¿Qué sería de la vida sin la anotación morfosintáctica?

¿Qué de la belleza, si perdiese su cualidad y fuese solo algo bello?

¿Qué de La Rioja, si las etiquetas se confundiesen y *vino* fuera solo un verbo?

¿Qué de la Filosofía, si no existiera la etiqueta *sustantivo*?

¿Qué de los que me rodean, si el POS *tagging* errase y no supieran si los amo o soy su amo?

En fin ¿qué sería del génesis mundial si al principio no hubiera sido el *verbo?*

1. Breve caracterización de SpaCy y FreeLing

El presente trabajo consiste en el análisis comparativo de la tokenización, lematización y etiquetado morfosintáctico de un breve texto en español mediante dos herramientas del Procesamiento del Lenguaje Natural (PLN): SpaCy y FreeLing.

Si bien ambas bibliotecas son de código abierto, solo SpaCy se ha ejecutado mediante lenguaje Python, mientras que FreeLing se ha empleado en su versión demostrativa *online.*

Desarrollada en sus orígenes por Explosion AI, SpaCy ha sido concebida, principalmente, para aplicaciones industriales, mientras que FreeLing, creada por el centro de Tecnologías y Aplicaciones del Lenguaje (TALP) de la Universitat Politècnica de Catalunya (UPC) se ha ideado para facilitar tareas de análisis lingüístico, no solo en el ámbito industrial, sino también en proyectos de investigación.

Aunque estas finalidades no son excluyentes, determinan parte de las diferencias fundamentales entre ambos etiquetadores.

SpaCy da cobertura a más de setenta y cinco idiomas de todo el mundo y el modelo de etiquetado POS que emplea para el español es el marco de anotación universal *Universal Dependencies*[1] (UD). En cambio, los idiomas soportados por FreeLing son lenguas europeas[2], algunas, como el galés o el asturiano, minoritarias, de ahí que siga, con alguna variante, el estándar, desarrollado por el *Expert Advisory Group on Language Engineering Standards* (EAGLES), en el marco del programa de Investigación Lingüística e Ingeniería de la CE para la anotación de textos en PLN.

Ambas herramientas coinciden en las principales tareas de PLN que desempeñan: tokenización, etiquetado *Part of Speech* (POS), desambiguación morfosintáctica, reconocimiento de entidades nombradas y análisis sintáctico. Sin embargo, entre las tareas específicas que ofrece SpaCy destacan la clasificación de texto y la evaluación de similitud semántica. Por su parte, entre los servicios exclusivos de FreeLing se encuentran la extracción de grafos semánticos y la codificación fonética.

Las dos aplicaciones permiten personalizar las tareas de PLN que ofrecen: SpaCy mediante el empleo de modelos basados en redes neuronales, como BERT o RoBERTa y FreeLing debido a su diseño modular que posibilita, en su versión instalada, adaptar componentes lingüísticos a necesidades específicas, como el estudio diacrónico del lenguaje.

En definitiva, mientras que SpaCy, es una librería de tercera generación, basada, principalmente, en modelos de aprendizaje automático y redes neuronales profundas que la adecúan para proyectos que requieren versatilidad, FreeLing es una herramienta

[1] El corpus en español en el que se basa SpaCy es AnCora.
[2] Las lenguas que cubre FreeLing son: inglés, español, portugués, francés, italiano, alemán, ruso, noruego, catalán, gallego, croata, esloveno, asturiano y galés.

de segunda generación, por lo que su combinación de reglas lingüísticas y modelos probabilísticos, la convierten en una aplicación idónea para el análisis lingüístico detallado de lenguas poco documentadas.

2. Precisión de cada herramienta

Etiquetado[3]

Tras realizar el análisis contrastivo del etiquetado morfosintáctico (POS *tagging*) del texto propuesto, efectuado por las dos herramientas, se evidencia una gran diferencia entre la especificad del conjunto de etiquetas *(tagset)* de FreeLing y la generalidad del de SpaCy.

Esta diferencia se explica a partir de los distintos propósitos, no excluyentes, ya descritos: empresarial y de carácter global, en el caso de SpaCy, y académico-europeo, en el de FreeLing.

Estos objetivos justifican, entre otros aspectos, las distintas elecciones, antes mencionadas, del conjunto de etiquetas: SpaCy utiliza el marco UD, que, como su nombre indica, persigue representar de modo general la morfología de numerosas lenguas mundiales, mientras que FreeLing emplea el estándar EAGLES, que pretende reflejar minuciosamente los accidentes gramaticales de las lenguas europeas.

[3] Enlaces a los *tagsets* empleados por SpaCy y FreeLing: https://universaldependencies.org/treebanks/es_ancora/index.html https://freeling-user-manual.readthedocs.io/en/latest/tagsets/tagset-es/

Por ejemplo, para el verbo *eran,* mientras que SpaCy, utiliza la etiqueta *AUX,* referida a la naturaleza auxiliar del verbo, FreeLing asigna una etiqueta que refleja detalles, no solo sobre la naturaleza verbal, sino también sobre sus morfemas flexivos: *VSII3P0*.

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
1	fue del	fue del	fue de el	ser del	ser de el	AUX ADP	VSIS3S0 SP DA0MS0
2	De repente se dio cuenta que	De repente se dio cuenta que	De_repente se dio_cuenta que	de repente él dar cuenta que	de_repente se dar_cuenta que	ADP NOUN PRON VERB NOUN SCONJ	RG P00CN00 VMIS3S0 PR0CN00
3	eran 21:15	eran 21:15	eran 21:15	ser 21:15	ser 21:15	AUX NUM	VSII3P0 Z

Detalle de la distinta especificidad entre los *tagset* de SpaCy y FreeLing

Unidades léxicas complejas

Debido a que el significado de ciertas expresiones no puede deducirse de la suma de sus componentes, el análisis de unidades léxicas complejas (*multiwords)* es un desafío para los etiquetadores morfosintácticos.

Según el texto analizado, FreeLing es una herramienta más precisa que SpaCy en la tokenización, lematización y etiquetado POS de estas unidades.

De este modo, los nombres propios compuestos, como *Los Halcones* son tokenizados como una unidad por FreeLing: *Los_halcones,* etiqueta POS: *NP00G00*[4] y segmentados por Spacy: *Los Halcones,* etiqueta POS: *DET, PROPN.*

Análogamente, frente a SpaCy que fragmenta las locuciones adverbiales y prepositivas, *de repente, a pesar de*, en sus componentes (*de repente*: *ADP, NOUN*) FreeLing les atribuye correctamente una sola etiqueta, aunque esta no es *locución adverbial o preposicional,* según corresponda*,* sino que se obvia el término exacto *locución* (*de_ repente: RG; a_pesar_de: SP).*

A pesar de su mayor precisión en el tratamiento de unidades léxicas complejas, FreLing, al igual que SpaCy, tampoco identifica ciertas locuciones[5] adverbiales, como *sin embargo, una vez, en un santiamén* o la locución adjetival coloquial *al loro*, que son tratadas como varios tokens y, en consecuencia, incorrectamente etiquetadas morfosintácticamente. Ej.: *sin embargo:* etiquetado

[4] Aun así, la etiqueta es incorrecta, pues se identifican *Los Halcones* como localización.
[5] La locución verbal *se dio cuenta* será comentada en el apartado posterior.

POS: FreeLing: *SP, NCMS000*; SpaCy *ADP, NOUN; una vez:* FreeLing*: DI0FS0 NCFS000;* SpaCy *DET, NOUN; en un santiamén:* FreeLing: *SP*, *DI0MS0, NCMS000*; SpaCy: *ADP, DET, NOUN; al loro:* FreeLing*: SP, DA0MS0, NCMS000*; SpaCy*: ADP, NOUN.*

Por otro lado, ejemplos textuales como *empezaron a cantar* o *seguir divirtiéndose* evidencian que ninguna de las dos aplicaciones identifica las perífrasis verbales como una unidad léxica, por lo que, al tokenizarse independientemente, el etiquetado POS es en ambas incorrecto y no se distingue entre verbo auxiliar y verbo auxiliado (*empezaron a cantar:* etiquetado POS: FreeLing: *VMIS3P0, SP, VMN0000;* SpaCy: *VERB, ADP, VERB*).

