

# Lingüística de corpus utilizada en la concepción de plataformas de formación de docentes de lenguas<sup>1</sup>

Jorge Mauricio Molina Mejía

LIDILEM – Université Grenoble Alpes, *Corpus Ex Machina* – Grupo de Estudios Sociolingüísticos – Universidad de Antioquia

In this article we present a study about corpus annotation, created in order to establish a Computer Assisted Language Learning system. This system is intended to the linguistic training of future French as a Foreign Language teachers in Colombian universities. We have a tagged corpora using Natural Language Processing methods for the morphological and syntactic aspects, and manual annotations to aim at the notions studied by text linguistics. The result of these annotations provide a serie of activities that are taken into account for ELiTe-[FLE]<sup>2</sup> informatic system. Those activities look for the improvement of the linguistic level of the Colombian training students through the study of linguistic notions contextualised in the texts.

**Keywords:** Corpus annotation, Corpus Linguistics, NLP (Natural Language Processing), FFL (French as a Foreign Language), CALL (Computer Assisted Language Learning), Teachers Training.

## 1. Introducción

“La linguistique de corpus a été développé pour extraire d’un corpus les connaissances linguistiques nécessaires à l’enseignement des langues ; un corpus étant un ensemble collecté et ordonné de données langagières réelles” (Teubert, 2009: 1)<sup>2</sup>.

<sup>1</sup> Este artículo nace a partir de una presentación en un coloquio internacional sobre lingüística de corpus, presentado en Orléans (Francia) en septiembre de 2015.

<sup>2</sup> “La lingüística de corpus fue desarrollada para extraer de un corpus los conocimientos lingüísticos necesarios para la enseñanza de las lenguas; un corpus es considerado como un conjunto coleccionado y ordenado de datos lingüísticos reales” (Nuestra traducción de Teubert, 2009: 1).

Según los criterios señalados por Teubert, y en el marco de realización de una plataforma para la formación de los futuros docentes de lengua francesa en Colombia (EliTe-[FLE]<sup>3</sup>)<sup>3</sup> que elaboramos (Molina Mejía, 2014; 2015; Molina Mejía & Antoniadis, 2014a; 2014b; 2014c; 2015); consideramos interesante comenzar la investigación a partir de la premisa “de enseñanza de conocimientos lingüísticos a partir de datos lingüísticos reales”. Es por ello que decidimos constituir un pequeño corpus de análisis lingüístico fundamentado en datos reales o auténticos. La plataforma informática que proponemos permite la construcción de actividades específicas para la formación de formadores, basadas en el enfoque teórico conocido bajo el nombre de lingüística textual. Con esta finalidad constituimos una muestra de análisis de fenómenos del lenguaje sacados a partir de este enfoque teórico, y para ello utilizamos textos auténticos sacados de antiguos exámenes del DELF<sup>4</sup> y del DALF<sup>5</sup>.

## 2. Lingüística de corpus dentro de una didáctica de las lenguas

Según algunos autores, la noción de corpus o de lingüística de corpus puede parecer algo vaga, sobretodo, en algunos campos como en el de la lingüística francesa (Souque, 2014: 64). Para otros autores, incluso, existe el dilema de saber si se trata o no de un enfoque teórico, o más bien nos hallamos frente a un uso de carácter metodológico (Gries, 2010: 327-328). No obstante esta situación, es innegable la importancia adquirida en los últimos decenios por la lingüística de corpus (Williams, 2006; Parodi, 2008), lo que le ha permitido convertirse en un paradigma bastante importante en los estudios del lenguaje. De esta forma, el trabajo sobre corpus ha conocido un éxito creciente en el campo de la lingüística y de la didáctica de las lenguas en estos últimos años, como lo muestran Guiliani y Hannachi (2010: 145).

### 2.1 Lingüística de corpus y enseñanza de lenguas extranjeras

La noción de corpus electrónicos no es nueva en la didáctica de lenguas. En efecto, los primeros trabajos en este campo datan de los años 1990, y vemos así

<sup>3</sup> “*Environnement informatique fondé sur la Linguistique Textuelle, pour la Formation Linguistique des futurs Enseignants de FLE*” (Ambiente informático fundamentado en la Lingüística Textual, para la Formación Lingüística de los futuros docentes de FLE).

<sup>4</sup> “*Diplôme d'Études en Langue Française*” (Diploma de Estudios de la Lengua Francesa).

<sup>5</sup> “*Diplôme Approfondi en Langue Française*” (Diploma de Profundización de la Lengua Francesa).

a autores como Guy Aston (1995) que hacen referencia a la utilidad de emplear este tipo de corpus en los cursos de idiomas. En este sentido, Aston considera que el empleo de la informática basada en corpus para la enseñanza debería permitir a los aprendientes de lengua el tener acceso a colecciones de textos libres, o a materiales muy elaborados (Aston, 1995: 257). En el campo de la didáctica de lenguas, la lingüística de corpus puede ser definida como un campo que permite “d’enseigner tous types de langues, pour peu que l’on soit à même de constituer un corpus informatisé de la langue que l’on souhaite soumettre aux apprenants” (Guiliani & Hannachi, 2010: 145)<sup>6</sup>.

De hecho, al comienzo de los años 1990 va a surgir una corriente asociada a la lingüística de corpus, y que Tim Johns denominaría como DDL “*Data-Driven Learning*”, o en francés ASC (*Apprentissage sur corpus*) (Boulton & Tyne, 2014: 6)<sup>7</sup>. El objetivo del ASC sería, según Johns, “de eliminar al intermediario [dicho de otra forma, al docente] en la medida de lo posible, para dar al aprendiente un acceso directo a los datos” (Johns, 1991, citado y traducido por Boulton, 2007: 38).

En 2011, la autora Corinne Landure propone la práctica del ASC como una alternativa dentro de la enseñanza y el aprendizaje de lenguas extranjeras (Molina Mejía, 2015: 146). En este sentido, aquella autora señala que:

“La démocratisation actuelle des nouvelles technologies offre d’autres perspectives pour l’enseignement et l’apprentissage d’une langue seconde et pourrait contribuer à changer en profondeur la façon de penser et de concevoir l’enseignement” (Landure, 2011: 164)<sup>8</sup>.

En efecto, según esta autora, la idea sería de pasar de una enseñanza de tipo tradicional a una enseñanza en la cual los docentes tendrían un nivel de compromiso mucho más elevado con su profesión y, para ello, estos podrían adquirir un número mínimo de competencias en el campo de la informática (Landure, 2011: 164). En tal sentido, consideramos que además de la formación en el dominio de la informática, los docentes de idiomas deberían también

<sup>6</sup> “La enseñanza de todos los tipos de lengua, siempre y cuando se considere la constitución de un corpus informatizado de la lengua que se desea presentar a los aprendientes” (Nuestra traducción de Giuliani & Hannachi, 2010: 145).

<sup>7</sup> Esto equivaldría en español a “Aprendizaje sobre corpus”, término que emplearemos a partir de este momento.

<sup>8</sup> “La democratización actual de las nuevas tecnologías ofrece otras perspectivas para la enseñanza y el aprendizaje de una segunda lengua, y podría contribuir a cambiar en profundidad la manera de pensar y de concebir la enseñanza” (Nuestra traducción de Landure, 2011: 164).

formarse en la utilización y en la constitución de corpus. Proponemos, en pocas palabras, formar a aquellos docentes al ASC, como a un nuevo enfoque y paradigma de investigación, tal y como lo proponen Boulton y Tyne (2014: 6).

## 2.2 Lingüística de corpus y formación de formadores de FLE

Los corpus concebidos específicamente para la formación de formadores en lenguas, y muy particularmente en el campo del francés como lengua extranjera, son bastante escasos en nuestros días, como lo pudimos comprobar en nuestro reciente trabajo (Molina Mejía, 2015). Esto contrasta con el hecho de la existencia, en el terreno de la formación de los aprendientes de lenguas, de un número cada vez más creciente de trabajos en el dominio del ASC, como lo señalan Boulton y Tyne (2014). Pero, ¿por qué este desinterés en emplear corpus en el campo específico de la formación de los futuros docentes de FLE? Según lo observado a lo largo de nuestro trabajo es más “fácil”, en cierta medida, constituir corpus de aprendientes para un público más extenso, que para un público mucho más específico como lo es el de los futuros docentes que se forman en medio *exolingüe* (es decir, formados en el terreno en el cual la lengua no es ni lengua materna ni lengua segunda) (Autor, 2015). En efecto, los problemas derivados de la formación de los futuros formadores dentro de un medio *exolingüe* son mucho más complejos que cuando los docentes se forman en un medio *endolingüe*<sup>9</sup> (o sea, la formación que tiene lugar en los lugares en los que la lengua a enseñar es lengua materna o lengua segunda). Para el primer tipo de formación se requiere no solamente de la formación en el terreno de la didáctica de la lengua objetiva a ser enseñada, sino que además es indispensable la formación lingüística que permita el dominio de esta misma lengua (Arismendi & Colorado, 2015; Molina Mejía, 2015). Todo esto demanda la realización de un trabajo muy enfocado en el desarrollo de la competencia lingüística, y esto a un nivel muy elevado. Es por esto que cuando hablamos de un corpus que está destinado a la formación de futuros docentes de FLE, para quienes el francés no es su lengua materna, ni tampoco es oficial en el país en el que aquellos están siendo formados, nos encontramos entonces frente a características muy específicas: el nivel de lengua es muy importante (los niveles con respecto al MCERL<sup>10</sup>, los futuros docentes de lenguas en Colombia,

<sup>9</sup> El proyecto *InUne* es un ejemplo de utilización de corpus para la formación de docentes de FLE en medio *endolingüe* (Guiliani & Hannachi, 2010). Este tipo de corpus tiene como fin ayudar en la formación de estudiantes para los cuales en su mayoría tienen el francés como lengua materna (L1).

<sup>10</sup> Marco Común Europeo de Referencia para las Lenguas.

por ejemplo, deben alcanzar los niveles B2 o C1 de este Marco para las lenguas), los aspectos gramaticales deben también estar comprendidos en la formación (morfología, sintaxis, ortografía, etc.).

No obstante, existen algunos trabajos que se interesan en la formación de los futuros docentes de lenguas, utilizando para ello el estudio de corpus lingüísticos y de corpus diferentes a los corpus lingüísticos. Se trata, en el primer caso, del trabajo sobre corpus para la formación de profesores de FLE de la Universidad de Bretagne Sud, el proyecto *IntUne* (Giuliani & Hannachi, 2010), y para los profesores de inglés en el LANSAD<sup>11</sup> de la Universidad París Diderot (Kübler, 2014). En este primer tipo de corpus existe un objetivo de formación por intermedio de nociones de orden lingüístico (gramática de la lengua objetiva, por ejemplo). En el segundo caso, se trata del análisis de corpus de diarios de aprendizaje (Cadet & Tellier, 2007) y de corpus que analizan las interacciones de la enseñanza en línea (Chanier & Ciekanski, 2010<sup>12</sup>). Este segundo tipo de corpus parte de la idea de la formación a partir de las interacciones entre los docentes y sus aprendientes. Hasta el momento no hemos encontrado corpus creados con la finalidad de realizar una formación lingüística de futuros profesores de FLE en un medio *exolingüe*.

