

**Lucía Sanz-Valdivieso
Belén López Arroyo**



**Generador de
Notas de Cata de Aceite
de Oliva 1.0:**

**Lingüística
Aplicada a la
Internacionalización
del Aceite de Oliva**

**IX Premio de Investigación
del Consejo Económico y Social
de la provincia de Jaén**

MENCIÓN ESPECIAL

**GENERADOR DE NOTAS DE
CATA DE ACEITE DE OLIVA 1.0:
LINGÜÍSTICA APLICADA A LA
INTERNACIONALIZACIÓN DEL
ACEITE DE OLIVA**

Lucía Sanz-Valdivieso
Belén López Arroyo

**GENERADOR DE NOTAS DE
CATA DE ACEITE DE OLIVA 1.0:
LINGÜÍSTICA APLICADA A LA
INTERNACIONALIZACIÓN DEL
ACEITE DE OLIVA**



Consejo Económico y Social
de la provincia de Jaén

Edita: DIPUTACIÓN PROVINCIAL DE JAÉN
Consejo Económico y Social de la provincia de Jaén

© Del autor
© De la presente edición: Diputación Provincial de Jaén

I.S.B.N.: 978-84-09-39668-9
Depósito Legal: J. 166 - 2022

Impreso en España | Unión Europea

Agradecimientos

Al Consejo Económico y Social de la provincia de Jaén, por brindarnos la oportunidad de compartir nuestra investigación con los sectores involucrados a través del IX Concurso de Investigación y por su reconocimiento a través de una Mención Especial en el mismo.

A la Diputación de Jaén por prestarse a publicar nuestro trabajo de investigación de forma que pueda llegar a los interesados.

Al Dr. Francisco José Palacios Burgos (Universidad Europea Miguel de Cervantes), por su generosidad y profesionalidad al desarrollar desinteresadamente las tareas de diseño informático necesarias para materializar la aplicación del estudio llevado a cabo.

A la Universidad de Valladolid, por la concesión del proyecto *Tech-writing tools: Herramientas para la ayuda a la escritura de textos especializados en lengua inglesa* (PROYEMER-2021-35), en el marco del cual se ha desarrollado el presente estudio.

Resumen

El aceite de oliva español, en general, y el jiennense, en particular, han recibido atención internacional debido a sus excepcionales y versátiles características. Este producto ha sido objeto de estudio desde las perspectivas fisicoquímicas, biomédicas, agrícolas y económicas. Sin embargo, el presente proyecto¹ pretende acercarse a este *oro líquido* a través de la lingüística de corpus centrándose en las notas de cata en lengua inglesa. El propósito es describir el género textual, su lenguaje para fines específicos propio y sus características a todos los niveles lingüísticos, para aplicar los resultados del análisis al diseño de un generador semiautomático de notas de cata de aceite de oliva en lengua inglesa (Generador de Notas de Cata de Aceite de Oliva 1.0). Este sistema, fiable, económico y de fácil uso, está pensado para ayudar a los profesionales de habla española, expertos y semi-expertos, legos en el uso del inglés especializado, a comunicar su producto a un mercado mayor, y contribuir así a la internacionalización económica y turística de este icónico pilar de nuestro patrimonio intangible.

PALABRAS CLAVE: género lingüístico, lenguajes para fines específicos, notas de cata de aceite de oliva, lingüística de corpus, lingüística aplicada, nuevas tecnologías, internacionalización.

Abstract

Olive oil has been widely studied from physic-chemical, bio-medical, agricultural, and economic perspectives. However, there are not very many studies from the linguistic perspective; this project will approach this *liquid gold* focusing on olive oil tasting notes in the English language from a functional approach and using a linguistic corpus. We aim at describing the genre, its particular LSP, and its micro-linguistics features. The results from the analysis will be used to develop a semi-automatic tool that generates olive oil tasting notes in English, designed to help Spanish-speaking olive oil professionals, experts and semi-experts, lay in the specialized English language, to communicate to a wider market, and therefore contribute to the economic and touristic internationalization of this flagship of our heritage.