Este aspecto, quizá, pueda remediarse mediante el empleo de diccionarios perifrásticos y se propone su tokenización guionizada: *empezaron_a_ cantar,* para preservar, simultáneamente, la posibilidad de movilidad de los constituyentes de la perífrasis y su unidad.

Clíticos: enclíticos y proclíticos

Enclíticos

En español los pronombres átonos, cuando aparecen como enclíticos tras la base verbal a la que se adjuntan, dependen morfofonológicamente de ciertos verbos, con los que constituyen una solo palabra gráfica, pero no gramatical.

SpaCy y FreeLing abordan de manera diferente la tokenización, la lematización y el etiquetado POS de estos enclíticos.

FreeLing los separa del verbo base sistemáticamente, con un enfoque granular que permite analizar cada unidad gramatical independientemente (por ejemplo, *reanimarla* es segmentado como dos tokens: *reanimar* y *la),* Sin embargo, esta tokenización,

basada en la separación sistemática de enclíticos, puede considerarse errada en contextos donde el pronombre constituye un morfema verbal, como en *zambulléndose (zambullendo se: VMG0000, PP3CN00) o en divirtiéndose (divirtiendo se: VMG0000, PP3CN00).*

En oposición, SpaCy trata siempre los enclíticos como una parte integral del verbo, por lo que, si bien con este análisis atina en la tokenización de verbos pronominales como *sintiéndose* (*sintiéndose: VERB*), no refleja la independencia gramatical del enclítico en otros casos, como en *reanimarla* (*reanimarla: VERB*).

Proclíticos

Debido a que no responden a los patrones precisos de los enclíticos, en las oraciones propuestas, las dos herramientas fragmentan la unidad de verbos o locuciones verbales pronominales o pronominales medios (*se dio cuenta, se reunió, se quitó*) y tokenizan y etiquetan segmentadamente formas verbales que debieran considerarse una unidad con el pronombre átono, pues, aunque esta descomposición dé cuenta de la morfología interna de esas formas, no respeta su integridad léxica, con el, a veces, consiguiente cambio semántico.

Ejemplos de este hecho los observamos, en *se dio cuenta*, locución tokenizada y anotada de modo más unitario por FreeLing (SpaCy la tokeniza y, en consecuencia, anota en unidades independientes; *se dio cuenta;* etiquetado POS*: PRON, VERB, NOUN*) pero no perfectamente, pues fragmenta el *se* de la forma verbal: *se dio_cuenta: P00CN00, VMIS3S0*, cuando *darse cuenta de,* no significa lo mismo que *dar cuenta de.*

Para el etiquetado de clíticos se propone, como solución integradora considerar los verbos pronominales como un solo token y en los demás clíticos preservar la unidad morfofonológica, por un lado, y la independencia gramatical de las unidades léxicas constituyentes, por otro, mediante el uso del guion. Ej.: *reanimar_la.*

Contracciones

A pesar de basarse en el estándar EAGLES, que considera las contracciones como una unidad, FreeLing las expande y separa (por ejemplo, token *del*: *de el;* token *al: a el*).

En contraste, SpaCy las trata como elementos contractos y las tokeniza unitariamente: *del, al*.

Lo que pudiera parecer falta de precisión por parte de FreeLing, no lo es, pues, entre las tareas de PLN que se ofrecen en su página web se encuentra esta expansión contractiva, dirigida a contextos en los que es esencial comprender los componentes gramaticales individualmente.

Por el contrario, la conservación de la contracción por parte de SpaCy, si bien implica mayor precisión en la tokenización, tiene por inconveniente la falta de especificidad en su etiquetado POS (*al, del: ADP*), hecho que puede limitar su precisión en determinados casos.

Signos de puntuación

Uno de los aspectos más divergentes entre ambas herramientas es en el tratamiento de los signos de puntuación.

En este sentido, FreeLing destaca por su alta precisión: identifica adecuadamente como tokens los signos de puntuación y la etiqueta POS correspondiente es específica.

Frente a esta exactitud, SpaCy, bajo la etiqueta POS, *PUNCT,* se refiere a cualquier signo de puntuación, independientemente del tipo. Además, en algunos casos, estos signos quedan sin tokenizar (por lo que el etiquetado POS correspondiente es erróneo) o con una etiqueta morfosintáctica equivocada.

Un ejemplo revelador de este dispar tratamiento de los signos de puntuación se observa en la oración siete, en la que aparecen signos como las comillas o el guion precediendo al estilo directo.

Como se observa en la imagen adjunta, FreeLing tokeniza adecuadamente las comillas y el guion. Las comillas se distinguen con etiquetas diferenciadas, según sean de apertura o cierre *(Fra, Frc*) y el guion se designa correctamente como *Fg (F. Hyphen).*

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
7	“ - Hay pasa - “	“ -Hay pasa- “	“ - Hay pasa - “	“ -haber pasa- “	“ - haber pasar - “	PROPN VERB PUNCT PROPN	Fra Fg VMIP3S0 VMIP3S0 Fg Frc

Detalle del tratamiento de los signos de puntuación por SpaCy y FreeLing

Por el contrario, SpaCy no tokeniza los guiones, que, según se encuentren a inicio o fin de cita, adjunta respectivamente al token posterior (etiquetándose entonces como la categoría gramatical de este), o al anterior (en cuyo caso ambas unidades son designadas con la etiqueta POS*, PUNCT).*

Por lo que respecta a las comillas, si bien son correctamente tokenizadas por SpaCy, la etiqueta morfológica asignada es errada: *PROPN*.

Homografía

Dado que los anotadores automáticos deben decidir cuál es la opción adecuada al contexto entre varias, la homografía es causa de ambigüedad en el análisis morfosintáctico.

Con respecto al texto analizado, tanto SpaCy como FreeLing asignan una etiqueta POS incorrecta a las voces homónimas *nada y fue.*

En la secuencia, *pero no nada* se asigna a *nada* la etiqueta POS de pronombre, en lugar de verbo: SpaCy: *PRON*; FreeLing, *PI0CS00* y, en consecuencia, se lematiza como tal.

Por otro lado, la etiqueta POS de *fue* se corresponde con un verbo auxiliar, *AUX,* en SpaCy y como uno semiauxiliar en FreeLing, *VSIS3S0,* por lo que las dos herramientas lo lematizan equivocadamente como el verbo *ser*.

Sin embargo, la homografía es resuelta correctamente por ambas herramientas en referencia a *cuarto,* que aparece en las oraciones segunda y tercera con distintos sentidos: en una como sustantivo, *en su cuarto,* y en otra como adjetivo numeral, *el cuarto*

piso. Ambos casos son desambiguados morfológicamente con las etiquetas correctas por los dos etiquetadores, debido, probablemente, a la presencia o ausencia de un sustantivo adyacente.

Reconocimiento de entidades temporales

Ambos recursos son imprecisos al asignar incorrectamente la etiqueta POS a la hora *21:15.* SpaCy la anota como *NUM* y FreeLing como *Z,* etiquetas genéricas para números.

Aunque FreeLing cuenta con un módulo para la detección de entidades temporales, es posible que la versión *online* no contemple patrones específicos para este formato o carezca de una *Augmented Transition Network* (ATN) específica para español. Por otro lado, SpaCy ofrece extensiones como *Date SpaCy,* diseñadas para procesar patrones temporales, que no han sido configuradas en este análisis.