Es por lo anterior que nuestra idea es la de trabajar a nivel de la anotación y de la puesta en ejecución de este tipo de corpus. Para ello, nos fundamentamos en un enfoque de aprendizaje sobre corpus, como el que fue expuesto en la parte anterior.

### 2.3 Lingüística de corpus y formación a la lingüística textual

El término de lingüística textual (o lingüística del texto) tiene su origen en los trabajos de Eugenio Coseriu de mediados de los años cincuenta, y sobre todo en un artículo inicialmente publicado en una revista alemana (Adam, 2010)<sup>13</sup>. En el cual nos habla de una lingüística del texto, emparentada con la estilística del habla, es decir, un campo que permitiría analizar específicamente el discurso hablado (Coseriu, 1969, p.289).

<sup>11</sup> “*Langues pour spécialistes d'autres disciplines*” (lenguas para especialistas de otras disciplinas).

<sup>12</sup> Se trata de un corpus multimodal de aprendizaje, recolectado en el marco del proyecto Mulce. <http://lrl-diffusion.univ-bpclermont.fr/mulce2/index.html>

<sup>13</sup> Nos referimos aquí al artículo “Determinación y entorno. Dos problemas de una lingüística del hablar”, publicado originalmente en 1955/1956 in *Romanistisches Jahrbuch* VII, pp. 29-54. Reeditado en 1969 en España, en la obra de Coseriu: “Teoría del lenguaje y lingüística general”, de la editorial Gredos.

Podemos trazar los primeros trabajos en el campo incipiente de la lingüística textual a partir de los años sesenta. Tenemos, en ese sentido, los primeros trabajos de Harald Weinrich (1964)<sup>14</sup>; este autor dedicó este primer trabajo al análisis de los tiempos verbales a nivel textual. Más adelante el semiólogo francés, Roland Barthes (1966), dedicará una parte de su trabajo al análisis discursivo del género relato. Los años setenta traerán algunos trabajos importantes en este campo: en primer lugar el de Teun Van Dijk (1972) sobre la gramática textual, y en segundo lugar el producido por Halliday y Hassan (1976) acerca del tema de la cohesión textual. Ambos trabajos abrirán las posibilidades para el estudio de la lengua desde el análisis de textos.

Es importante destacar todos los trabajos realizados a partir de los años ochenta y noventa, en los cuales se pretende, además, que sirvan para la enseñanza y aprendizaje de lenguas extranjeras y maternas. En tal sentido podemos citar los trabajos de Lita Lundquist (1980, 1990), el mismo Van Dijk (1983) y Jean-Michel Adam (1990). Los recientes trabajos de Lundquist y Adam se convertirán en referentes importantes, a nivel del estudio de la lengua francesa, a partir de la lingüística textual (Lundquist, 2013; 2014; Adam, 2011a; 2011b).

En lo referente a la relación entre la lingüística de corpus y la lingüística textual, algunos autores como Rastier (2004), Malrieu (2004), y Molina Mejía (2015), pretenden establecer un vínculo en común entre ambos enfoques teóricos. Es decir, de la lingüística de corpus como una herramienta que debería ser utilizada en el análisis de la lengua a partir del estudio de fenómenos de orden textual (*cf.* 3.2.2 § 3.3).

La parte que sigue presenta, *grosso modo*, la constitución del corpus para el sistema informático que concebimos.

### **3. Constitución de un corpus para la formación de docentes de FLE**

Desde el comienzo de la constitución de nuestro corpus, estuvimos confrontados con la dificultad de encontrar documentos auténticos clasificados con respecto a los niveles del MCERL. La tarea se hizo aún más complicada cuando comenzamos a buscar un corpus que no solamente estuviese clasificado por niveles, sino que además se encontrara ya anotado de forma morfosintáctica. Luego de una larga búsqueda encontramos en la página web del CIEP<sup>15</sup> una

<sup>14</sup> Aparecido en francés en 1973, bajo el nombre de “Temps”.

<sup>15</sup> “Centre International d’Études Pédagogiques” (Centro Internacional de Estudios Pedagógicos), <http://www.ciep.fr/>

pequeña muestra de antiguos exámenes clasificados por niveles<sup>16</sup>. En realidad, el problema esencial de la búsqueda de textos clasificados por niveles radica en el hecho de que estos deben ser libres de derechos de autor. Contrariamente a los exámenes del CIEP que se pueden utilizar libremente con fines educativos, existen algunos textos que clasifican documentos auténticos por niveles, como es el caso de la obra de Barthe y Chovelon (2009), pero que, infortunadamente, se encuentran bajo derechos de autor. En cuanto al hecho de tener un corpus que incluya textos y que se encuentren anotados siguiendo las características antes señaladas, no pudimos encontrar nada. Tuvimos, por lo tanto, que encontrar los textos y anotarlos nosotros mismos.

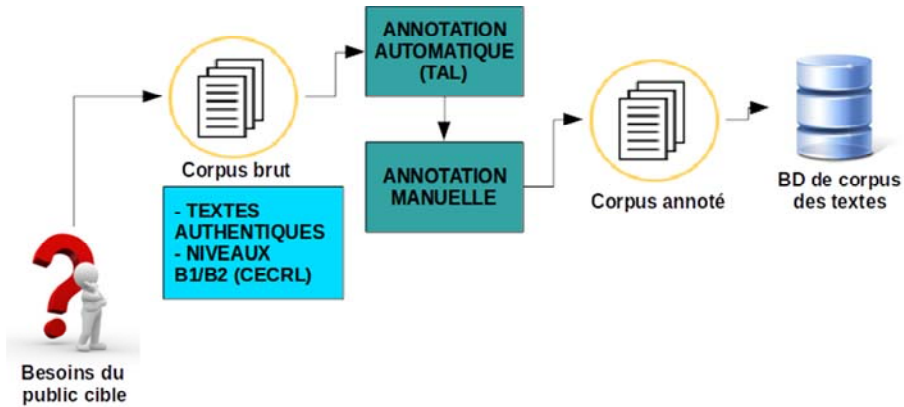
### 3.1 ¿Cuál corpus?

Para la creación (automática) de actividades, el sistema informático que concebimos (*cf.* sección 3) se apoya en las anotaciones que han sido introducidas en los textos. Es así como comenzamos por constituir una pequeña muestra con unos quince documentos extraídos de periódicos, de revistas, de obras literarias y de Internet. La idea era tener documentos auténticos y variados, que nos permitiesen el poder ofrecer a nuestro público objetivo, textos representativos de la lengua francesa. Poseemos, por el momento, 5 textos de nivel B1, 5 textos de nivel B2, y 5 textos de nivel C1<sup>17</sup>. Los textos comprenden: extractos de novelas, artículos periodísticos, textos de divulgación científica, extractos de ensayos, etc.

Nuestro corpus contiene alrededor de 5000 *tokens* (unidades) para el conjunto de textos de los tres niveles del MCERL. Durante el proceso de anotación, características tales como el tipo de texto, la fuente (nombre de la revista, periódico, etc.), el autor, etc., fueron indicadas en el “*entête*” (encabezamiento) de cada texto (*cf.* figura 3). En efecto, estos metadatos son esenciales durante la concepción de las secuencias pedagógicas, con el fin de permitirle a los docentes el poder tener todas las informaciones concernientes a los textos que se hallan en el sistema.

<sup>16</sup> Los textos encontrados en la página web del CIEP son de libre utilización con fines pedagógicos, y permiten el entrenamiento para los exámenes DELF y DALF.

<sup>17</sup> Según el MCERL, los niveles B1 y B2 corresponden al “usuario independiente” y el nivel C1 al “usuario experimentado”.



**Figura 1.** Anotaciones del corpus textual (Molina Mejía, 2015: 265).

### 3.2 ¿Cuáles anotaciones?

Luego de haber seleccionado los documentos, procedimos a anotarlos siguiendo dos tipos de enfoque: un primer enfoque basado en el TAL (Tratamiento Automático del Lenguaje natural), con el analizador *Cordial*; lo que permite después realizar un segundo etiquetado manual gracias al formato de balisado XML<sup>18</sup> (figura 1).

En la parte siguiente explicaremos brevemente los dos procedimientos de anotación, para ello nos apoyamos en capturas de pantalla que muestran, por un lado, las etiquetas del analizador morfosintáctico *Cordial* y, por otro lado, las anotaciones en XML. Finalmente, mostraremos la DTD<sup>19</sup> que posee la gramática de base de nuestro corpus. Una vez el corpus esté anotado y verificado gracias a la DTD, pasa a ser almacenado en una BD<sup>20</sup> de corpus textuales.

#### 3.2.1 Anotaciones operadas automáticamente con la ayuda del analizador *Cordial*

Para el primer tipo de anotaciones, se trata de asociar a cada unidad textual las informaciones morfosintácticas correspondientes (la categoría gramatical, el lema, el género, el número, el tiempo, etc.). Algunas verificaciones semiautomáticas y manuales nos permitieron corregir algunos de los pocos errores que *Cordial* produce en salida, y adaptar las informaciones a nuestros

<sup>18</sup> *eXtensible Markup Language* (“lenguaje de balizas extensible”, en español).

<sup>19</sup> *Document Type Definition* (“Definición del Tipo de Documento”, en español).

<sup>20</sup> BD: Base de datos.



objetivos. Este primer tipo de anotaciones nos posibilitaron luego el poder focalizar de mejor manera las anotaciones manuales. En la imagen siguiente (figura 2, Molina Mejía, 2014: 25), por ejemplo, mostramos algunas de las etiquetas que Cordial produce en salida. Si tomamos la oración que aparece destacada en la imagen:

C'est un phénomène unique dans l'histoire française, né d'une conjonction particulièrement favorable de la démographie et de l'économie plus rare que l'on sait déjà qu'il ne pourra pas durer.