KEYWORDS: genre analysis, language for specific purposes, olive oil tasting notes, corpus linguistics, applied linguistics, new technologies, internationalization.

¹ Estudio desarrollado dentro del proyecto *Tech-writing tools: Herramientas para la ayuda a la escritura de textos especializados en lengua inglesa* (PROYEMER-2021-35) concedido por la Universidad de Valladolid.

1. Introducción

La gastronomía y el turismo lideran el atractivo internacional de nuestro país. España está a la cabeza de la exportación mundial de aceite de oliva, siendo la provincia de Jaén la mayor productora oleícola del país y del mundo. En el año 2019, último año del que se disponen datos, Jaén contaba con 586.921 hectáreas de superficie de plantación de olivar de aceituna de almazara y con 2.084 nuevas plantaciones desde el año anterior. La provincia produjo, en 2019, 396.096 toneladas de aceite de oliva, doblando con creces a la siguiente mayor productora andaluza, Córdoba, que contó con una producción de 195.148 toneladas. De hecho, Jaén aporta el 44,21% de la producción de aceite de oliva de Andalucía y el 35,08% de la producción nacional (Gobierno de España, 2019; Junta de Andalucía, 2019). Además, el aceite de oliva es el segundo producto que más exporta la provincia de Jaén, con 248 millones de euros; lo que supone un 19,90% de las exportaciones provinciales, comportando un aumento en la exportación de este producto en 2019 de un 20,40% en relación al año 2018 (CES, 2020, p. 46).

El atractivo internacional de este producto jiennense no se reduce a la exportación, sino que el aspecto turístico y cultural alrededor del aceite de oliva también reciben interés de visitantes extranjeros. En la última década, se ha acuñado un nuevo vocablo para referirse a las actividades turísticas y culturales desarrolladas alrededor del aceite de oliva: el oleoturismo, que, según Aceites de Oliva de España (2017), abarca desde visitas guiadas a almazaras hasta paseos educativos por los olivares locales, pasando por las catas de aceite de oliva. Es en la cata en la actividad oleícola específica en la que se centrará el presente estudio de investigación, aunque el resultado (Generador de Notas de Cata de Aceite de Oliva 1.0) puede ser aplicado no sólo en las catas turísticas, sino también en etiquetado y marketing internacional del producto, entre otros. La cata de aceite de oliva, objeto de estudio del presente trabajo en su dimensión lingüística y comunicativa, se refiere a la experiencia de la apariencia, aroma y sabor del producto en primera persona. Orgaz Agüera et al. (2017) argumentan que una de las mayores razones por las que el turista se ve atraído a la cata de aceite es la búsqueda de la individua-

lidad de la experiencia, y, por ende, de exclusividad y superioridad social. En este sentido, y, en parte, gracias a la promoción de su producto turístico OleotourJaén, la provincia vuelve a acaparar la actividad turística y cultural que tiene al aceite como protagonista: en 2019, Jaén recibió a 531.700 viajeros, de los cuales 72.468 declararon ser residentes en el extranjero, es decir, 13,63% del total (Fundación, 2021, p. 32). Del total de viajeros ese mismo año, 24.195 visitaron almazaras jienenses adheridas a OleotourJaén, y se contabilizaron 17.352 visitas a los Museos y Centros de Interpretación de OleotourJaén, según comunicaciones telefónicas mantenidas con la Fundación Estrategias para el Desarrollo de la Provincia de Jaén y Diputación de Jaén. No hay datos disponibles sobre la nacionalidad de estos oleoturistas en el año 2019, pero, según La Vanguardia, en 2018, el 70% de los visitantes a almazaras jienenses provenían del extranjero, especialmente de Estados Unidos y países de Europa central.

