpus de aquí a diez o quince años más?**

R: La LC, como ya lo hemos comentado en este congreso aquí en Uberlândia, tiene varios desafíos, como partí diciendo al inicio de esta entrevista, cuando comencé a trabajar con LC yo decía “denme las palabras, y borren todo el resto”, yo creo que el gran descubrimiento, más bien el redescubrimiento, o como algunos lo llaman: “el descubrimiento de lo obvio”, es el descubrimiento de la multimodalidad de los textos, o sea que los textos no son sólo palabras. Esto puede parecer una cosa obvia: que los textos no son sólo palabras no es nada nuevo, porque claro, cuando yo decía “denme sólo las palabras y borren todo el resto”, estaba aludiendo a que yo veía que en los textos habían otras cosas más que palabras, pero mi motivación y la de mi equipo eran tan grande en ese momento por construir corpus digitales marcados, por tener un etiquetador y una interfaz de consulta, que estábamos obsesionados con la idea de las palabras y de poder decir: este es un sustantivo y esto es un pronombre de primera persona singular; poder tener el conteo de las formas y de las categorías gramaticales, que cuando decíamos denme todas las palabras y borren todo el resto, visualizábamos que los textos eran más que palabras, pero nuestro cometido estaba con las palabras. Pero, hace ya algunos años, también se me enfrentó un texto que me persiguió y me hizo ver que lo que estábamos borrando era parte integral de los textos, por lo tanto los textos no son sólo palabras obviamente, pero sobre todo si nosotros pensamos que los seres humanos hablamos en textos y, más aun, hablamos en géneros y nos comunicamos en textos y en géneros, entonces los textos y los géneros tienen mucho más que palabras. Como sabemos, para los seres humanos el centro de la comunicación es la construcción y transmisión de significados, y los significados no los podemos construir sólo de palabras. Y esto quiere decir que si alguien está interesado como yo en los textos y los géneros, y por supuesto desde la LC, tenemos el gran desafío de poder dar cuenta de cómo son realmente los textos. Esto quiere decir que la LC va a tener que desafiar a sí misma y tendrá que pensar en una LC

Multimodal, o una visión multisemiótica del lenguaje desde la LC. Así, el sistema verbal constituye sólo uno de los sistemas constitutivos de los textos, por lo tanto si la LC quiere hacerse efectivamente cargo de los textos, los textos no son sólo palabras. Tenemos, entonces, que poder desarrollar herramientas que logren identificar los otros sistemas, los sistemas gráficos, sistemas visuales, sistemas matemáticos, sistemas del color, que es tan fundamental por ejemplo si pensamos en los gráficos.

Todo esto de la multimodalidad quiere decir mucho para la LC. Así, si queremos dar cuenta del significado que está en los textos, yo creo que un desafío maravilloso para la LC es pensarse como una LC Multimodal, y eso creo que son unos diez o quince, o hasta treinta años por lo menos. Es cierto que ya hay especialistas buscando crear programas computacionales que logren identificar un gráfico, una tabla, una ilustración, de modo que puedan tener un marcaje; algunos piensan que el XML puede ser una forma de trabajo para ayudarse con esto. Pero la verdad es que el desafío es de proporciones y más vale tener claro el escenario. Aún no se logra dar cuenta efectiva de las relaciones semánticas entre oraciones de un texto y, por ende, automatizar este tipo de procesos constituye un desafío tremendo que no está logrado aún. Esa es mi opinión en este momento. Creo que aún no logramos capturar sistemáticamente los mecanismos textuales de, por ejemplo, la construcción de la coherencia textual. Por lo tanto, pensar en cuáles son las relaciones semánticas que conectan una ilustración con las palabras de un determinado texto, o sea como el sistema verbal interactúa con los sistemas gráficos, visuales, matemáticos, entre otros. Por ejemplo, un texto de economía donde mucho de lo que se dice está en fórmulas y gráficos, no puede analizar solamente las palabras. Se debe analizar los gráficos y las tablas y se debe analizar, además, como se construye el significado en la intersemiosis entre esos gráficos y esas tablas y el sistema verbal. Suena maravilloso, o sea, hay una tarea enorme, mucho trabajo por delante. El futuro de la LC es tremendo.

**7. Con respecto al interfaz computacional El Grial, pensando en la perspectiva del aprendizaje y enseñanza de la lengua española, tanto como lengua materna como lengua extranjera, y en la investigación, ¿nos podría contar un poco como funciona dentro de este ámbito?**

R: El Grial tiene fortalezas tremendas, pero en el contexto mundial, está pensado solo para una lengua que es el español y entonces pensando en las lenguas extranjeras o la investigación en otras lenguas El Grial no ofrece posibilidades en este momento, y tampoco puede hacer nada con respecto de los textos multimodales, o sea, sólo trabaja con las palabras, tal vez eso es una fortaleza tremenda, y que lo haga bien y que se haga cargo de eso, yo creo que es un tremendo desarrollo y hay que perfeccionarlo y mejorarlo, pero también es interesante que El Grial debe desafiarse a sí mismo.