N°	MOT	LEMME	P	Prop.	Fonction	Groupe	S+G	Type	Type détaillé	Sémantique
1	C'	ce	1	independante	Sujet	Groupe pronominal	1/1	POD	PRON Dén. Sing	
2	est	être	1	independante	Verbe	Groupe nominal	2/2	VINGP3S	Indef. PRESENT	existence/in-évement/évit
3	un	un	1	independante	Attribut du sujet	Groupe nominal	4/4	DETM3S	ART Ind. Masc. Sing	
4	phénomène	phénomène	1	independante	Attribut du sujet	Groupe nominal	4/4	NOMS	NOM Masc. Sing	polyémique : fait observable/événement/événement
5	unique	unique	1	independante	Attribut du sujet	Groupe nominal	4/4	ADJ3S	ADJ Sing Ind. Genre	polyémique : original/spécial/original/spécial
6	dans	dans	1	independante	Attribut du sujet	Groupe nominal prépositionnel	4/8	PREP	PREPOSITION	
7	l'	le	1	independante	Attribut du sujet	Groupe nominal prépositionnel	4/8	DET3FS	ART Dnf. Fem. Sing	
8	histoire	histoire	1	independante	Attribut du sujet	Groupe nominal prépositionnel	4/8	NCF3S	NOM Fem. Sing	polyémique : étude historique/potential/évolution/continuum
9	française	français	1	independante	Attribut du sujet	Groupe nominal prépositionnel	4/8	ADJ3S	ADJ Fem. Sing	citoyen/européen
10			1	independante	punctuation/able					
11	né	naître	1	independante	Apposition	Groupe adjectival	11/11	VIP3RPM3S	Participle PASSE M.	polyémique : manifestation/origine/acte/début/begin
12	d'	de	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/14	PREP	PREPOSITION	
13	une	un	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/14	DETF3S	ART Ind. Fem. Sing	
14	conjonction	conjonction	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/14	NCF3S	NOM Fem. Sing	polyémique : réunion/union/connection
15	particulièrement	particulièrement	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/14	ADV	ADVERBE	démarquement
16	favorable	favorable	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/14	ADJ3S	ADJ Sing Ind. Genre	polyémique : prospecte/avantageux/prospicieux
17	de	de	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/19	PREP	PREPOSITION	
18	la	le	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/19	DET3FS	ART Dnf. Fem. Sing	
19	démographie	démographie	1	independante	Complément d'objet indirect	Groupe nominal prépositionnel	14/19	NCF3S	NOM Fem. Sing	théories/généralques/démographie; domaine = démographie
20	et	et	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	COJ	CONJ. Coordin.	
21	de	de	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	PREP	PREPOSITION	
22	r'	le	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	DET3FS	ART Dnf. Fem. Sing	
23	économie	économie	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	NCF3S	NOM Fem. Sing	polyémique : (s)carpe/substrat/Prat/Leaving
24	plus	plus	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	ADV	ADVERBE	différence/quantité/discrète/interlocution
25	rare	rare	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	ADJ3S	ADJ Sing Ind. Genre	petite fréquence/harélé
26	que	que	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	ADV	ADVERBE	negative
27	l'	l'	2	coordonnée	Complément d'objet indirect	Groupe nominal prépositionnel	19/23	NCF3S	NOM Fem. Sing	lettre
28	on	on	2	coordonnée	Sujet	Groupe pronominal	20/28	PPER3S	PRON Pers. 3e S	polyémique : avoir conscience/realize/be aware/know
29	sait	savoir	2	coordonnée	Verbe	Groupe nominal	29/29	VINGP3S	Indef. PRESENT	date antérieure
30	déjà	déjà	2	coordonnée	Complément circonstanciel de...	Groupe nominal	30/30	ADV	ADVERBE	
31	qu'	que	3	subordonnée	Sujet	Groupe pronominal	31/31	SUB	CONJ. Subord.	
32	il	il	3	subordonnée	Sujet	Groupe pronominal	32/32	PPER3S	PRON Pers. 3e S	
33	ne	ne	3	subordonnée	Verbe avec négation	Groupe nominal	34/34	ADV	ADVERBE	
34	peut	peuvent	3	subordonnée	Verbe avec négation	Groupe nominal	34/34	VINGF3S	Indef. FUTUR 3e	polyémique : avoir la possibilité/can/could
35	pas	pas	3	subordonnée	Verbe avec négation	Groupe nominal	34/34	ADV	ADVERBE	
36	dure	dure	4	infinive	Verbe	Groupe nominal	36/36	VINF	INFINITIF	polyémique : occuper un espace temps/être en contact/possibilité

Figura 2. Captura de pantalla con las etiquetas del analizador Cordial.

Podemos observar que cada elemento dispone de una etiqueta particular según el contexto en el cual se encuentra el elemento. Observamos, por otra parte, que esta oración compuesta (o compleja), dispone en la quinta columna de un

etiquetado que señala las proposiciones<sup>21</sup> que se encuentran en la oración. Además de esta noción de proposición, tenemos de igual forma nociones tales como grupos nominales, pronominales y preposicionales; lo que nos permite focalizar de mejor manera los correferentes textuales, por ejemplo.

Como lo podemos apreciar en dicha figura, a parte de las informaciones morfológicas y sintácticas, *Cordial* propone, además, las informaciones textuales que acabamos de analizar (división en oraciones y en proposiciones), e informaciones semánticas para cada elemento. A partir de estas informaciones, las anotaciones manuales pueden realizarse de una manera mucho más fácil.

### 3.2.2 Anotaciones añadidas manualmente y que permiten el balizado de las características concernientes a la lingüística textual

A las anotaciones presentadas en la parte anterior, procedimos a agregar otras anotaciones y, para ello, utilizamos el formato de balizas extensibles de XML. La idea es poder anotar todos los fenómenos que pertenecen al campo de la lingüística textual, y que no pueden ser anotados de forma automática: los correferentes textuales, los conectores y marcadores lógico-temporales y discursivos, etc.

La particularidad del lenguaje XML es justamente el permitir la creación de sus propias entidades y anotarlas, así como lo podemos apreciar en la imagen siguiente (figura 3, Molina Mejía, 2014: 25).

Para los elementos, decidimos conservar las etiquetas propuestas por *Cordial*, por ejemplo “DETDMS” (*determinantes, masculinos, singulares*), y los “NCMS” (*nombres comunes, masculinos, singulares*), etc. Esto nos permitió la subsiguiente anotación de los “GN” (*grupos nominales*) durante el análisis de los correferentes de tipo nominal. Por otra parte, la anotación de los elementos de tipo “PPER” (*pronombre personal*) nos permitió el poder anotar de manera más fácil algunos correferentes anafóricos de tipo “*pronominal*”.

A las anotaciones del analizador *Cordial* agregamos, entonces, las anotaciones de la correferencia textual, tales como: “*Referente*”, “*Correferente*”, “*Anáforas*”, “*Catáforas*”, los “*id*” (identificadores unidos a cada referente o correferente, o que identifican cada una de las cadenas de correferencia textual, etc.).

En efecto, todas estas anotaciones nos permiten aislar, como en el caso anterior, los referentes principales, los correferentes anafóricos o catafóricos, los diferentes tipos de anáforas (fieles, infieles, asociativas, etc.). Ellas nos

<sup>21</sup> Utilizamos el término de “proposición” en francés “*proposition*” tal y como está descrito en la “*Grammaire méthodique du français*” (Riegel *et al.*, 2009).

permiten, de igual modo, el poder identificar y señalar las cadenas de coreferencia (una o varias según el texto anotado).

```

- <phrase type="complexe">
- <prop type="subordonnee" sous-type="principale">
- <coreferent type="Nom" sous-type="NC" classe="fidele" idn="c9" chaine="r2" position="Anaphore" reference="c6">
  <element type="NCMP" cat="NC" genre="M" nombre="P">Préjugés</element>
</coreferent>
- <coreferent type="Pron" sous-type="Rel" classe="totale" idn="c10" chaine="r2" position="Anaphore" reference="c9">
  <element type="PRI">que</element>
</coreferent>
- <coreferent type="Nom" sous-type="GN" classe="infidele" idn="c11" chaine="r3" position="Anaphore" reference="c4">
  <element type="DETFDS" cat="DETD" genre="F" nombre="S">la</element>
  <element type="ADJFS">nouvelle</element>
  <element type="NCFS" cat="NC" genre="F" nombre="S">semaine</element>
  <element type="PREP">pour</element>
  <element type="DETDMS" cat="DETD" genre="M" nombre="S">|</element>
  <element type="NCMS" cat="NC" genre="M" nombre="S">emploi</element>
</coreferent>
  <element type="PREP">en</element>
- <coreferent type="Nom" sous-type="GN" classe="associative" idn="c12" chaine="r1" position="Anaphore" reference="c8">
  <element type="DETPOSS">leur</element>
  <element type="NCFS" cat="NC" genre="F" nombre="S">faveur</element>
</coreferent>
  <element type="PCTFAIB">,</element>
- <prop type="subordonnee" sous-type="relative">
- <coreferent type="Pron" sous-type="Rel" classe="totale" idn="c13" chaine="r3" position="Anaphore" reference="c11">
  <element type="PRI">qui</element>
</coreferent>
  <element type="VINDP3S" cat="V" mode="IND" temps="P" pers="3" nombre="S">vient</element>
  <element type="PREP">de</element>
  <element type="VINP" cat="V" mode="INF">débuter</element>
  <element type="PCTFAIB">,</element>
</prop>
  <element type="VINDP3S" cat="V" mode="IND" temps="P" pers="3" nombre="S">entend</element>
  <element type="VINP" cat="V" mode="INF">combattre</element>
</prop>
<element type="PCTFORTE">.</element>
</phrase>

```

**Figura 3.** Anotación manual en formato XML del corpus textual (fragmento).

### 3.2.3 DTD del corpus

La DTD es el documento que normaliza todas las anotaciones de un corpus. Esta “define y costriñe las balizas y los atributos que aquellos [los documentos XML] deben utilizar para ser válidos” (Souque, 2014: 85). Para Habert *et al.* (1998: 64), la DTD precisa el número de veces y el orden en el que los elementos balizados pueden aparecer en el documento normalizado.

En nuestro caso, la anotación global de los textos es descrita y normalizada por la DTD correspondiente. En efecto, la DTD específica que hemos constituido, nos permitió la verificación de la coherencia de nuestras anotaciones, gracias a “una gramática de base que contiene todos los elementos etiquetados de manera automática y que luego fueron balizados manualmente” (Molina Mejía, 2015: 288). Este documento permite establecer cuántas veces un elemento debe aparecer, o no, en el texto, el orden de los elementos y la manera en la que aquellos han sido constituidos.

La DTD nos permitió, igualmente, la realización informática de las actividades pedagógicas. Es decir, la DTD fue utilizada con el fin de permitir la elaboración del sistema informático a partir de su contenido. Es decir, que

gracias a esta los programadores pudieron encontrar los elementos que fueron empleados en el proceso de constitución de las actividades, y a partir de ellos extraer las informaciones referentes a la lengua que procesan las actividades.

Por ejemplo, la imagen siguiente (figura 4, Molina Mejía, 2015) nos permite apreciar los distintos elementos que fueron anotados y, de una manera global, nos permite percibir la manera en la que funciona el sistema a partir del corpus anotado. Para ello, la DTD contiene tres tipos de informaciones importantes que permiten su explotación informática: los elementos (*ELEMENT*) que fueron anotados y descritos en el documento XML (cf. figura 3), las listas de atributos de cada elemento (*ATTLIST*), y los valores de dichos atributos que se encuentran insertados al interior de las “*ATTLIST*”.

Ejemplo: Para el elemento “*paragraphe*” (párrafo), podemos ver que este se encuentra constituido por “*phrases*” (oraciones) o por “*éléments*” (elementos). En cuanto a las “*phrases*” (oraciones), estas pueden tener el atributo “*type*” (tipo), y los valores de tipo de oración son “*simple*” o “*complexe*” (compleja).