3.Puntos de similitud y divergencia

Similitudes

Tratamiento de términos inusuales

Ambas aplicaciones muestran precisión en la identificación de *SAMUR* como un nombre propio, incluso cuando podría haber sido un término fuera de vocabulario (OOV).

Divergencias

Normalización de las mayúsculas en la tokenización y lematización

Según la muestra analizada, mientras que en la tokenización las dos herramientas preservan la mayúscula inicial, característica de los nombres propios en español, en la lematización, FreeLing opta por normalizar todos los tokens, incluidos los nombres propios, a minúsculas. Tal es el caso de *Santiago, Los Halcones* o *SAMUR* que FreeLing lematiza respectivamente como *santiago*, *los_halcones y samur*.

De este modo, asegura la compatibilidad con tareas del PLN en las que esa normalización a minúsculas es necesaria, al tiempo que conserva la información sobre la capitalización de los términos en los tokens o en el etiquetado POS (lema: *samur:* etiqueta POS: *NP00000*).

Por otro lado, SpaCy conserva la mayúscula inicial en la lematización de nombres propios*: Santiago, el Halcones, SAMUR,* aspecto útil en tareas en las que los lemas, almacenados en bases de datos, se vinculan a sistemas externos. De esta manera, el reconocimiento de entidades nombradas mediante el lema garantiza una correspondencia exacta con las entidades existentes, necesaria para la creación, por ejemplo, de grafos de conocimiento.

Lematización de los pronombres personales átonos

Mientras que SpaCy lematiza no solo los pronombres personales tónicos, sino también los átonos, con la forma tónica *él,* FreeLing establece una correspondencia entre el token y el lema de los pronombres átonos con distintas formas, dependiendo de la función sintáctica que desempeñan o de si se trata de la forma*,* invariable en género y número, *se*: la →lo; le →le; se →se.

Así, por ejemplo, la lematización de SpaCy de *se la llevaron→ él él llevar*, contrasta con la de FreeLing: *se la llevaron → se lo llevar*.

A pesar de que el empleo de la forma tónica como lema puede ser más compatible con otros idiomas que no diferencien entre estos pronombres, en lenguas como el español, donde los pronombres átonos tienen una función gramatical importante, mantener la forma átona facilita tareas como la desambiguación gramatical o el análisis de concordancias.

Lematizacion de *cuarto*

Mientras que SpaCy lematiza el adjetivo numeral ordinal *cuarto* como tal*,* FreelIng lo hace como el número entero *4.*

Esta diferencia, probablemente, se base en la pretensión por parte de FreeLing de simplificar la representación de información numérica, frente a la priorización de la fidelidad lingüística por la que se decanta SpaCy, en este caso.

Etiquetado POS del verbo *ser*

Por su escaso contenido léxico, el verbo copulativo *ser* es asimilado por FreeLing y SpaCy como semiauxiliar y auxiliar, respectivamente. Por ejemplo, en *eran las 21:15,* el etiquetado POS es: FreeLing: *VSII3P0*; SpaCy: *AUX.* En *creían ser,* la anotación es: FreeLing: *VSN0000*; SpaCy: *AUX.*

Esta pequeña divergencia ejemplifica el mayor nivel de especificidad que caracteriza al etiquetado de FreeLing.

4. Problemáticas específicas

En la primera y cuarta oraciones FreeLing identifica erróneamente los nombres propios de persona y organización como localizaciones: *Santiago, Los Halcones:* etiqueta POS: *NP00G00.* (Sin embargo, en la oración séptima el antropónimo *Santiago,* aparece bien etiquetado: *NP00SP0*).

En la segunda oración, en el contexto: …*se dio cuenta de que*… FreeLing considera equivocadamente la conjunción *que* como pronombre, *PR0CN00*.

En la sexta oración FreeLing etiqueta erróneamente el adjetivo indefinido *todos* en *Todos ellos…* como pronombre: *PI0MP00.*

En la séptima oración, en…*todo lo que*… el etiquetado POS de *lo* (*PRON*), asignado por SpaCy es incorrecto.

En la décima oración, FreeLing etiqueta el verbo *estar* en *estaba pendiente* como principal (*VMII3S0)*, frente al criterio de considerar el verbo copulativo *ser* como semiauxiliar.

Aspectos problemáticos identificados

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
1	fue del	fue del	fue de el	ser del	ser de el	AUX ADP	VSIS3S0 SP DA0MS0
2	De repente se dio cuenta que	De repente se dio cuenta que	De_repente se dio_cuenta que	de repente él dar cuenta que	de_repente se dar_cuenta que	ADP NOUN PRON VERB NOUN SCONJ	RG P00CN00 VMIS3S0 PR0CN00
3	eran 21:15	eran 21:15	eran 21:15	ser 21:15	ser 21:15	AUX NUM	VSII3P0 Z

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
4	Una vez se reunió Los Halcones	Una vez se reunió Los Halcones	Una vez se reunió Los_halcones	uno vez él reunir el Halcones	uno vez se reunir los_halcones	DET NOUN PRON VERB DET PROPN	DI0FS0 NCFS000 P00CN00 VMIS3S0 NP00G00
5	al ser	al ser	a el ser	al ser	a el ser	ADP AUX	SP DA0MS0 VSN0000
6	empezaron a cantar	empezaron a cantar	empezaron a cantar	empezar a cantar	empezar a cantar	VERB ADP VERB	VMIS3P0 SP VMN0000

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
7	Sin embargo Santiago al loro lo “ - Hay nada le pasa - “	Sin embargo Santiago al loro lo “ -Hay nada le pasa- “	Sin embargo Santiago a el loro lo “ - Hay nada le pasa - “	sin embargo Santiago al loro él “ -haber nada él pasa- “	sin embargo santiago a el loro el “ - haber nada le pasar - “	ADP NOUN PROPN ADP NOUN PRON PROPN VERB PRON PRON PUNCT PROPN	SP NCMS000 NP00SP0 SP DA0MS0 NCMS000 DA00S0 Fra Fg VMIP3S0 PI0CS00 PP3CSD0 VMIP3S0 Fg Frc

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
8	Sintiéndose	Sintiéndose	Sintiendo se	sentir él	sentir se	VERB	VMG0000 PP3CN00
	del	del	de el	del	de el	ADP	SP DA0MS0
	se quitó	se quitó	se quitó	él quitar	se quitar	PRON	P00CN00
	zambulléndose	zambulléndose	zambullendo se	zambullendo él	zambullir se	VERB	VMG0000 PP3CN00

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	POS FreeLing
9	A pesar de reanimarla	A pesar de reanimarla	A_pesar_de reanimar la	a pesar de reanimar él	a_pesar_de reanimar lo	ADP NOUN ADP VERB	SP VMN0000 PP3FSA0
10	estaba al SAMUR	estaba al SAMUR	estaba a el SAMUR	estar al SAMUR	estar a el samur	AUX ADP PROPN	VMII3S0 SP DA0MS0 NP00000
11	en lugar de seguir divirtiéndose	en lugar de seguir divirtiéndose	en_lugar_de seguir divirtiendo se	en lugar de seguir divertir él	en_lugar_de seguir divertir se	ADP NOUN ADP VERB VERB	SP VMN0000 VMG0000 PP3CN00

5. Valoración global y aplicaciones de la anotación morfosintáctica

Tras el análisis textual realizado se ha comprobado que, ambas herramientas son, en general, precisas (en torno, al 83%), tanto en relación con la lematización, la tokenización y el etiquetado morfosintáctico.

Debido a que ciertos fenómenos lingüísticos no siguen patrones fijos, de la comparación textual se deduce que los principales problemas a los que se enfrentan estas aplicaciones son las unidades léxicas complejas, la homografía y el tratamiento de los clíticos,

De los anotadores comparados se han detectado aspectos positivos y carencias, determinados (además de por las dificultades computacionales señaladas, a causa del dinamismo y de la falta de sistematicidad inherente al lenguaje) por el fin para el que fueron diseñadas

De este modo, el *tagest* de carácter universal en que se basa SpaCy (UD) tiene por objetivo la unificación categorial para permitir análisis interlingüísticos, lo que implica la simplificación, mientras que el conjunto de etiquetas en que se fundamenta FreeLing persigue la especificidad.