En este contexto, y con el propósito de desarrollar el potencial demostrado a través de las cifras anteriores, el presente estudio pretende trabajar en la línea de algunas de las consideraciones y recomendaciones incluidas en la *Memoria 2019 sobre la situación socioeconómica y laboral de la provincia de Jaén* publicada por el Consejo Económico y Social (CES) de la Provincia de Jaén. En concreto, el Generador creado está destinado a cumplir la recomendación 29 del Capítulo 2, *Tejido Empresarial*, al tratar de “apoyar la internacionalización mediante la cooperación interempresarial, la formación de profesionales expertos y el diseño de plataformas tecnológicas” (2020, p. 85). Asimismo, la introducción de la herramienta de ayuda a la escritura supone “intensificar las acciones de comunicación con la finalidad de que siga aumentando la demanda mundial de aceites de oliva vírgenes” (2020, p. 88). Igualmente, contar con una herramienta como el Generador aumenta las posibilidades de realización de “acciones de comunicación en forma de ‘catas’, destinadas, tanto a los consumidores potenciales y actuales, como a los propios olivares provinciales, al objeto de lograr un mayor conocimiento de los distintos aceites de oliva y de sus características diferenciales, así como un mayor reconocimiento y valoración de los esfuerzos y el coste de la calidad” (2020, p. 88). En última instancia, este trabajo pretende también conseguir una transferencia de conocimiento desde la Universidad y la acción investigadora hacia las empresas y el sector productivo de aceite de oliva de Jaén.

Más concretamente, el Generador está diseñado con el objetivo último de servir de herramienta que abra el sector jienense al mercado mundial y globalizado. De hecho, el CES de la Provincia de Jaén, en sus recomendaciones relativas al turismo (Capítulo 6) aboga por la promoción internacional del patrimonio a través de la traducción y la digitalización de los recursos de cara al turismo internacional (2020, p. 94). En este sentido, la Fundación Estrategias para el Desarrollo de la Provincia de Jaén ha publicado el *II Plan Estratégico de la provincia de Jaén*, donde reconoce el “[i]mportante potencial de crecimiento en el turismo

extranjero” (Fundación, 2020, p. 103). Además, el oleoturismo ha recibido, en los últimos años, interés y financiación también por parte de las entidades públicas (Fundación, 2020, p. 99). Precisamente, uno de los objetivos de la Estrategia 5 *Jaén, paraíso interior*, parte de la “elevada potencialidad de la cultura del aceite y el olivo: ‘Oleoturismo’” (2020, p. 103), para lo cual se propone como reto “[c]onsolidar y mantener las inversiones puestas en marcha e incrementar los retornos de las mismas” (2020, p. 143). Asimismo, dentro de la Estrategia 6, *Jaén, centro mundial del aceite de oliva*, se propone el reto “desarrollar el oleoturismo” (2020, p. 114). En este contexto, la presente investigación pretende contribuir con las acciones previstas en el Proyecto 6.1. de los Proyectos Estructurantes del *II Plan*, que tiene como uno de sus objetivos el desarrollo de un “Plan de Internacionalización del Sector Oleícola” (2020, p. 224).

En un mundo globalizado de estas características, la comunicación internacional entre profesionales es importante en todos los niveles y se produce, principalmente, en lengua inglesa. Sin embargo, muchos expertos españoles no tienen la competencia lingüística necesaria para producir textos especializados de su campo en inglés. Por ese motivo, es necesario desarrollar herramientas específicas que les ayuden en la comunicación profesional, favoreciendo, al mismo tiempo, la internacionalización de sus productos. El presente estudio pretende desarrollar una herramienta de escritura semi-automática de notas de cata de aceite de oliva en lengua inglesa que ayude a los profesionales de habla española, expertos y semi-expertos, legos en el uso del inglés especializado, a poner en valor su producto en un mercado mayor, y contribuir así a la internacionalización económica, cultural y turística de este icónico pilar de nuestro patrimonio intangible. Para ello, y como paso previo, es necesario un análisis lingüístico integral de este género textual (léxico, sintáctico-gramatical, semántico y discursivo) con el propósito de describir el género, su lenguaje para fines específicos (LFE) y sus características a todos los niveles lingüísticos, para aplicar los resultados del análisis al diseño de la mencionada herramienta de ayuda a la escritura.