### 3.3 ¿Para cuáles actividades?

Una vez que los textos han sido anotados, y que la DTD ha sido establecida, nosotros podemos comenzar a realizar las secuencias pedagógicas de formación. Para ello, pasamos a dividir las en series de actividades o de ejercicios. Así, las actividades elaboradas en este caso toman en cuenta los fenómenos ligados a la lingüística textual. Estas tienen que ver con:

- *La estructura lógica del texto*: El análisis de las oraciones principales en los textos; el estudio de la estructuración textual; el análisis de los diferentes tipos de texto y/o de secuencias textuales.
- *La coherencia y la cohesión textuales*: El estudio de la correferencia textual (anáforas y catáforas, referentes y correferentes, cadenas de correferencia, etc.); el análisis de los conectores y de los marcadores lógico-temporales y discursivos en los textos.
- *La progresión temática*: El trabajo alrededor de las nociones de tema y de rema, y los diferentes tipos de progresión temática (lineal, con tema constante o derivada).

```

1 <!ELEMENT DOCUMENTTEXTE (ENTETE, CONTENU)>
2 <!--ATTLIST DOCUMENTTEXTE id_doc ID #REQUIRED-->
3 <!ELEMENT ENTETE (titre_doc, auteur, source, date, type, niveau)>
4 <!--ELEMENT titre_doc (#PCDATA)-->
5 <!--ELEMENT auteur (#PCDATA)-->
6 <!--ELEMENT source (#PCDATA)-->
7 <!--ATTLIST source type (magazine | journal | livre | internet | roman | Inconnu) #REQUIRED-->
8 <!--ATTLIST source periodicit (hebdomadaire | journaliere | mensuelle | bimensuelle | annuelle | Inconnue) #IMPLIED-->
9 <!--ELEMENT date (#PCDATA)-->
10 <!--ELEMENT type (#PCDATA)-->
11 <!--ELEMENT niveau (#PCDATA)-->
12 <!--ELEMENT activite (#PCDATA)-->
13 <!--ELEMENT CONTENU (titre?, (sous-titre?, paragraphe)+)-->
14 <!--ELEMENT titre (phrase+)-->
15 <!--ELEMENT sous-titre (phrase+)-->
16 <!--ELEMENT paragraphe (phrase+ | element)+-->
17 <!--ELEMENT phrase (#PCDATA | prop | element)*-->
18 <!--ATTLIST phrase type (simple | complexe) #REQUIRED-->
19 <!--ELEMENT ellipse EMPTY-->
20 <!--ELEMENT prop (#PCDATA | referent | coreferent | element | prop)*-->
21 <!--ELEMENT referent (#PCDATA | element)*-->
22 <!--ELEMENT coreferent (#PCDATA | element)*-->
23 <!--ELEMENT element (#PCDATA)*-->
24 <!--ATTLIST prop type (principale | Independante | subordonnee | averbale | incise) #REQUIRED-->
25 <!--ATTLIST prop sous-type (principale | participiale | complementive | coordonnee | relative | juxtaposee | verbale |
  Infrinitive | conjonctive | averbale) #IMPLIED-->
26 | DETDMS | NCFP | ADJFP | VINDR3P | ADJHP | PCTFORTE | DETDMS | ADJ3RD | DETPOSS | PPER3P | PPER35 | ADJFS | NCSIG | PCTFAIB | ADJNRH |
  NCMIN | PRI | NPFS | COD | ADJMS | NPI | NCHMS | SUB | NCFIN | ADJIND | VCONP35 | VPARPMS | NCPG | NPMS | DETDEH | PIPG | VPARPMP |
  VPARPFS | VPARPFS | PIMS | VINDF35 | ADJPG | NPFIN | NPSIG | NPMSIG | VINDI3P | NPMIN | VINDI35 | PIMP | VPARPFP | PPER2P | VINDP2P |
  PIFP | NCI | PDP | ADJINT | PPERIP | VINDF2P | ADJINV | PII | VIMPP2P | VSUBP35 | VINDF3P | VINDP35 | INT | PISIG | VINDP35P | PPER25 |
  PPERIS | VINDIIP | VINDP25 | VINDP15 | ADJMIN | VIMPP1P | PRFS | PRMS | VIMPP25 | VCONP3P | PIFS | VINDP1P | ADJHFS | ADJHMS | VINDI25 |
  ELLIP) #REQUIRED-->
27 <!--ATTLIST elenen cat (NC | V | DETD | ADJ | PPER | DETI | NP | PD | PI | PR) #IMPLIED-->
28 <!--ATTLIST elenen node (INF | IND | CON | PAR | IMP | SUB) #IMPLIED-->
29 <!--ATTLIST element temps (P | F | PRES | I | PS) #IMPLIED-->
30 <!--ATTLIST element pers (1 | 2 | 3) #IMPLIED-->
31 <!--ATTLIST element genre (M | F | I) #IMPLIED-->
32 <!--ATTLIST elenen nombre (S | P | I) #IMPLIED-->
33 <!--ATTLIST referent type (Nom) #IMPLIED-->
34 <!--ATTLIST referent sous-type (GN | NP | NC) #IMPLIED-->
35 <!--ATTLIST referent idn ID #REQUIRED-->
36 <!--ATTLIST coreferent type (Nom | Pron) #IMPLIED-->
37 <!--ATTLIST coreferent sous-type (GN | COD | Pers | Rel | N | Compl | Pronom | COI | NP | NC) #IMPLIED-->
38 <!--ATTLIST coreferent classe (fidele | Infidele | determinative | totale | partielle | integrante | associative)
  #IMPLIED-->
39 <!--ATTLIST coreferent idn ID #REQUIRED-->
40 <!--ATTLIST coreferent chaine IDREF #REQUIRED-->
41 <!--ATTLIST coreferent position (Anaphore | Cataphore) #IMPLIED-->
42 <!--ATTLIST coreferent reference IDREF #REQUIRED-->
43
44 |

```

Figura 4. DTD con sus diferentes elementos, atributos y valores.

Para un determinado nivel, los textos asociados a las actividades son intercambiables; estos hacen parte del parametrizado de las actividades. Las secuencias de las actividades (figura 5) se hallan constituidas por cuatro fases (ejercicios iniciales de localización, explicaciones a través de soportes teóricos, ejercicios de sistematización y ejercicios de aplicación de conocimientos), las cuales el docente puede también parametrizar (cf. figura 6 § figura 7).

Un aspecto interesante del sistema es el hecho de proponer el almacenado del soporte teórico (es decir, de un corpus teórico, fase 2). Por el momento los docentes pueden almacenar en la “*BD de corpus du support théorique*” documentos en formato PDF o tratamiento de texto (*Word, LibreOffice Writer*, etc.), pero la idea es la de proponer una interfaz en HTML o en Javascript que permita la explicación de los fenómenos que tienen relación con la lingüística textual.

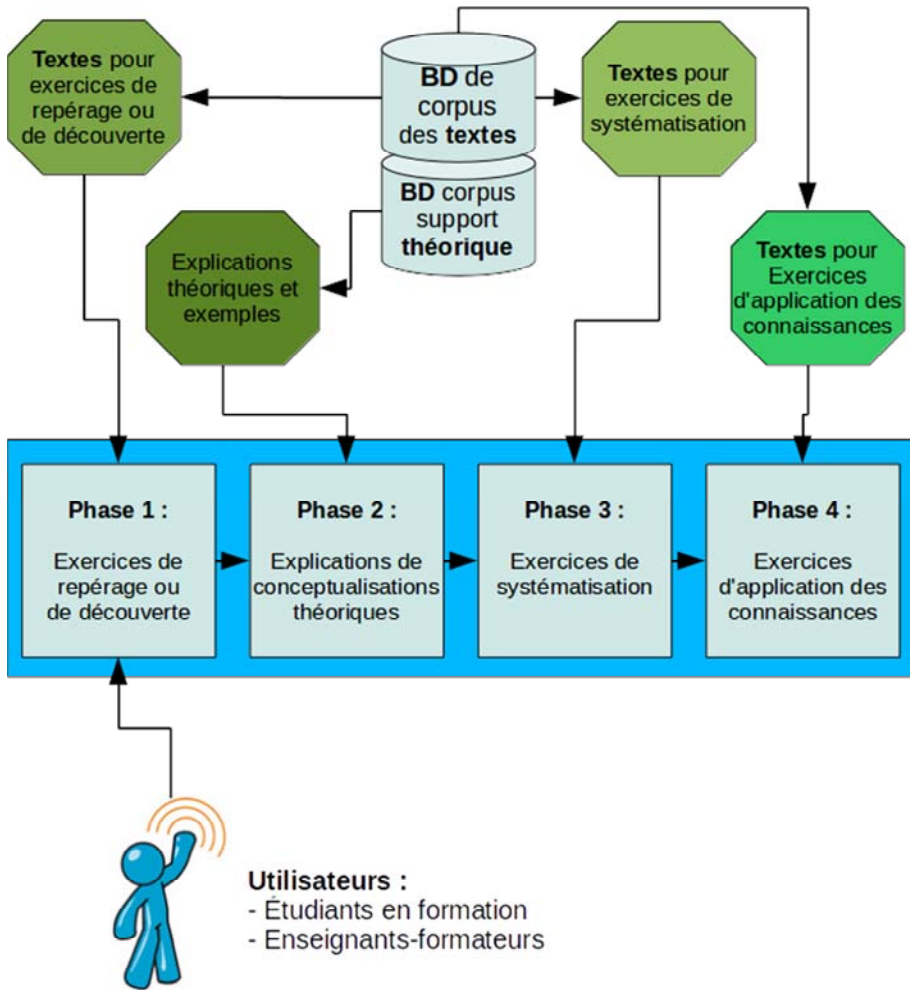


Figura 5. Diferentes fases de una secuencia didáctica (Molina Mejía, 2015).

#### 4. Un sistema informático fundamentado en corpus

El sistema de ALAO (Aprendizaje de Lenguas Asistido por Ordenador), ELiTe-[FLE]<sup>2</sup>, que nosotros proponemos se halla fundamentado en un trabajo realizado sobre corpus. Este funciona a partir de corpus etiquetados y anotados, gracias a las técnicas extraídas del TAL y a los tratamientos efectuados de forma manual, de la manera en que lo presentamos en la sección anterior.

##### 4.1 ELiTe-[FLE]<sup>2</sup> y la formación de formadores

La idea de poseer un sistema de ALAO como lo es ELiTe-[FLE]<sup>2</sup>, es el de ofrecer a los docentes-formadores y a sus estudiantes en formación, un ambiente informático que puede, a la vez, ser utilizado en un curso presencial, así como en autonomía guiada. Dicho sistema debería servir dentro del proceso de formación lingüística de los formadores de lengua francesa (Molina Mejía, 2015).

ELiTe-[FLE]<sup>2</sup> funciona a partir de las actividades que se encuentran fundamentadas en el corpus lingüístico que fue presentado en la parte anterior. Algunas actividades permiten ayudar a comprender nociones tales como la coherencia y la cohesión textuales y discursivas, las nociones de anáforas, de catáforas, la progresión temática, etc. En este sentido, consideramos que la correcta adquisición de dichas nociones pasa por el aprendizaje de otras nociones tales como: GN, COD, COI, etc. Dicho de otra forma, de las nociones que los futuros formadores deberán luego enseñar a sus aprendientes de lengua extranjera.

El sistema propuesto permite así, el poder preparar unas secuencias didácticas de formación y, para ello, dicho sistema utiliza los textos que han sido almacenados en la base de datos del corpus textual. La siguiente imagen (figura 6, Molina Mejía, 2015) nos da una visión general del sistema que permite la creación de secuencias didácticas por parte de los docentes-formadores.

En dicha figura podemos observar las distintas posibilidades que ofrece la plataforma.