Frente al tratamiento simplificado de estructuras lingüísticas complejas, la falta de especificidad en contracciones y en signos de puntuación o la simplificación en la lematización pronominal que presenta el etiquetado de SpaCy, el de FreeLing desataca por un intento de rigor y exhaustividad en estas cuestiones.

Sin embargo, al tratarse de un etiquetador de segunda generación, su uso presenta más errores en el etiquetado categorial.

¿Buscas un etiquetador para trabajar con grandes volúmenes de datos en numerosos idiomas? SpaCy es la mejor opción. ¿Prefieres la minuciosidad filológica y el estudio del galés del s. XVIII?, decántate por FreeLing.

Con respecto a las aplicaciones de la anotación morfosintáctica en mi campo de estudio, gestores de corpus como *Scketch Engine* evidencian cómo la lingüística computacional y la lingüística de corpus se retroalimentan: por un lado, los etiquetadores morfosintácticos han sido diseñados para el procesamiento del lenguaje computacionalmente, por otro, la anotación morfosintáctica sirve apara etiquetar corpus que más tarde entrenarán anotadores automáticos.

El hecho de que los etiquetados POS revelen patrones morfosintácticos permite conocer no solo la naturaleza abstracta del lenguaje o las frecuencias combinatorias de sus elementos, sino también la forma discursiva y el registro de los textos. Asimismo, su empleo posibilita la extracción terminológica y la creación de tesauros (al permitir la búsqueda de lemas de la misma categoría.)

Para mejorar estas herramientas sería interesante añadir a sus funciones el análisis morfológico derivativo y así, quizá, poco a poco, la necesaria supervisión humana del etiquetado automático deje de serlo en un futuro próximo.

7. Análisis de un breve texto en italiano

Por ser el italiano una lengua romance, permite una comparación estructural sencilla de los aspectos revelados como problemáticos en el análisis anterior: *multiwords,* contracciones, clíticos, signos de puntuación, OOV, nombres propios y homografías.

Para ello se ha selecciona el siguiente texto: “Di colpo, l'amica di Ilda si accorse che il suo collega aveva dimenticato di inviare l'email al SUEM del Piemonte: «Piano, c'è qualcosa che non va», si disse”, cuyo análisis se sintetiza en la imagen y tablas adjuntas:

Aspectos problemáticos detectados

# Oración	Texto original	Token SpaCy	Token FreeLing	Lema SpaCy	Lema FreeLing	POS SpaCy	TAG SpaCy	POS FreeLing
1	Di colpo , l’ Ilda si accorse al SUEM del Piemonte : « Piano c’ che	Di colpo , l’ Ilda si accorse al SUEM del Piemonte : « Piano c’ che	Di colpo , l’ Ilda si accorse al SUEM_del_Piemonte : « Piano c’ che	di colpo , il Ilda si accorgere a il SUEM di il Piemonte : « piano ci che	di colpo , il ilda si accorgere al suem_del_piemonte : « piano c’ che	ADP NOUN PUNCT DET PROPN PRON VERB ADP PROPN ADP PROPN PUNCT PUNCT PROPN PRON SCONJ	E S FF RD SP PC V E_RD SP E_RD SP FC FB SP PC CS	SPS00 NCMS000 Fc DA0MS0 NP00000 PP3CN00 VMIS3S0 SPCMS NP00000 Fd Fra NCMN000 RG PT00000

Aspecto	SpaCy	FreeLing
Multiwords	Tokeniza segmentadamente *di colpo* y *si accorse.*	Tokeniza segmentadamente *di colpo* y *si accorse.*
Homografías	Etiqueta el adverbio *piano* como nombre propio.	Etiqueta el adverbio *piano* como nombre común.
Contracciones	Expandidas en el lema.	No expandidas.
OOV y nombres propios	Identifica los nombres propios y el posible OOV.	Considera erróneamente *SUEM y Piemonte* como una unidad léxica.
Signos de puntuación	Correctamente identificados y etiquetados.	Correctamente identificados y etiquetados.
Categoría gramatical	Anota incorrectamente *che* como conjunción.	Etiqueta *che* como pronombre, pero erróneamente como interrogativo.
Lematización	Correcta lematización de *c'* en *ci.*	Se equivoca en la lematización de *c'.*

Del análisis se deduce:

que SpaCy y FreeLing muestran diferente precisión, según el idioma del texto analizado. De este modo, en la muestra en italiano la precisión de SpaCy es mayor que en español, por ejemplo, en lo relativo a los signos de puntuación o a la correcta lematización de los pronombres átonos, debido a que el corpus de base, el *Italian Stanford Dependency Treebank (ISDT),* conserva etiquetas más detalladas, en su conversión a UD, que el corpus español, UD AnCora.

Tagset it::isdt

This table is courtesy by Maria Simi.

Tagset `it::isdt`, total 264 tags.

A	A	num=n|gen=n	=> ADJ	_	*biposto, ultra, bene, best, live, running*
A	A	num=p|gen=f	=> ADJ	Gender=Fem|Number=Plur	*violente, personalissime, lisce, paritetiche, vive, immense*
A	A	num=p|gen=m	=> ADJ	Gender=Masc|Number=Plur	*romani, disoccupati, austeri, disposti, radioattivi, corrotti*
A	A	num=p|gen=n	=> ADJ	Number=Plur	*immobili, danesi, principali, presidenziali, lievi, ormonali*
A	A	num=s|gen=f	=> ADJ	Gender=Fem|Number=Sing	*complicata, lontana, missilistica, esigua, immobiliare, famosa*
A	A	num=s|gen=m	=> ADJ	Gender=Masc|Number=Sing	*misterioso, accentuato, preventivo, fondato, fisico, curioso*
A	A	num=s|gen=n	=> ADJ	Number=Sing	*insostenibile, corresponsabile, militare, vergine, keniota, laziale*
A	AP	num=n|gen=n	=> DET	Poss=Yes|PronType=Prs	*loro, altrui, my*
A	AP	num=p|gen=f	=> DET	Gender=Fem|Number=Plur|Poss=Yes|PronType=Prs	*mie, vostre, sue, nostre, proprie*
A	AP	num=p|gen=m	=> DET	Gender=Masc|Number=Plur|Poss=Yes|PronType=Prs	*vostri, propri, tuoi, miei, suoi, nostri*
A	AP	num=s|gen=f	=> DET	Gender=Fem|Number=Sing|Poss=Yes|PronType=Prs	*vostra, mia, sua, propria, nostra, tua*
A	AP	num=s|gen=m	=> DET	Gender=Masc|Number=Sing|Poss=Yes|PronType=Prs	*proprio, tuo, nostro, suo, mio, vostro*
B	B	_	=> ADV	_	*peggio, unicamente, legittimamente, forte, solennemente, carponi*
B	BN	_	=> ADV	PronType=Neg	*no, mica, non, nemmeno, neanche, neppure*
C	CC	_	=> CONJ	_	*bensì, nè, sia, et, mentre, come*
C	CS	_	=> SCONJ	_	*benché, affinché, quale, dopo, poiché, sebbene*
D	DD	num=n|gen=m	=> DET	Gender=Masc|PronType=Dem	*tal*
D	DD	num=p|gen=f	=> DET	Gender=Fem|Number=Plur|PronType=Dem	*queste, quelle*
D	DD	num=p|gen=m	=> DET	Gender=Masc|Number=Plur|PronType=Dem	*quegli, questi, quei*
D	DD	num=p|gen=n	=> DET	Number=Plur|PronType=Dem	*tali*

Imagen del mapeo y conservación del *tagset* del corpus ISDT

Además, aunque similares, la diferente naturaleza de estos idiomas implica la aplicación en ciertos contextos de reglas distintas, hecho que puede influir, por ejemplo, en la no expansión contractiva de FreeLing.