Esta herramienta nos ha apoyado de manera muy importante y ha dado impacto a nuestras investigaciones en LC. Sin El Grial, muy probablemente no habríamos alcanzado los entusiasmos y compromisos de muchos colaboradores y estudiantes. Ha sido, como ya decía, un antes y un después en nuestra investigación en la Pontificia Universidad Católica de Valparaíso. Sin duda, también nos ha aportado mucho a la visualización de los que hemos hecho y hacemos en la ELV. Espero en sus siguientes fases, El Grial pueda ser una herramienta que también apoye los procesos de enseñanza aprendizaje del español y pueda acoger corpus de aprendientes, por ejemplo.

### **8. ¿Podía hacernos un breve resumen de qué se trata El Grial?**

R: El Grial es una interfaz computacional que está en internet y que, desde uno de sus objetivos, se presenta como de acceso abierto y gratuito. Básicamente lo que hace es que permite etiquetar morfosintácticamente los textos que se suben en formato plano; también permite colgar estos textos allí dentro y permite acceder a estos textos a través de diferentes herramientas de búsqueda para recuperar información en formato de consultas específicas. Es posible hacerlo por medio de formas o de categorías, o de cadenas de categorías o de cadenas de formas, entre otras. O sea, permite a través de la interfaz, en lo que denominamos la Búsqueda Compleja o, a través de lo que llamamos El Manchador de Textos, buscar alguna forma determinada, por ejemplo puedo buscar la palabra “para” específicamente, pero puedo buscar la palabra “para” mezclada con un adjetivo y mezclada en la misma cadena, seguida de un adjetivo y de un sustantivo por ejemplo. Entonces estoy definiendo en la cadena de búsqueda una forma y dos categorías gramaticales. Esto solo a modo de ejemplo y muy resumidamente.

### **8.1. Pensando en las ventajas para el aprendizaje de lengua, específicamente del español como lengua materna y como lengua extranjera**

R: En este momento yo diría que El Grial, tal y como está, no es una herramienta que potencie el aprendizaje y enseñanza de lenguas. Es una herramienta más bien de investigación; está proyectado, como decía anteriormente, que a finales del 2015 tenga alguna contribución en torno a los llamados corpus de aprendices o aprendientes, y vamos a comenzar a hacerlo con una disciplina que es la disciplina de economía, pero en este momento tal y como está es una herramienta para que el investigador pueda testar sus hipótesis, comprobar, corroborar o indagar sus hipótesis, tener datos empíricos y a partir de estos datos que recupera un investigador. En virtud de esto, un investigador puede crear material o tomar decisiones metodológicas de enseñanza aprendizaje, pero no es una herramienta en sí misma -en este momento- construida para que un aprendiz pueda acercarse a ella y obtener la información para el desarrollo de su lengua materna o español como lengua extranjera. Sin duda, existe entonces un nicho importante que se puede desarrollar y que estamos pensando comenzar a explorar muy tentativamente.

### **9. ¿Qué mensaje le dejaría a un estudiante que se comienza a interesar por la LC?**

R: Estoy seguro de que un camino certero es mostrarles que existe un asunto que a mí me motivó a comenzar a estudiar lingüística. Yo era ayudante del curso de gramática del inglés, con mi maestra, y observé en la lengua inglesa la existencia de regularidades, de tendencias pero también de variaciones. Y esto me obsesionó, me apasionó y lo hace hasta el día de hoy. Creo que una cosa que les podemos enseñar a los estudiantes noveles, que le podemos mostrar a los alumnos que están recién ingresando, es cómo la LC, como una herramienta, como una metodología de trabajo, puede permitir descubrir la maravilla que hay en el lenguaje, en las tendencias, en las regularidades, pero al mismo tiempo de las variaciones y como ciertas tendencias y regularidades varían. Cambiar los registros, los géneros, las disciplinas, y así observar cómo se construyen los discursos en ciertas disciplinas, en ciertos contextos, cómo esos contextos estimulan la variación de los rasgos léxico-gramaticales en la construcción del significado. Estoy convencido de que si un estudiante que está recién ingresando a la investigación visualiza las regularidades, la idea de encontrar tendencias, pero a la vez encontrar también “no tendencias” o de encontrar variaciones entre las tendencias, yo creo que se ve un escenario muy interesante de investigación, y yo creo que eso le permite mostrar

la riqueza y la variedad de las lenguas, lo cual a mi me parece que es una cosa fundamental. Mostrar que cada lengua en todas sus dimensiones es muy rica, es inconmensurable. Aportar así a la formación de las nuevas generaciones de profesionales y de futuros investigadores constituye un desafío muy motivador en el que desplegar todas las potenciales en conjunto es una proyección potente.

Al concluir esta última pregunta, deseo agradecer de manera especial a Ariel Novodvorski por su generosidad y compañerismo al invitarme a Uberlândia a dos magníficos eventos académicos a través de los cuales he podido conocer personas magníficas y, al mismo tiempo, aprender muchas nuevas ideas acerca de la LC y de los desarrollos en Brasil en esta línea. También y no menor, le agradezco a Ariel haber organizado esta entrevista que me brinda la posibilidad de poner en palabras orales y luego por escrito algunas nociones y reflexiones en torno a un área de las ciencias del lenguaje, con la que tanto él como yo estamos muy comprometidos en nuestras investigaciones: la LC. Les dejo desde este foro, un abrazo fraterno.

Uberlândia, Brasil, 5 de noviembre, 2014.