Ya sea en el menú que se encuentra en la parte superior, ya sea en el menú que encontramos al frente de cada secuencia creada, este permite a los docentes el realizar varias acciones: un docente puede crear una secuencia didáctica, utilizando para ello las opciones que le ofrece el sistema (*cf.* figura 7). Él puede ensayar la secuencia que él mismo ha creado, es decir, desarrollarla con el fin de observar cómo dicha secuencia funciona antes de utilizarla con sus estudiantes. El docente puede, también, modificar o suprimir una secuencia ya existente,

siempre y cuando esta haya sido creada por él. Finalmente, el docente puede enviar la secuencia que él haya seleccionado a los correos electrónicos de sus estudiantes en formación.

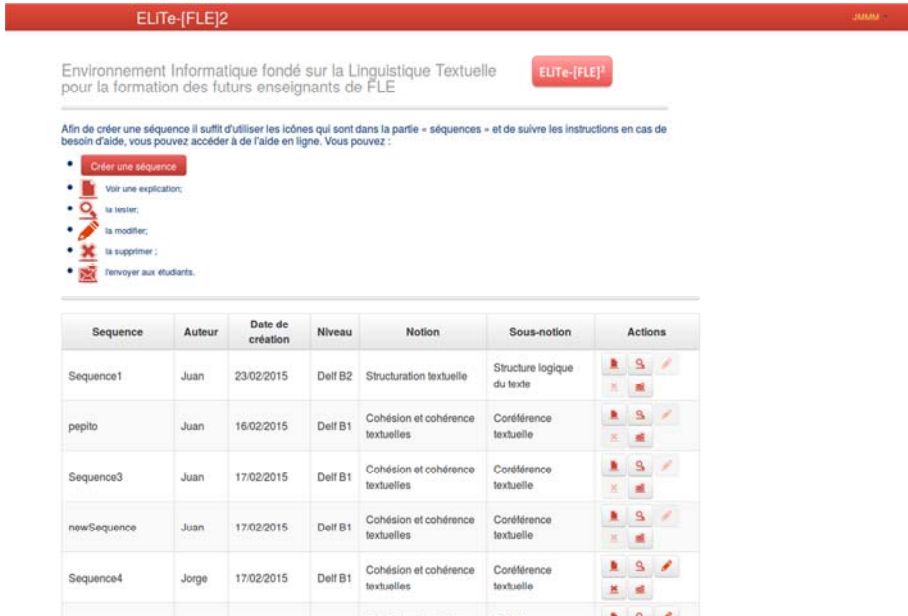


Figura 6. Módulo de gestión de secuencias de formación ELiTe-[FLE]<sup>2</sup>.

Los docentes pueden, así, crear sus propias secuencias, utilizar las secuencias creadas por sus colegas o crear nuevas secuencias al modificar las secuencias ya existentes. Con el fin de realizar todas estas acciones, ellos poseen un módulo de gestión de secuencias (figura 7, Molina Mejía, 2015).

Este módulo permite a los docentes seleccionar una de las tres nociones a ser estudiadas y una sub-noción (cf. 2.3). Ellos pueden además escoger el nivel con relación a la actividad (B1, B2 o C1), y según el nivel seleccionado hay varios textos que son propuestos por la plataforma. Los textos cambian según el tipo de actividad (ejercicios iniciales de localización y ejercicios de sistematización). El docente puede, por otro lado, decidir si quiere emplear, o no, el soporte pedagógico. Puede además escribir las instrucciones para la actividad de aplicación de conocimientos (la cual aparecerá al final, en cuanto las demás actividades hayan sido resueltas). Finalmente, el docente podrá



también describir la secuencia con la finalidad de compartir esta información con sus colegas o para recordar él mismo las actividades que él ya ha creado.

The screenshot shows the 'Composez votre séquence didactique' interface. At the top, there is a red header with 'ELiTe-[FLE]²' on the left and 'ADDA' on the right. Below the header, the form is organized into several sections:

- Nom \***: A text input field with a placeholder 'Écrivez un nom à la séquence'.
- Activité de repérage**: A section with several dropdown menus:
  - Notion**: 'Structuration textuelle'
  - Sous-notion**: 'Phrases principales'
  - Niveau**: 'Def B1'
  - Texte de repérage \***: 'Aider les handicapés dans le mo...' (with a small icon)
  - Activité de repérage**: 'Activité 1'
- Support linguistique et didactique**: A section with a question 'Souhaitez-vous utiliser des supports linguistiques et/ou didactiques ?' and two radio buttons: 'Oui' (selected) and 'Non'.
- Activité de systématisation**: A section with a text input field for 'Texte de systématisation \*' containing 'Les Thibault, tome 1, le cahier gr...' (with a small icon) and a dropdown menu for 'Activité de systématisation' set to 'Activité 2'.
- Activité d'application de connaissances**: A large text input field with the instruction 'Écrivez dans cette zone vos consignes de travail. (2500 caractères)'.
- Explication de la séquence**: Another large text input field with the instruction 'Écrivez dans cette zone une explication de la séquence. (2500 caractères)'.

At the bottom of the form, there are two red buttons: 'Finir' and 'Suivant'.

**Figura 7.** Módulo de creación de secuencias de formación del sistema.

## 4.2 ELiTe-[FLE]² y la formación de formadores

En un primer momento decidimos probar el trabajo desarrollado a nivel del corpus para las actividades al rededor de la correferencia textual. Mostramos una actividad desarrollada a partir del sistema ELiTe-[FLE]², tal como la describimos en nuestro trabajo (cf. Molina Mejía, 2015 § capítulo 11) y en un artículo recientemente publicado (Molina Mejía & Antoniadis, 2015).

En la sección siguiente, mostramos solamente las actividades que utilizan el corpus, es decir, las actividades iniciales de localización (o actividades de descubrimiento) y las actividades de sistematización.

#### 4.2.1 Ejercicios iniciales de localización o de descubrimiento

El corpus anotado permite el desarrollo de actividades en las cuales los estudiantes en formación pueden localizar en un texto los elementos generales a partir de una noción específica a ser estudiada. El siguiente ejemplo (figura 8, Molina Mejía & Antoniadis, 2015) nos muestra la manera en la cual los estudiantes, siguiendo un procedimiento inductivo, van a descubrir la noción de correferencia textual con la guía del docente-formador. En esta primera actividad se trata de deslizar desde el texto que se encuentra en la parte superior hacia la segunda columna situada en zona inferior, las unidades textuales (en este caso específico se trata de anáforas), y hacia la tercera columna el referente principal que completa la serie. La idea es de descubrir los referentes y los correferentes textuales que hacen parte de la noción de correferencia textual.

Gracias a las anotaciones provistas por el corpus, una retro-alimentación aparece en la validación de cada actividad. En este caso, tenemos una retro-alimentación que nos permite dar a los estudiantes en formación algunos elementos de orden morfo-sintáctico, y que les permiten de adquirir algunas de las nociones ligadas a la formación lingüística (figura 9, Molina Mejía & Antoniadis, 2015). En la imagen podemos ver a la izquierda una retro-alimentación en el que las nociones que aparecen son GN, mientras que lo que se esperaba era otro tipo de unidad textual (Pronombre, por ejemplo). La idea es de ayudar a los futuros docentes a que realicen un proceso de reflexión, y esto siempre bajo una perspectiva de enfoque inductivo, empleando para ello algunas de las nociones ligadas a la lingüística textual, tales como: “correferentes de tipo nominal” o “correferentes de tipo pronominal”, y así sucesivamente.

En la imagen de la derecha podemos observar el caso de una retro-alimentación de tipo “positivo”, por el momento esta no da informaciones de orden meta-lingüístico, pero la idea es de mejorar en el futuro este aspecto, y dar así a los aprendientes el tipo de correferente encontrado (GN o Pron).

En una segunda actividad, los estudiantes en formación serán conducidos a descubrir las nociones ligadas a la lengua objetiva. Es así como en el caso que vemos en la siguiente imagen (figura 10, Molina Mejía & Antoniadis, 2015), los estudiantes trabajan acerca de la noción de anáforas nominales (GN, N), y pronominales (Pron). La actividad consiste en desplazar los elementos que se encuentran en la parte superior, llevándolos hacia la sección inferior; la idea es de hacerlos separar los elementos en dos partes, en la parte superior los estudiantes van a dejar solamente los GN y los N, y en la parte inferior quedarán todos los Pron.

Una vez que todos los elementos han sido desplazados, obtenemos una retro-alimentación en la cual puede observarse una zona gris, esto hasta que el

estudiante encuentre todos los elementos similares, GN y N en la parte superior y Pron en la parte inferior (figura 11, Molina Mejía & Antoniadis, 2015). La idea es de hacer descubrir a los futuros docentes de francés que existen algunos correferentes de orden nominal, mientras que hay otros de tipo pronominal. En otras actividades será posible trabajar otras nociones, como es el caso de los tipos de anáforas (fieles, infieles, asociativas, etc. Al final, todas las zonas deben quedar en color blanco para que la actividad sea validada.

**Activité de Repérage**

En suivant la même démarche, trouvez maintenant la suite des éléments pour le groupe écrit en vert et trouvez ensuite le groupe qui convient à la liste d'éléments écrits en bleu.

**Titre :** Aider les handicapés dans le monde du travail  
**Auteur :** Comité « tous ensemble au travail »  
**Source :** Inconnue  
**Date :** novembre 2002  
**Type :** Argumentatif  
**Niveau :** B1 -CEFR

**Aider les handicapés dans le monde du travail**

Les personnes handicapées restent victimes de **forts préjugés** dans le monde du travail. Le système Semaire pour l'emploi en **leur faveur** entend les combattre. **Cela** est marqué par une reprise de la polémique sur les Centres d'Aide par le Travail.

Aujourd'hui, **26 % des personnes handicapées** sont au chômage. En cause, selon beaucoup, les préjugés dont **elles** sont victimes. Privilèges que **le nouveau système pour favoriser** en leur faveur, au vu de **leur handicap**, entend combattre. L'association pour la gestion du fonds pour l'insertion professionnelle des personnes handicapées (Apef) et la Ligue pour l'Adaptation du handicap physique au travail (Adapt) veulent mobiliser autour du slogan Agir, c'est réussir.

Le handicap, lorsqu'il est visible est encore trop souvent associé à l'ineptitude. La mobilisation est plus que jamais nécessaire pour modifier le regard sur le handicap, équiper les deux associations. Aujourd'hui, **repellent ethics**, **215 000 personnes handicapées** sont à la recherche d'un emploi et restent, en moyenne, deux fois plus longtemps sans activité.

Pourtant, 87 % des entreprises qui emploient **des travailleurs handicapés** s'en disent satisfaites et 62 % des entreprises qui **ne** emploient pas estiment qu'une telle expérience pourrait se dérouler de manière satisfaisante, selon un sondage réalisé en vue de **la semaine d'emploi**. On observe ainsi que le passage à l'acte est déterminant pour les chefs d'entreprises, puisqu'ils sont majoritairement satisfaits lorsqu'ils ont effectué une embauche, à commencer le directeur général de l'Adapt, Philippe Vietu. D'autres des dirigeants de chefs d'entreprise recrutés dans le Guide France Info **Le salarié handicapé** dans l'entreprise, **ce dernier** est un salarié comme les autres, avec en plus, la volonté de s'en sortir. Souvent plus productif que les autres, il crée un effet fédérateur dans une équipe de travail où les petits problèmes courants sont résolus.