Dado que las problemáticas comunes a las dos lenguas analizadas son, principalmente, el tratamiento de la homografía y de las unidades léxicas complejas se propone:

a) que el corpus en que se base SpaCy conserve un etiquetado morfológico específico, aunque se mapee a UD para aplicarse a numerosas lenguas.
b) que se expliciten patrones morfosintácticos para resolver ambigüedades homonímicas.
c) que se incorporen diccionarios con las *multiwords* más frecuentes.
d) que se priorice la correcta tokenización, pues de ella dependen el etiquetado POS y la lematización.

ANEXOS

Referencias consultadas:

Moreno-Sandoval, Antonio. 2022. Etiquetadores morfosintácticos para corpus en español, Giovanni Parodi, Pascual Cantos, Chad Howe (eds.). *Lingüística de corpus en español*, 404-418. New York: Routledge.

Asociación de Academias de la Lengua Española. 2020. *Nueva gramática básica de la lengua española.* Madrid: Espasa Libros.

Imágenes de la salida del texto en español en formato *txt* dada por SpaCy

#Oración 1

Token	POS	Etiqueta Detallada	Lema
La	DET	DET	el
noche	NOUN	NOUN	noche
de	ADP	ADP	de
las	DET	DET	el
hogueras	NOUN	NOUN	hoguera
,	PUNCT	PUNCT	,
Santiago	PROPN	PROPN	Santiago
,	PUNCT	PUNCT	,
un	DET	DET	uno
chico	NOUN	NOUN	chico
corriente	ADJ	ADJ	corriente
,	PUNCT	PUNCT	,
fue	AUX	AUX	ser
a	ADP	ADP	a
la	DET	DET	el
playa	NOUN	NOUN	playa
a	ADP	ADP	a
celebrar	VERB	VERB	celebrar
su	DET	DET	su
nuevo	ADJ	ADJ	nuevo
logro	NOUN	NOUN	logro
:	PUNCT	PUNCT	:
había	AUX	AUX	haber
terminado	VERB	VERB	terminar
la	DET	DET	el
carrera	NOUN	NOUN	carrera
del	ADP	ADP	del
todo	PRON	PRON	todo
.	PUNCT	PUNCT	.
	SPACE	SPACE	

#Oración 2

De	ADP	ADP	de
repente	NOUN	NOUN	repente
,	PUNCT	PUNCT	,
se	PRON	PRON	él
dio	VERB	VERB	dar
cuenta	NOUN	NOUN	cuenta
de	ADP	ADP	de
que	SCONJ	SCONJ	que
había	AUX	AUX	haber
olvidado	VERB	VERB	olvidar
el	DET	DET	el
papel	NOUN	NOUN	papel
con	ADP	ADP	con
su	DET	DET	su
deseo	NOUN	NOUN	deseo
en	ADP	ADP	en
su	DET	DET	su
cuarto	NOUN	NOUN	cuarto
.	PUNCT	PUNCT	.

#Oración 3

No	ADV	ADV	no
quiso	VERB	VERB	querer
volver	VERB	VERB	volver
,	PUNCT	PUNCT	,
pues	SCONJ	SCONJ	pues
eran	AUX	AUX	ser
las	DET	DET	el
21:15	NUM	NUM	21:15
y	CCONJ	CCONJ	y
subir	VERB	VERB	subir
hasta	ADP	ADP	hasta
el	DET	DET	el
cuarto	ADJ	ADJ	cuarto
piso	NOUN	NOUN	piso
supondría	VERB	VERB	suponer
llegar	VERB	VERB	llegar
tarde	ADV	ADV	tarde
a	ADP	ADP	a
la	DET	DET	el
cita	NOUN	NOUN	cita
.	PUNCT	PUNCT	.

#Oración 4

Una	DET	DET	uno
vez	NOUN	NOUN	vez
en	ADP	ADP	en
la	DET	DET	el
playa	NOUN	NOUN	playa
,	PUNCT	PUNCT	,
se	PRON	PRON	él
reunió	VERB	VERB	reunir
con	ADP	ADP	con
su	DET	DET	su
grupo	NOUN	NOUN	grupo
de	ADP	ADP	de
amigos	NOUN	NOUN	amigo
,	PUNCT	PUNCT	,
Los	DET	DET	el
Halcones	PROPN	PROPN	Halcones
.	PUNCT	PUNCT	.

#Oración 5

Habían	AUX	AUX	haber
elegido	VERB	VERB	elegir
este	DET	DET	este
nombre	NOUN	NOUN	nombre
por	ADP	ADP	por
su	DET	DET	su
amor	NOUN	NOUN	amor
al	ADP	ADP	al
atletismo	NOUN	NOUN	atletismo
,	PUNCT	PUNCT	,
creían	VERB	VERB	creer
ser	AUX	AUX	ser
los	DET	DET	el
más	ADV	ADV	más
veloces	ADJ	ADJ	veloz
corredores	NOUN	NOUN	corredor
.	PUNCT	PUNCT	.

#Oración 6

Todos	DET	DET	todo
ellos	PRON	PRON	él
empezaron	VERB	VERB	empezar
a	ADP	ADP	a
cantar	VERB	VERB	cantar
,	PUNCT	PUNCT	,
bailar	VERB	VERB	bailar
y	CCONJ	CCONJ	y
beber	VERB	VERB	beber
,	PUNCT	PUNCT	,
creyendo	VERB	VERB	creer
que	SCONJ	SCONJ	que
nada	PRON	PRON	nada
pasaba	VERB	VERB	pasar
a	ADP	ADP	a
su	DET	DET	su
alrededor	NOUN	NOUN	alrededor
.	PUNCT	PUNCT	.

#Oración 7

Sin	ADP	ADP	sin
embargo	NOUN	NOUN	embargo
,	PUNCT	PUNCT	,
Santiago	PROPN	PROPN	Santiago
estaba	AUX	AUX	estar
al	ADP	ADP	al
loro	NOUN	NOUN	loro
de	ADP	ADP	de
todo	DET	DET	todo
lo	PRON	PRON	él
que	PRON	PRON	que
sucedía	VERB	VERB	suceder
,	PUNCT	PUNCT	,
y	CCONJ	CCONJ	y
observó	VERB	VERB	observar
que	SCONJ	SCONJ	que
algo	PRON	PRON	algo
no	ADV	ADV	no
iba	VERB	VERB	ir
bien	ADV	ADV	bien
.	PUNCT	PUNCT	.
“	PROPN	PROPN	“
-Hay	VERB	VERB	-haber
una	DET	DET	uno
chica	NOUN	NOUN	chica
en	ADP	ADP	en
el	DET	DET	el
agua	NOUN	NOUN	agua
,	PUNCT	PUNCT	,
pero	CCONJ	CCONJ	pero
no	ADV	ADV	no
nada	PRON	PRON	nada
,	PUNCT	PUNCT	,
algo	PRON	PRON	algo
le	PRON	PRON	él
pasa-	PUNCT	PUNCT	pasa-
”	PROPN	PROPN	”
dijo	VERB	VERB	decir
.	PUNCT	PUNCT	.

#Oración 8

Sintiéndose	VERB	VERB	sentir él
el	DET	DET	el
más	ADV	ADV	más
valiente	ADJ	ADJ	valiente
del	ADP	ADP	del
lugar	NOUN	NOUN	lugar
,	PUNCT	PUNCT	,
se	PRON	PRON	él
quitó	VERB	VERB	quitar
la	DET	DET	el
ropa	NOUN	NOUN	ropa
en	ADP	ADP	en
un	DET	DET	uno
santiamén	NOUN	NOUN	santiamén
y	CCONJ	CCONJ	y
,	PUNCT	PUNCT	,
zambulléndose	VERB	VERB	zambullendo él
en	ADP	ADP	en
el	DET	DET	el
agua	NOUN	NOUN	agua
,	PUNCT	PUNCT	,
rescató	VERB	VERB	rescatar
a	ADP	ADP	a
la	DET	DET	el
bañista	PROPN	PROPN	bañista
.	PUNCT	PUNCT	.

#Oración 9

A	ADP	ADP	a
pesar	NOUN	NOUN	pesar
de	ADP	ADP	de
sus	DET	DET	su
intentos	NOUN	NOUN	intento
por	ADP	ADP	por
reanimarla	VERB	VERB	reanimar él
,	PUNCT	PUNCT	,
no	ADV	ADV	no
consiguió	VERB	VERB	conseguir
que	SCONJ	SCONJ	que
respirase	VERB	VERB	respirar
.	PUNCT	PUNCT	.

#Oración 10

Uno	PRON	PRON	uno
de	ADP	ADP	de
los	DET	DET	el
allí	ADV	ADV	allí
presentes	ADJ	ADJ	presente
que	PRON	PRON	que
también	ADV	ADV	también
estaba	AUX	AUX	estar
pendiente	ADJ	ADJ	pendiente
llamó	VERB	VERB	llamar
al	ADP	ADP	al
SAMUR	PROPN	PROPN	SAMUR
y	CCONJ	CCONJ	y
se	PRON	PRON	él
la	PRON	PRON	él
llevaron	VERB	VERB	llevar
al	ADP	ADP	al
hospital	NOUN	NOUN	hospital
.	PUNCT	PUNCT	.

#Oración 11

La	DET	DET	el
noche	NOUN	NOUN	noche
terminó	VERB	VERB	terminar
de	ADP	ADP	de
forma	NOUN	NOUN	forma
inesperada	ADJ	ADJ	inesperado
para	ADP	ADP	para
los	DET	DET	el
chicos	NOUN	NOUN	chico
que	PRON	PRON	que
,	PUNCT	PUNCT	,
sin	ADP	ADP	sin
otro	DET	DET	otro
quehacer	NOUN	NOUN	quehacer
,	PUNCT	PUNCT	,
decidieron	VERB	VERB	decidir
regresar	VERB	VERB	regresar
a	ADP	ADP	a
casa	NOUN	NOUN	casa
en	ADP	ADP	en
lugar	NOUN	NOUN	lugar
de	ADP	ADP	de
seguir	VERB	VERB	seguir
divirtiéndose	VERB	VERB	divertir él
.	PUNCT	PUNCT	.

Imágenes de la salida del texto en español en formato *CONLL* dada por FreeLing

#Oración 1

▼ CONLL format

```
1  La          el          DA0FS0  DA  - -      - - - - - -
2  noche       noche       NCFS000 NC  - -      - - - - - -
3  de          de          SP      SP  - -      - - - - - -
4  las         el          DA0FP0  DA  - -      - - - - - -
5  hogueras    hoguera     NCFP000 NC  - -      - - - - - -
6  ,           ,           Fc      Fc  - -      - - - - - -
7  Santiago    santiago    NP00G00 NP  - B-LOC  - - - - - -
8  ,           ,           Fc      Fc  - -      - - - - - -
9  un          uno         DI0MS0  DI  - -      - - - - - -
10 chico       chico       NCMS000 NC  - -      - - - - - -
11 corriente   corriente   AQ0CS00 AQ  - -      - - - - - -
12 ,           ,           Fc      Fc  - -      - - - - - -
13 fue         ser         VSIS3S0 VSI - -      - - - - - -
14 a           a           SP      SP  - -      - - - - - -
15 la          el          DA0FS0  DA  - -      - - - - - -
16 playa       playa       NCFS000 NC  - -      - - - - - -
17 a           a           SP      SP  - -      - - - - - -
18 celebrar    celebrar    VMN0000 VMN - -      - - - - - -
19 su          su          DP3CSN  DP  - -      - - - - - -
20 nuevo       nuevo       AQ0MS00 AQ  - -      - - - - - -
21 logro       logro       NCMS000 NC  - -      - - - - - -
22 :           :           Fd      Fd  - -      - - - - - -
23 había       haber       VAII3S0 VAI - -      - - - - - -
24 terminado   terminar     VMP00SM VMP - -      - - - - - -
25 la          el          DA0FS0  DA  - -      - - - - - -
26 carrera     carrera     NCFS000 NC  - -      - - - - - -
27 de          de          SP      SP  - -      - - - - - -
28 el          el          DA0MS0  DA  - -      - - - - - -
29 todo        todo        PI0MS00 PI  - -      - - - - - -
30 .           .           Fp      Fp  - -      - - - - - -
```

#Oración 2

▼ CONLL format

```
1  De_repente de_repente RG      RG  - - - - - - - -
2  ,          ,          Fc      Fc  - - - - - - - -
3  se         se         P00CN00 P0  - - - - - - - -
4  dio_cuenta dar_cuenta VMIS3S0 VMI - - - - - - - -
5  de         de         SP      SP  - - - - - - - -
6  que        que        PR0CN00 PR  - - - - - - - -
7  había      haber      VAII3S0 VAI - - - - - - - -
8  olvidado   olvidar    VMP00SM VMP - - - - - - - -
9  el         el         DA0MS0  DA  - - - - - - - -
10 papel      papel      NCMS000 NC  - - - - - - - -
11 con        con        SP      SP  - - - - - - - -
12 su         su         DP3CSN  DP  - - - - - - - -
13 deseo      deseo      NCMS000 NC  - - - - - - - -
14 en         en         SP      SP  - - - - - - - -
15 su         su         DP3CSN  DP  - - - - - - - -
16 cuarto     cuarto     NCMS000 NC  - - - - - - - -
17 .          .          Fp      Fp  - - - - - - - -
```

#Oración 3

▼ CONLL format

```
1  No         no       RN       RN  - - - - - - - -
2  quiso      querer   VMIS3S0  VMI - - - - - - - -
3  volver     volver   VMN0000  VMN - - - - - - - -
4  ,          ,        Fc       Fc  - - - - - - - -
5  pues       pues     CS       CS  - - - - - - - -
6  eran       ser      VSII3P0  VSI - - - - - - - -
7  las        el       DA0FP0   DA  - - - - - - - -
8  21:15      21:15    Z        Z   - - - - - - - -
9  y          y        CC       CC  - - - - - - - -
10 subir      subir    VMN0000  VMN - - - - - - - -
11 hasta      hasta    SP       SP  - - - - - - - -
12 el         el       DA0MS0   DA  - - - - - - - -
13 cuarto     4        AO0MS00  AO  - - - - - - - -
14 piso       piso     NCMS000  NC  - - - - - - - -
15 supondría  suponer  VMIC3S0  VMI - - - - - - - -
16 llegar     llegar   VMN0000  VMN - - - - - - - -
17 tarde      tarde    RG       RG  - - - - - - - -
18 a          a        SP       SP  - - - - - - - -
19 la         el       DA0FS0   DA  - - - - - - - -
20 cita       cita     NCFS000  NC  - - - - - - - -
21 .          .        Fp       Fp  - - - - - - - -
```

#Oración 4

▼ CONLL format

```
1  Una          uno          DI0FS0  DI  - -     - - - - - -
2  vez          vez          NCFS000 NC  - -     - - - - - -
3  en           en           SP      SP  - -     - - - - - -
4  la           el           DA0FS0  DA  - -     - - - - - -
5  playa        playa        NCFS000 NC  - -     - - - - - -
6  ,            ,            Fc      Fc  - -     - - - - - -
7  se           se           P00CN00 P0  - -     - - - - - -
8  reunió       reunir       VMIS3S0 VMI - -     - - - - - -
9  con          con          SP      SP  - -     - - - - - -
10 su           su           DP3CSN  DP  - -     - - - - - -
11 grupo        grupo        NCMS000 NC  - -     - - - - - -
12 de           de           SP      SP  - -     - - - - - -
13 amigos       amigo        NCMP000 NC  - -     - - - - - -
14 ,            ,            Fc      Fc  - -     - - - - - -
15 Los_Halcones los_halcones NP00G00 NP  - B-LOC - - - - - -
16 .            .            Fp      Fp  - -     - - - - - -
```