Le Semaire s'est ouvert lundi sur le parvis du Trocadéro à Paris, par un événement symbolique, en présence des adhérents pour les personnes handicapées au maire de Paris. Les Franciliens sont invités par les organisateurs à marquer leur soutien à l'intégration **des personnes handicapées** dans le monde du travail, en apposant l'empreinte colorée de leurs mains sur des livres géants. Tout au long de **la semaine**, 16 régions se mobilisent et organisent près de 80 événements. Forums, tables rondes, pièces de théâtre, matchs de tennis (tennis adapté) entre une équipe de **athlètes visuels** et des chefs d'entreprise, tout sera bon pour lever les freins psychologiques et culturels. (...)

**les handicapés**

Les personnes handicapées  
leur faveur

26 % des personnes handicapées  
elles

leur faveur

215 000 personnes handicapées  
des travailleurs handicapés

en

le

les personnes handicapées  
des personnes handicapées

différents outils

**forts préjugés**

elles

le nouveau système pour favoriser  
leur

le nouveau système pour favoriser  
le

le nouveau système pour favoriser  
le

le nouveau système pour favoriser  
le

**?**

elles

le nouveau système pour favoriser  
leur

le nouveau système pour favoriser  
le

le nouveau système pour favoriser  
le

le nouveau système pour favoriser  
le

estimer qu'une telle expérience pourrait se dérouler de manière satisfaisante, selon un sondage réalisé en vue de la semaine d'emploi. On observe ainsi que le passage à l'acte est déterminant pour les chefs d'entreprises, puisqu'ils sont majoritairement satisfaits lorsqu'ils ont effectué une embauche, à commencer le directeur général de l'Adapt, Philippe Vietu. D'autres des dirigeants de chefs d'entreprise recrutés dans le Guide France Info Le salarié handicapé dans l'entreprise, ce dernier est un salarié comme les autres, avec en plus, la volonté de s'en sortir. Souvent plus productif que les autres, il crée un effet fédérateur dans une équipe de travail où les petits problèmes courants sont résolus.

Le Semaire s'est ouvert lundi sur le parvis du Trocadéro à Paris, par un événement symbolique, en présence des adhérents pour les personnes handicapées au maire de Paris. Les Franciliens sont invités par les organisateurs à marquer leur soutien à l'intégration des personnes handicapées dans le monde du travail, en apposant l'empreinte colorée de leurs mains sur des livres géants. Tout au long de la semaine, 16 régions se mobilisent et organisent près de 80 événements. Forums, tables rondes, pièces de théâtre, matchs de tennis (tennis adapté) entre une équipe de athlètes visuels et des chefs d'entreprise, tout sera bon pour lever les freins psychologiques et culturels. (...)

**Attention vous vous êtes trompé.**

Le groupe nominal que vous avez choisi n'est pas le bon choix : un attribut peut-être un autre type d'unité syntaxique. Essayez encore!

**Excellent travail, Mikaltons !**

**Continuer**

**les handicapés**

Les personnes handicapées  
leur faveur

26 % des personnes handicapées  
elles

leur faveur

215 000 personnes handicapées  
des travailleurs handicapés

en

**les handicapés**

Les personnes handicapées  
leur faveur

26 % des personnes handicapées  
elles

leur faveur

215 000 personnes handicapées  
des travailleurs handicapés

en

Figura 9. Retro-alimentación correspondiente al ejercicio inicial de localización 1 (extraído del texto).

### 4.2.2 Ejercicios de sistematización

Este tipo de ejercicios debería permitir a los estudiantes de dominar de una mejor manera las nociones aprendidas gracias a la realización de las actividades, que les permiten la puesta en marcha de los conocimientos adquiridos. Dicho de otra forma, una vez que los estudiantes hayan hecho los ejercicios iniciales de localización, y hayan además obtenido las explicaciones de parte de sus docentes-formadores, entonces en dicho momento podrán ir más allá en el proceso de dominio del tema.

ELITE-[FLE]2
AMM

---

#### Activité de Repérage

En suivant la même démarche, trouvez maintenant la suite des éléments pour le groupe écrit en vert et trouvez ensuite le groupe qui convient à la liste d'éléments écrits en bleu.

---

Titre : Aider les handicapés dans le monde du travail  
 Auteur : Comité « tous ensemble au travail »  
 Source : inconnue  
 Date : novembre 2002  
 Type : Argumentatif  
 Niveau : B1-CECR

#### Aider les handicapés dans le monde du travail

Les personnes handicapées restent victimes de **forts préjugés** dans le monde du travail. La sixième Semaine pour l'emploi en **tour France** entend les combattre. Elle est marquée par une reprise de la polémique sur les Centres d'Aide par le Travail.

Aujourd'hui, **25 % des personnes handicapées** sont au chômage. En cause, selon beaucoup, les **préjugés** dont elles sont victimes. **Préjugés** que la **seizième semaine** peut tenter en **tour France**, qui vient de débuter, entend combattre. L'Association pour la gestion du fonds pour l'insertion professionnelle des personnes handicapées (Aofaph) et la Ligue pour l'Adaptation du milieu physique au travail (Adapt) veulent mobiliser autour du slogan Agir, c'est réussir.

Le handicap, lorsqu'il est visible est encore trop souvent associé à l'incapacité. La mobilisation est plus que jamais nécessaire pour modifier le regard sur le handicap, expliquer les deux associations. Aujourd'hui, rappelez-elles, **210 000 personnes handicapées** sont à la recherche d'un emploi et restent, en moyenne, deux fois plus longtemps sans activité.

Pourtant, **87 % des entreprises** qui emploient **des travailleurs handicapés** s'en disent satisfaites et **62 % des entreprises** qui **ne** emploient pas estiment qu'une telle expérience pourrait se dérouler de manière satisfaisante, selon un sondage réalisé en vue de la **seizième semaine**. On observe ainsi que le passage à l'acte est déterminant pour les chefs d'entreprises, puisqu'ils sont majoritairement satisfaits lorsqu'ils ont effectué une embauche, à condition que le directeur général de l'Adapt, Philippe Velut. D'après des témoignages de chefs d'entreprise recueillis dans le Guide France Info **Le travailleur handicapé** dans l'entreprise, **ce dernier** est un salarié comme les autres, avec en plus, la volonté de s'en sortir. Souvent plus productif que les autres, il crée un effet fédérateur dans une équipe de travail où les petits problèmes courants sont rationalisés.

La **seizième** est ouverte lundi sur le parvis du Trocadero à Paris, par un événement symbolique, en présence des acquies pour les personnes handicapées au maire de Paris. Les Franciliens sont invités par les organisateurs à manifester leur soutien à l'intégration **des personnes handicapées** dans le monde du travail, en apposant l'empreinte colorée de leurs mains sur des livres géants. Tout au long de la **seizième**, 16 régions se mobilisent et organisent près de 10 événements. Forums, tables rondes, pièces de théâtre, match de football (football adapté) entre une équipe de **athlètes visuels** et des chefs d'entreprise, tout sera bon pour lever les freins psychologiques et culturels. (...)

les handicapés

Les personnes handicapées  
qui forment  
**25 % des personnes** candidates  
elles  
leur forment  
**210 000 personnes** handicapées  
**des travailleurs** handicapés  
et  
Le **salarié** handicapé  
ou **handicapé**  
**87 % personnes** handicapées  
**des entreprises** handicapées  
**difficiles visuels**

forts préjugés

?

?

Elle  
la **seizième semaine** pour l'emploi  
ou  
la **seizième** d'action  
La **seizième**  
le **salarié**

Figura 10. Actividad inicial de localización 2.

En la figura anterior (figura 12, Molina Mejía & Antoniadis, 2015), tenemos un ejemplo de este tipo de ejercicio. En esta actividad se trata de llevar a los estudiantes a realizar algunas inferencias, a partir de la noción de correferencia. Ellos deberán, primero que todo, seleccionar los colores en función del contexto en el que se se encuentra el elemento que hace falta. Para ello, los estudiantes disponen de un menú desplegado, en el que se encuentran varias opciones para cada elemento. Los colores de los correferentes deben ser seleccionados en función de los referentes principales.

Pourtant, 87 % des entreprises qui emploient **des travailleurs handicapés** s'en disent satisfaites et 62 % des entreprises qui n'en emploient pas estiment qu'une telle expérience pourrait se dérouler de manière satisfaisante, selon un sondage réalisé en vue de **la semaine d'action**. On observe ainsi que le passage à l'acte est déterminant pour les **chefs d'entreprises**, puisqu'ils sont majoritairement satisfaits lorsqu'ils ont effectué une embauche, a commenté le directeur général de l'Adapt, Philippe Velut. D'après des témoignages de **chefs d'entreprise** recueillis dans le Guide France Info **Le salarié handicapé** dans l'entreprise, **ce dernier** est un salarié comme les autres, avec en plus, la volonté de s'en sortir. Souvent plus productif que les autres, **il** crée un effet fédérateur dans une équipe de travail où les petits problèmes courants sont relativisés.

**La Semaine s'** est ouverte lundi sur le parvis du Trocadéro à Paris, par un événement symbolique, en présence des adjointes pour **les personnes handicapées** au maire de Paris. Les **Franciliens** sont invités par les organisateurs à manifester leur soutien à l'intégration **des personnes handicapées** dans le monde du travail, en apposant l'empreinte colorée de leurs mains sur des livres géants. Tout au long de **la semaine**, 16 régions se mobilisent et organisent près de 80 événements. Forums, tables rondes, pièces de théâtre, match de torball (football adapté) entre une équipe de **déficients visuels** et des **chefs d'entreprise**, tout sera bon pour lever les freins psychologiques et culturels. (...)



**Figura 11.** Retro-alimentación para la actividad inicial de localización 2.

Hay tres colores que corresponden a las tres cadenas de correferencia (naranja para los correferentes de la cadena 1, azul para los correferentes de la cadena 2, y verde para los correferentes de la cadena 3). La idea es que para cada elemento, siempre existen estas tres posibilidades de elección.

Una vez que los estudiantes han terminado de realizar el ejercicio de sistematización 1, ellos pasan a una nueva fase. En un segundo ejercicio de este tipo (figura 13, Molina Mejía & Antoniadis, 2015), ellos deberán seleccionar los elementos adecuados con relación a las cadenas de correferencia. Los estudiantes deberán inferir, una vez más, sobre cual elemento, según el contexto y el co-texto en el cual aquel se encuentra, es el más adecuado; y esto teniendo como referencia a los referentes principales. Para cada grupo de correferentes, les corresponde un color en particular. La idea es de hacerles trabajar dichas

nociones, teniendo en cuenta el contexto derecho y el contexto izquierdo en el cual se hallan los elementos que faltan en el texto.

En cada espacio que se debe llenar, tan sólo se hallará una lista de elementos pertenecientes a la cadena correferencial que ha sido anotada.

**ELITE-[FLE]2**

Niveau : B1-B2 CEFR

Au coin de la rue de Vaugirard, comme ils longeaient déjà les bâtiments de l'École, **M. Thibault**, [ ] pendant le trajet n'avait pas adressé la parole à [ ] [ ] arrêta brusquement :

Ah, cette fois, **Antoine**, non, cette fois, ça dépasse ! [ ] ne répondit pas.