#Oración 5

▼ CONLL format

```
1  Habían      haber      VAII3P0 VAI - - - - - - - -
2  elegido     elegir     VMP00SM VMP - - - - - - - -
3  este        este       DD0MS0  DD  - - - - - - - -
4  nombre      nombre     NCMS000 NC  - - - - - - - -
5  por         por        SP      SP  - - - - - - - -
6  su          su         DP3CSN  DP  - - - - - - - -
7  amor        amor       NCMS000 NC  - - - - - - - -
8  a           a          SP      SP  - - - - - - - -
9  el          el         DA0MS0  DA  - - - - - - - -
10 atletismo   atletismo  NCMS000 NC  - - - - - - - -
11 ,           ,          Fc      Fc  - - - - - - - -
12 creían      creer      VMII3P0 VMI - - - - - - - -
13 ser         ser        VSN0000 VSN - - - - - - - -
14 los         el         DA0MP0  DA  - - - - - - - -
15 más         más        RG      RG  - - - - - - - -
16 veloces     veloz      AQ0CP00 AQ  - - - - - - - -
17 corredores  corredor   NCMP000 NC  - - - - - - - -
18 .           .          Fp      Fp  - - - - - - - -
```

#Oración 6

▼ CONLL format

```
1  Todos      todo       PI0MP00 PI  - - - - - - - -
2  ellos      ellos      PP3MP00 PP  - - - - - - - -
3  empezaron  empezar    VMIS3P0 VMI - - - - - - - -
4  a          a          SP      SP  - - - - - - - -
5  cantar     cantar     VMN0000 VMN - - - - - - - -
6  ,          ,          Fc      Fc  - - - - - - - -
7  bailar     bailar     VMN0000 VMN - - - - - - - -
8  y          y          CC      CC  - - - - - - - -
9  beber      beber      VMN0000 VMN - - - - - - - -
10 ,          ,          Fc      Fc  - - - - - - - -
11 creyendo   creer      VMG0000 VMG - - - - - - - -
12 que        que        CS      CS  - - - - - - - -
13 nada       nada       PI0CS00 PI  - - - - - - - -
14 pasaba     pasar      VMII3S0 VMI - - - - - - - -
15 a          a          SP      SP  - - - - - - - -
16 su         su         DP3CSN  DP  - - - - - - - -
17 alrededor  alrededor  NCMS000 NC  - - - - - - - -
18 .          .          Fp      Fp  - - - - - - - -
```

#Oración 7

```
1  Sin       sin       SP       SP  - -      - - - - - - -
2  embargo   embargo   NCMS000  NC  - -      - - - - - - -
3  ,         ,         Fc       Fc  - -      - - - - - - -
4  Santiago  santiago  NP00SP0  NP  - B-PER  - - - - - - -
5  estaba    estar     VMII3S0  VMI - -      - - - - - - -
6  a         a         SP       SP  - -      - - - - - - -
7  el        el        DA0MS0   DA  - -      - - - - - - -
8  loro      loro      NCMS000  NC  - -      - - - - - - -
9  de        de        SP       SP  - -      - - - - - - -
10 todo      todo      DI0MS0   DI  - -      - - - - - - -
11 lo        el        DA00S0   DA  - -      - - - - - - -
12 que       que       PR0CN00  PR  - -      - - - - - - -
13 sucedía   suceder   VMII3S0  VMI - -      - - - - - - -
14 ,         ,         Fc       Fc  - -      - - - - - - -
15 y         y         CC       CC  - -      - - - - - - -
16 observó   observar  VMIS3S0  VMI - -      - - - - - - -
17 que       que       CS       CS  - -      - - - - - - -
18 algo      algo      PI0CS00  PI  - -      - - - - - - -
19 no        no        RN       RN  - -      - - - - - - -
20 iba       ir        VMII3S0  VMI - -      - - - - - - -
21 bien      bien      RG       RG  - -      - - - - - - -
22 .         .         Fp       Fp  - -      - - - - - - -
23 “         “         Fra      Fra - -      - - - - - - -
24 -         -         Fg       Fg  - -      - - - - - - -
25 Hay       haber     VMIP3S0  VMI - -      - - - - - - -
26 una       uno       DI0FS0   DI  - -      - - - - - - -
27 chica     chico     NCFS000  NC  - -      - - - - - - -
28 en        en        SP       SP  - -      - - - - - - -
29 el        el        DA0MS0   DA  - -      - - - - - - -
30 agua      agua      NCCS000  NC  - -      - - - - - - -
31 ,         ,         Fc       Fc  - -      - - - - - - -
32 pero      pero      CC       CC  - -      - - - - - - -
33 no        no        RN       RN  - -      - - - - - - -
34 nada      nada      PI0CS00  PI  - -      - - - - - - -
35 ,         ,         Fc       Fc  - -      - - - - - - -
36 algo      algo      PI0CS00  PI  - -      - - - - - - -
37 le        le        PP3CSD0  PP  - -      - - - - - - -
38 pasa      pasar     VMIP3S0  VMI - -      - - - - - - -
39 -         -         Fg       Fg  - -      - - - - - - -
40 ”         ”         Frc      Frc - -      - - - - - - -
41 dijo      decir     VMIS3S0  VMI - -      - - - - - - -
42 .         .         Fp       Fp  - -      - - - - - - -
```

#Oración 8

▼ CONLL format

```
1  Sintiendo   sentir     VMG0000 VMG - - - - - - - -
2  se          se         PP3CN00 PP  - - - - - - - -
3  el          el         DA0MS0  DA  - - - - - - - -
4  más         más        RG      RG  - - - - - - - -
5  valiente    valiente   AQ0CS00 AQ  - - - - - - - -
6  de          de         SP      SP  - - - - - - - -
7  el          el         DA0MS0  DA  - - - - - - - -
8  lugar       lugar      NCMS000 NC  - - - - - - - -
9  ,           ,          Fc      Fc  - - - - - - - -
10 se          se         P00CN00 P0  - - - - - - - -
11 quitó       quitar     VMIS3S0 VMI - - - - - - - -
12 la          el         DA0FS0  DA  - - - - - - - -
13 ropa        ropa       NCFS000 NC  - - - - - - - -
14 en          en         SP      SP  - - - - - - - -
15 un          uno        DI0MS0  DI  - - - - - - - -
16 santiamén   santiamén  NCMS000 NC  - - - - - - - -
17 y           y          CC      CC  - - - - - - - -
18 ,           ,          Fc      Fc  - - - - - - - -
19 zambullendo zambullir  VMG0000 VMG - - - - - - - -
20 se          se         PP3CN00 PP  - - - - - - - -
21 en          en         SP      SP  - - - - - - - -
22 el          el         DA0MS0  DA  - - - - - - - -
23 agua        agua       NCCS000 NC  - - - - - - - -
24 ,           ,          Fc      Fc  - - - - - - - -
25 rescató     rescatar   VMIS3S0 VMI - - - - - - - -
26 a           a          SP      SP  - - - - - - - -
27 la          el         DA0FS0  DA  - - - - - - - -
28 bañista     bañista    NCCS000 NC  - - - - - - - -
29 .           .          Fp      Fp  - - - - - - - -
```