L'École était fermée. C'était dimanche, et il était neuf heures du soir. **Un portier** entrouvrit le guichet.

Savez [ ] où est mon frère ? cria [ ] [ ] écarquilla les yeux. [ ] trappa du pied. Allez chercher l'abbé **Binot**.

[ ] précéda les deux hommes jusqu'au parloir, tira de [ ] un rat-de-cave, et alluma le lustre.

Quelques minutes passèrent. [ ] essouffé, [ ] était laissé choir sur une chaise ; [ ] murmura de nouveau, les dents serrées :

Cette fois, [ ] sais, non, cette fois !

Excusez-nous, [ ], dit l'abbé **Binot** qui venait d'entrer sans bruit. Il était fort petit et dut se dresser pour poser la main sur l'épaule d' [ ].

Bonjour, [ ] ! Qu'y a-t-il donc ?

Où est mon frère ? **Jacques** ? Il n'est pas rentré de la journée ! [ ] cria [ ] [ ] [ ] était levé.

Mais, où était-il allé ? fit l'abbé, sans trop de surprise. Ici, parbleu ! À la consigne ! L'abbé glissa ses mains sous sa ceinture :

**Jacques** n'était pas consigné. Quoi ? **Jacques** n'a pas paru à l'École aujourd'hui.

L'affaire se corsait. [ ] ne quittait pas du regard la figure du prêtre. [ ] secoua les épaules, et tourna vers l'abbé [ ] bouffi, dont les lourdes paupières ne se soulevaient presque jamais :

**Jacques** nous a dit hier qu'il avait quatre heures de consigne. Il est parti, ce matin, à l'heure habituelle, et puis, vers onze heures, pendant que nous étions tous à la messe, il est revenu, paraît-il : il n'a trouvé que la cuisinière ; il a dit qu'il ne reviendrait pas déjeuner parce qu'il avait huit heures de consigne au lieu de quatre.

Pure invention, appuya l'abbé.

[ ] ai dû sortir à la fin de l'après-midi, continua [ ], pour porter [ ] à la Revue des Deux Mondes. Le directeur recevait, [ ] ne suis rentré que pour le dîner. **Jacques** n'avait pas reparu. Huit heures et demie, personne. [ ] ai pris peur, [ ] ai envoyé chercher [ ] [ ] était de garde à [ ]. Et nous voilà.

qui  
s'  
M. Thibault  
M. Thibault  
-

Figura 12. Ejercicio de sistematización 1.

## ELITE-[FLE]2

Titre : Les Thibault, tome 1, le cahier gris (fragment)

Auteur : Roger Martin du Gard

Source : Livre numérisé

Date : 1922

Type : Narrative et dialogue

Niveau : B1-CECR

Au coin de la rue de Vaugirard, comme ils longeaient déjà les bâtiments de l'École, M. Thibault,  pendant le trajet n'avait pas adressé la parole à ,  arrêta brusquement :

Ah, cette fois, Antoine, non, cette fois, ça dépasse !  ne répondit pas.

L'École était fermée. C'était dimanche, et il était neuf heures du soir. Un portier entrouvrit le guichet.

Savez  où est mon frère ? cria ,  écarquilla les yeux.  trappa du pied.

Allez chercher l'abbé Binot.

précéda les deux hommes jusqu'au parloir, tira de  un rat-de-cave, et alluma le lustre.

Quelques minutes passèrent. , essoufflé,  était laissé choir sur une chaise :  murmura de nouveau, les dents serrées :

Cette fois,  sais, non, cette fois !

Excusez-nous, , dit l'abbé Binot qui venait d'entrer sans bruit. Il était fort petit et dut se dresser pour poser la main sur l'épaule d' . Bonjour,  ! Ou'y a-t-il donc ?

Où est mon frère ? Jacques ?

Il n'est pas rentré de la journée !  écria ,   était levé.

Mais, où était -il allé ? fit l'abbé, sans trop de surprise. Ici, parbleu ! À la consigne ! L'abbé glissa ses mains sous sa ceinture :

Figura 13. Ejercicio de sistematización 2.

La última actividad de sistematización debería permitir a los estudiantes en formación de adquirir de una mejor manera las nociones de anáforas nominales y pronominales. La siguiente imagen (figura 14, Molina Mejía & Antoniadis, 2015) nos permite observar en que forma los estudiantes pueden dividir los elementos en dos tipos diferentes, según su pertenencia a los correferentes nominales (GN ou N), y a los correferentes pronominales (Pron).

## 5. Análisis y resultados con relación al corpus

Con el fin de focalizar de una mejor manera las actividades que acabamos de presentar, decidimos adaptar las nociones y el etiquetado de nuestro corpus en función de las nociones seleccionadas a partir de la lingüística textual. Es por ello, que una tipología de nociones ligadas a este enfoque lingüístico fue

establecida. En nuestro caso partimos, por ejemplo, no de la noción de oración (muy común en las gramáticas de tipo estructuralista, y otras), sino de la noción de proposición, más acorde con el enfoque de la lingüística textual (Adam, 2011a), y de la lingüística enunciativa (Benveniste, 1970).

**ELITE-[FLE]2**

Ah, cette fois, **Antoine**, non, cette fois, ça dépasse ! **Le jeune homme** ne répondit pas.

L'École était fermée. C'était dimanche, et il était neuf heures du soir. **Un portier** entrouvrit le guichet. Savez-vous où est mon frère ? cria **Antoine**.

L'autre écarquilla les yeux. **M. Thibault** frappa du pied. Allez chercher l'abbé Binot.

**Le portier** précéda les deux hommes jusqu'au parloir, tira de sa poche un rat-de-cave, et alluma le lustre.

Quelques minutes passèrent. **M. Thibault**, essoufflé, s'était laissé choir sur une chaise ; il murmura de nouveau, les dents serrées :

Cette fois, je sais, non, cette fois !

Excusez-nous, **Monsieur**, dit l'abbé Binot qui venait d'entrer sans bruit. Il était fort petit et dut se dresser pour poser la main sur l'épaule d'Antoine. Bonjour, jeune docteur ! Qu'y a-t-il donc ?

Où est mon frère ? Jacques ? Il n'est pas rentré de la journée ! s'écria **M. Thibault**, qui s'était levé.

Mais, où était-il allé ? fit l'abbé, sans trop de surprise. Ici, parbleu ! À la consigne ! L'abbé glissa ses mains sous sa ceinture :

Jacques n'était pas consigné. Quoi ? Jacques n'a pas paru à l'École aujourd'hui.

L'affaire se corsait. **Antoine** ne quittait pas du regard la figure du prêtre. **M. Thibault** secoua les épaules, et tourna vers l'abbé son visage bouffi, dont les lourdes paupières ne se soulevaient presque jamais.

Jacques nous a dit hier qu'il avait quatre heures de consigne. Il est parti, ce matin, à l'heure habituelle, et puis, vers onze heures, pendant que nous étions tous à la messe, il est revenu, paraît-il ; il l'a trouvée que la cuisinière ; il a dit qu'il ne reviendrait pas déjeuner parce qu'il avait huit heures de consigne au lieu de quatre.

Pure invention, appuya l'abbé.

Il ai dû sortir à la fin de l'après-midi, continua **M. Thibault**, pour porter ma chronique à la Revue des Deux Mondes. Le directeur recevait, je ne suis rentré que pour le dîner. Jacques n'avait pas reparu. Huit heures et demie, personne. Il ai pris peur, j'ai envoyé chercher Antoine qui était de garde à son régime. Et nous voilà.

Attention vous vous êtes trompé 7 fois: 1- (1)de catégorie pron 2- (5) de catégorie GN

	M. Thibault	Antoine	Un portier
<input checked="" type="radio"/> Coref <input type="radio"/> GN <input type="radio"/> ou N <input type="radio"/> Coref <input type="radio"/> Pron	qui s' M. Thibault M. Thibault s' I Monsieur s' M. Thibault qui s' M. Thibault son visage é M. Thibault ma chronique et J. I	qui s' Le jeune homme Antoine s' Antoine qui son visage	s' L'autre Le portier sa poche Antoine jeune docteur

Figura 14. Ejercicio de sistematización 3.

Esto nos permitió aligerar un poco más nuestro corpus, haciéndolo además más dinámico, ya que en lugar de oraciones (cuya noción y delimitación son en ocasiones algo difíciles de cernir), o de chunks (para los cuales algunos lingüistas aún no logran ponerse de acuerdo acerca de su valor en el campo de la lingüística<sup>22</sup>), nosotros preferimos emplear la noción de proposición, en la forma

<sup>22</sup> La noción de chunk es analizada por A. Souque (2014: 32-33). Según esta autora: “Para Abney (1991) los chunks delimitan un grupo de palabras, para Giguet (1998) se trata de



en que esta es analizada por Jean-Michel Adam (2011a), y de la manera en la que ha sido descrita en la “*Grammaire méthodique du français*” (Gramática metódica del francés) (Riegel *et al.* 2009). En efecto, la proposición (o proposición-enunciado) es una unidad mínima portadora de sentido, ya que se trata de: “una unidad de base, efectivamente realizada y producida por un acto de enunciación, y por lo tanto como un enunciado mínimo” (Adam, 2011a).

Luego, nos dedicamos a anotar otras nociones en función de los fenómenos estudiados (*cf.* sección 2.3). Por ejemplo:

En el caso de la correferencia textual, anotamos los referentes principales [type → Nom], luego los correferentes [position → anaphores | cataphores]; [type → Nom | Pron]; [sous-type → GN | COD | Pers | Rel | N | Compl | Pronom | COI | NP | NC]; [classe → fidèle | infidèle | déterminative | totale | partielle | intégrante | associative]. Le asignamos a cada referente y correferente unos identificadores que nos permiten luego localizar las cadenas de correferencia textuales.

La DTD relacionada con nuestro corpus y las anotaciones nos permitieron la realización de actividades en las cuales las unidades textuales pueden ser desplazadas, gracias a lenguajes de programación como Java<sup>23</sup>, lo que permite aislar las unidades anotadas en XML, y de transformarlas en entidades que pueden “ser movidas”, y que además son reconocidas por el sistema (es decir, pueden ser validadas en términos informáticos). Esto último es de extrema importancia, ya que la validación de las actividades pasa por esta etapa, lo que se convierte en uno de los puntos fuertes de nuestro sistema, ya que hace automático este proceso para cualquier tipo de texto, teniendo como único requisito de estar en conformidad con la DTD.

---

'sintagmas mínimos', Vergne y Giguet (1998) hablan de 'sintagmas no recursivos', Lebarbé (2002) define el chunk como una unidad constituida por una palabra lexical, rodeada por una costelación de palabras funcionales (determinantes, pronombres, adverbios, etc.)”. Finalmente para Souque (2014: 33): “Se trata de una unidad, que llamaremos de manera indiferente chunk o sintagma, definida no en función de su contenido, sino en función de sus fronteras, principalmente de palabras gramaticales [...], de marcas morfológicas [...] y de puntuaciones”.

<sup>23</sup> Java es un lenguaje de programación informático orientado a objetos creado a mediados de los años 90 por James Gosling y Patrick Naughton. Este lenguaje, gracias a su singularidad, permite convertir las unidades textuales en objetos y, de este modo, permitir que se muevan dentro del texto, y que sean validadas por el sistema.

## 6. Conclusión y perspectivas

Le corpus que hemos constituido con textos auténticos, y anotados en concordancia con las nociones que deseábamos ayudar a aprender en el campo de la lingüística textual, nos permitió en un primer momento de realizar un primer prototipo operacional del sistema ELiTe-[FLE]<sup>2</sup>. Hemos desarrollado por ahora algunas actividades para el análisis de los correferentes textuales, pero esperamos en un segundo instante, el poder desarrollar otras actividades en relación con las otras nociones anotadas.

Una vez el sistema se halle completado, con todas las anotaciones y con todas las actividades almacenadas en la base de datos, tenemos la fiel intención de ensayar el sistema en los cursos de formación de futuros docentes de FLE. En efecto, la idea es de mejorar el sistema en función de la evaluación que los usuarios (docentes-formadores y estudiantes en formación) realizarán en las universidades colombianas que mayor impulso le han dado a este tipo de formación (Universidad de Antioquia y Universidad Nacional de Colombia).

Consideramos, finalmente, que este modelo puede utilizarse para la formación en otras lenguas extranjeras. De hecho, existe en la actualidad un proyecto que lideramos en el grupo Corpus Ex Machina de la Universidad de Antioquia para la creación de un sistema de ALAO, basado en corpus lingüístico, para la enseñanza y el aprendizaje del Español Lengua Extranjera (ELE). La idea es hacer un trabajo parecido al propuesto en el marco de la formación de docentes de FLE, pero en esta ocasión aplicado a aprendientes en general del español.

## Referencias

- Adam, J.-M. 1990. *Éléments de linguistique textuelle : Théorie et pratique de l'analyse textuelle*. Liège: Pierre Mardaga.
- Adam, J.-M. 2010. L'émergence de la Linguistique Textuelle en France : entre perspective fonctionnelle de la phrase, grammaires et linguistiques du texte et du discours. In *Revista Investigações* 23(2): 11-47.
- Adam, J.-M. 2011a. *La linguistique textuelle*. Paris: Armand Colin.
- Adam, J.-M. 2011b. *Les textes : types et prototypes*, 3<sup>rd</sup> edition. Paris: Armand Colin.
- Aston, G. 1995. Corpora in Language Pedagogy: Matching Theory and Practice. In G. Cook & B. Seidlhofer (eds), *Principle and Practice in Applied Linguistics: Studies in Honour of H. G. Widdowson*. Oxford: Oxford University Press, Chap. 17, 257-270.
- Arismendi, F. & Colorado, D. 2015. La formation des enseignants de FLE en Colombie : panorama et cas de l'université d'Antioquia. In *Dialogues et cultures* 61 (*La formation initiale des enseignants de français langue étrangère*): 44-61.

- Barthe, M. & Chovelon, B. 2009. *Le français par les textes : Volume 1*. Grenoble: Presses Universitaires de Grenoble.
- Barthès, R. 1966. Introduction à l'analyse structurale des récits. In *Communications, Recherches sémiologiques : l'analyse structurale du récit* 8: 1-27.
- Benveniste, E. 1970. *Problèmes de linguistique générale II*. Paris: Gallimard.
- Boulton, A. 2007. Esprit de corpus : Promouvoir l'exploitation de corpus en apprentissage des langues. In G. Williams (ed.), *Actes des cinquièmes journées de la linguistique de corpus, Texte et Corpus*, Vol. 3. Lorient: Université de Bretagne Sud, 37-46.
- Boulton, A. & Tyne, H. 2014. *Des documents authentiques aux corpus : démarches pour l'apprentissage des langues*. Paris: Didier.
- Cadet, L. & Tellier, M. 2007. Le geste pédagogique dans la formation des enseignants de LE : Réflexions à partir d'un corpus de journaux d'apprentissage. In *Les cahiers de Théodile* 7: 67-80.
- Chanier, T. & Ciekanski, M. 2010. Utilité du partage des corpus pour l'analyse des interactions en ligne en situation d'apprentissage : un exemple d'approche méthodologique autour d'une base de corpus d'apprentissage. In *Alsic* (online) 13 | 2010. DOI: 10.4000/alsic.1666. <http://alsic.revues.org/1666> (accessed December 1, 2016).
- Coseriu, E. 1969. *Teoría del lenguaje y lingüística general*. Editorial Gredos S.A., Madrid.
- Giguët, E. 1998. Méthode pour l'analyse automatique de structures formelles sur documents multilingues. PhD diss., Université de Caen.
- Giuliani, D. & Hannachi, R. 2010. Linguistique de corpus et didactique du F.L.E. Une exploitation du corpus IntUne. In *Cahiers de praxématique* (online), 54-55 | 2010, document 8. <http://praxématique.revues.org/1136> (accessed December 1, 2016).
- Gries, S.Th. 2010. Corpus linguistics and theoretical linguistics: A love-hate relationship? Not necessarily... In *International Journal of Corpus Linguistics* 15(3): 327-343.
- Habert, B., Fabre, C. & Issac, F. 1998. *De l'écrit au numérique : Constituer, normaliser et exploiter les corpus électroniques*. Paris: Masson.
- Halliday, M.A.K. & Hasan, R. 1976. *Cohesion in English*. London: Longman.
- Johns, T. 1991. From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. In *English language research journal* 4 (*Classroom concordancing*, T. Johns & P. King eds): 27-45.
- Kübler, N. 2014. Mettre en œuvre la linguistique de corpus à l'université : Vers une compétence utile pour l'enseignement/apprentissage des langues ? In *Recherches en didactique des langues et des cultures : Les Cahiers de l'Acedle* 11(1): 37-77.
- Landure, C. 2011. Data-Driven Learning : apprendre et enseigner à contre-courant. In *Mélanges CRAPEL* 32: 163-178.
- Lebarbé, T. 2002. Hiérarchie inclusive des unités linguistiques en analyse syntaxique coopérative. PhD diss., Université de Caen.
- Lundquist, L. 1980. *La cohérence textuelle : syntaxe, sémantique, pragmatique*. Copenhagen: Nyt NordiskForlag.
- Lundquist, L. 1990. *L'analyse textuelle : méthode, exercices*, 2<sup>nd</sup> edition. Copenhagen: DJOF Publishing.
- Lundquist, L. 2013. *Lire un texte académique en français*. Paris: Éditions Ophrys.
- Lundquist, L. 2014. La didactisation du Français Langue Étrangère : discours, genre et type de texte. In M. Monte & G. Philippe (eds), *Genres & textes : déterminations, évolutions*,

- confrontations. Études offertes à Jean-Michel Adam.* Lyon: Presses Universitaires de Lyon, 287-296.
- Malrieu, D. 2004. Linguistique de corpus, genres textuels, temps et personnes. In *Langages 153 (Les genres de la parole, S. Bouquet ed.): 73-85.*
- Molina Mejía, J. M. 2014. Conception d'un environnement informatique fondé sur la linguistique textuelle et destiné à la formation des enseignants de FLE en Colombie. In P.-A. Caron & R. Champagnat (eds), *Actes des 5ièmes Rencontres des Jeunes Chercheurs en Environnements Informatiques pour l'Apprentissage Humain*, 23-29. <https://apps.univ-lr.fr/cgi-bin/WebObjects/Colloque.woa/wa/colloque?code=1041> (accessed December 1, 2016).
- Molina Mejía, J.M. 2015. *EliTe-[FLE]<sup>2</sup> : Un environnement d'ALAO fondé sur la linguistique textuelle, pour la formation linguistique des futurs enseignants de FLE en Colombie.* PhD diss., Université Grenoble Alpes.
- Molina Mejía, J.M. & Antoniadis, G. 2014a. Toward the Constitution of a Hybrid Learning Environment for the FFL Teacher's Training in Colombian Universities Based on Text Linguistics. In G. Quiroz & P. Patiño (eds), *LSP in Colombia: Advances and Challenges.* Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien: Peter Lang, Chap. 152 33-249.
- Molina Mejía, J.M. & Antoniadis, G. 2014b. Conception d'un environnement informatique fondé sur la linguistique textuelle pour la formation initiale des futurs enseignants de FLE en Colombie. In *Adjectif* (online): Article 293. <http://www.adjectif.net/spip/spip.php?article293> (accessed December 1, 2016).
- Molina Mejía, J.M. & Antoniadis, G. 2014c. Conception et élaboration d'un environnement informatique fondé sur la linguistique textuelle et destiné à la formation des futurs enseignants de FLE en Colombie. In F. Olmo Cazevielle & J.-M. Mangiante (eds.), *II Coloquio franco-español de análisis del discurso y enseñanza de lenguas para fines específicos. Lenguas, comunicación y tecnologías digitales.* Valencia: Editorial Universitat Politècnica de Valencia, 143-158.
- Molina Mejía, J.M. & Antoniadis, G. 2015. Un environnement informatique pour la formation des formateurs en FLE Colombiens, fondé sur la linguistique textuelle. In *Alsic* (online), 18(2). DOI: 10.4000/alsic.2843. <http://alsic.revues.org/2843> (accessed December 1, 2016).
- Parodi, G. 2008. Lingüística de corpus: Una introducción al ámbito. In *RLA. Revista de Lingüística Teórica y Aplicada*, 46(1): 93-119.
- Rastier, F. 2004. Enjeux épistémologiques de la linguistique de corpus. In *Texto!* (online). [http://www.revue-texto.net/Inedits/Rastier/Rastier\\_Enjeux.html](http://www.revue-texto.net/Inedits/Rastier/Rastier_Enjeux.html) (accessed December 1, 2016).
- Riegel, M., Pellat, J.-C. & Rioul, R. 2009. *Grammaire méthodique du français.* Paris: Presses Universitaires de France.
- Souque, A. 2014. *Modèle de vérification grammaticale automatique gauche-droite.* PhD diss., Université de Grenoble.
- Teubert, W. 2009. La linguistique de corpus : une alternative. In *Semen* (online) 27. <http://semen.revues.org/8914> (accessed December 1, 2016).
- Van Dijk, T.A. 1972. *Some Aspects of Texts Grammars.* The Hague: Mouton.
- Van Dijk, T.A. 1983. *La ciencia del texto.* Barcelona: Paidós Comunicación.

- Vergne, J. & Giguët, E. 1998. Regards théoriques sur le "tagging". In *Actes de la 5ème conférence sur le Traitement Automatique des Langues Naturelles* (TALN 1998). Paris: Association pour le Traitement Automatique des Langues.  
[http://www.atala.org/taln\\_archives/TALN/TALN-1998/taln-1998-long-010.pdf](http://www.atala.org/taln_archives/TALN/TALN-1998/taln-1998-long-010.pdf)  
(accessed December 1, 2016).
- Weinrich, H. 1973. *Temps*. Paris: Éditions du Seuil.
- Williams, G. 2006. La linguistique de corpus, une affaire prépositionnelle. In *Texto!* (online).  
<http://www.revue-texto.net/Parutions/Livres-E/Albi-2006/Williams.pdf>  
(accessed December 1, 2016).