#Oración 9

▼ CONLL format

```
1  A_pesar_de a_pesar_de SP      SP  - - - - - - - -
2  sus        su         DP3CPN  DP  - - - - - - - -
3  intentos   intento    NCMP000 NC  - - - - - - - -
4  por        por        SP      SP  - - - - - - - -
5  reanimar   reanimar   VMN0000 VMN - - - - - - - -
6  la         lo         PP3FSA0 PP  - - - - - - - -
7  ,          ,          Fc      Fc  - - - - - - - -
8  no         no         RN      RN  - - - - - - - -
9  consiguió  conseguir  VMIS3S0 VMI - - - - - - - -
10 que        que        CS      CS  - - - - - - - -
11 respirase  respirar   VMSI3S0 VMS - - - - - - - -
12 .          .          Fp      Fp  - - - - - - - -
```

#Oración 10

▼ CONLL format

```
1  Uno        uno        PI0MS00 PI  - -      - - - - - -
2  de         de         SP      SP  - -      - - - - - -
3  los        el         DA0MP0  DA  - -      - - - - - -
4  allí       allí       RG      RG  - -      - - - - - -
5  presentes  presente   AQ0CP00 AQ  - -      - - - - - -
6  que        que        PR0CN00 PR  - -      - - - - - -
7  también    también    RG      RG  - -      - - - - - -
8  estaba     estar      VMII3S0 VMI - -      - - - - - -
9  pendiente  pendiente  AQ0CS00 AQ  - -      - - - - - -
10 llamó      llamar     VMIS3S0 VMI - -      - - - - - -
11 a          a          SP      SP  - -      - - - - - -
12 el         el         DA0MS0  DA  - -      - - - - - -
13 SAMUR      samur      NP00O00 NP  - B-ORG  - - - - - -
14 y          y          CC      CC  - -      - - - - - -
15 se         se         P00CN00 P0  - -      - - - - - -
16 la         lo         PP3FSA0 PP  - -      - - - - - -
17 llevaron   llevar     VMIS3P0 VMI - -      - - - - - -
18 a          a          SP      SP  - -      - - - - - -
19 el         el         DA0MS0  DA  - -      - - - - - -
20 hospital   hospital   NCMS000 NC  - -      - - - - - -
21 .          .          Fp      Fp  - -      - - - - - -
```

#Oración 11

▼ CONLL format

```
1  La           el           DA0FS0  DA  - - - - - - - -
2  noche        noche        NCFS000 NC  - - - - - - - -
3  terminó      terminar     VMIS3S0 VMI - - - - - - - -
4  de           de           SP      SP  - - - - - - - -
5  forma        forma        NCFS000 NC  - - - - - - - -
6  inesperada   inesperado   AQ0FS00 AQ  - - - - - - - -
7  para         para         SP      SP  - - - - - - - -
8  los          el           DA0MP0  DA  - - - - - - - -
9  chicos       chico        NCMP000 NC  - - - - - - - -
10 que          que          PR0CN00 PR  - - - - - - - -
11 ,            ,            Fc      Fc  - - - - - - - -
12 sin          sin          SP      SP  - - - - - - - -
13 otro         otro         DI0MS0  DI  - - - - - - - -
14 quehacer     quehacer     NCMS000 NC  - - - - - - - -
15 ,            ,            Fc      Fc  - - - - - - - -
16 decidieron   decidir      VMIS3P0 VMI - - - - - - - -
17 regresar     regresar     VMN0000 VMN - - - - - - - -
18 a            a            SP      SP  - - - - - - - -
19 casa         casa         NCFS000 NC  - - - - - - - -
20 en_lugar_de  en_lugar_de  SP      SP  - - - - - - - -
21 seguir       seguir       VMN0000 VMN - - - - - - - -
22 divirtiendo  divertir     VMG0000 VMG - - - - - - - -
23 se           se           PP3CN00 PP  - - - - - - - -
24 .            .            Fp      Fp  - - - - - - - -
```

Imagen de la salida del texto en italiano en formato *txt* dada por SpaCy

Token	POS	Etiqueta Detallada	Lema	TAG
Di	ADP	E	di	E
colpo	NOUN	S	colpo	S
,	PUNCT	FF	,	FF
l'	DET	RD	il	RD
amica	NOUN	S	amica	S
di	ADP	E	di	E
Ilda	PROPN	SP	Ilda	SP
si	PRON	PC	si	PC
accorse	VERB	V	accorgere	V
che	SCONJ	CS	che	CS
il	DET	RD	il	RD
suo	DET	AP	suo	AP
collega	NOUN	S	collega	S
aveva	AUX	VA	avere	VA
dimenticato	VERB	V	dimenticare	V
di	ADP	E	di	E
inviare	VERB	V	inviare	V
l'	DET	RD	il	RD
email	NOUN	S	email	S
al	ADP	E_RD	a il	E_RD
SUEM	PROPN	SP	SUEM	SP
del	ADP	E_RD	di il	E_RD
Piemonte	PROPN	SP	Piemonte	SP
:	PUNCT	FC	:	FC
«	PUNCT	FB	«	FB
Piano	PROPN	SP	piano	SP
,	PUNCT	FF	,	FF
c'	PRON	PC	ci	PC
è	VERB	V	essere	V
qualcosa	PRON	PI	qualcosa	PI
che	SCONJ	CS	che	CS
non	ADV	BN	non	BN
va	VERB	V	andare	V
»	PUNCT	FB	»	FB
,	PUNCT	FF	,	FF
si	PRON	PC	si	PC
disse	VERB	V	dire	V
.	PUNCT	FS	.	FS

Imagen de la salida del texto en italiano en formato *CONLL* dada por FreeLing

▼ CONLL format

```
1  Di                  di                  SPS00    SP  - - - - - - - -
2  colpo               colpo               NCMS000  NC  - - - - - - - -
3  ,                   ,                   Fc       Fc  - - - - - - - -
4  l'                  il                  DA0MS0   DA  - - - - - - - -
5  amica               amica               NCFS000  NC  - - - - - - - -
6  di                  di                  SPS00    SP  - - - - - - - -
7  Ilda                ilda                NP00000  NP  - - - - - - - -
8  si                  si                  PP3CN00  PP  - - - - - - - -
9  accorse             accorgere           VMIS3S0  VMI - - - - - - - -
10 che                 che                 CS       CS  - - - - - - - -
11 il                  il                  DA0MS0   DA  - - - - - - - -
12 suo                 suo                 AP0MS3S  AP  - - - - - - - -
13 collega             collega             NCCS000  NC  - - - - - - - -
14 aveva               avere               VAII3S0  VAI - - - - - - - -
15 dimenticato         dimenticare         VMP00SM  VMP - - - - - - - -
16 di                  di                  SPS00    SP  - - - - - - - -
17 inviare             inviare             VMN0000  VMN - - - - - - - -
18 l'                  il                  DA0MS0   DA  - - - - - - - -
19 email               email               NCMN000  NC  - - - - - - - -
20 al                  al                  SPCMS    SP  - - - - - - - -
21 SUEM_del_Piemonte   suem_del_piemonte   NP00000  NP  - - - - - - - -
22 :                   :                   Fd       Fd  - - - - - - - -
23 «                   «                   Fra      Fra - - - - - - - -
24 Piano               piano               NCMN000  NC  - - - - - - - -
25 ,                   ,                   Fc       Fc  - - - - - - - -
26 c'                  c'                  RG       RG  - - - - - - - -
27 è                   essere              VMIP3S0  VMI - - - - - - - -
28 qualcosa            qualcosa            PI0FS00  PI  - - - - - - - -
29 che                 che                 PT00000  PT  - - - - - - - -
30 non                 non                 RG       RG  - - - - - - - -
31 va                  andare              VMIP3S0  VMI - - - - - - - -
32 »                   »                   Frc      Frc - - - - - - - -
33 ,                   ,                   Fc       Fc  - - - - - - - -
34 si                  si                  PP3CN00  PP  - - - - - - - -
35 disse               dire                VMIS3S0  VMI - - - - - - - -
36 .                   .                   Fp       Fp  - - - - - - - -
```