Desde el punto de vista lingüístico, los antecedentes que motivan este estudio son diversos. Este trabajo se enmarca dentro de las áreas de Análisis de Género y Análisis de Lenguajes para Fines Específicos (LFE), aplicados a la LFE particular de la cata de aceite de oliva y al género de la nota de cata de aceite de oliva en lengua inglesa. La motivación de este trabajo de investigación es, como se ha mencionado anteriormente, el creciente interés internacional en el aceite de oliva español y, sobre todo, jiennense. En este sentido, también hallamos, dentro de los últimos trabajos enmarcados en el análisis de género y de LFE, una falta de atención científica al objeto de nuestro estudio. Hasta donde sabemos, no se ha realizado ningún trabajo que estudie el LFE particular de la cata de aceite de oliva en lengua inglesa, a diferencia del LFE del aceite de oliva en general, que aporte un análisis multidimensional y completo de dicho LFE y su género en inglés. Por

otra parte, el mencionado tipo de ayuda a la escritura, similar al que este trabajo tiene como objetivo final, ya ha sido desarrollado por el Grupo de Investigación ACTRES en otras ocasiones, habiendo creado generadores de *abstract* científicos, descripciones de productos electrónicos, informes técnicos, y notas de cata de vino, entre otros. Dada la utilidad y relevancia de dichos proyectos, este trabajo pretende adaptar dicha aplicación de la lingüística y las nuevas tecnologías al LFE de las notas de cata de aceite de oliva.

Para alcanzar estos objetivos, se realizará una descripción del género textual a partir del análisis de un corpus lingüístico de notas de cata de aceite de oliva en lengua inglesa compilado exclusivamente para este fin. El examen y categorización de las muestras permitirá trazar un esquema de las características del género a todos los niveles lingüísticos: léxico, sintáctico-gramatical, semántico y discursivo. A partir de esta información, se podrá desarrollar un modelo lingüístico en la que se basará el Generador, incluyendo las denominadas líneas modelo (estructuras gramaticales y sintácticas) y descriptores (términos y unidades fraseológicas). Estos resultados serán necesarios para la composición de la arquitectura de la herramienta digital destinada a facilitar la redacción de este tipo de textos por parte de profesionales del aceite de oliva hablantes de español, de modo que los textos resultantes sean aceptables dentro de la comunidad discursiva meta. En definitiva, el Generador creado supone un esfuerzo realista, fiable, económico y de uso intuitivo en aras de la promoción internacional del aceite de oliva jiennense.

Habiendo introducido el tema y el objetivo, el presente trabajo continúa con una revisión del estado de la cuestión (sección 2) en cuanto al análisis de los lenguajes para fines específicos y géneros textuales (sección 2.1 y 2.2) y a su aplicación al estudio del lenguaje para fines específicos relativo al aceite de oliva y a su cata (sección 2.3). En la sección 3 se describirá el material empleado (sección 3.1) y la metodología seguida para la consecución del objetivo (sección 3.2), cuyos resultados (sección 4) a todos los niveles lingüísticos se presentan en las secciones 4.1, 4.2, 4.3 y 4.4, con especial énfasis en su aplicación en el desarrollo de la herramienta de ayuda a la escritura (sección 4.5). Finalmente, describiremos el Generador, su construcción y su modo de empleo en la sección 5, para concluir el trabajo en la sección 6.

2. Estado de la cuestión

2.1. Lenguajes para Fines Específicos (LFE)

Los lenguajes para fines específicos (LFE) son “variedades funcionales de la lengua que se han desarrollado como instrumentos de comunicación en un determinado ámbito científico o profesional, [basadas] en las estructuras fonéticas, gramaticales y léxicas de la lengua natural y [caracterizadas] por la formación de una terminología propia, así como por la preferencia por ciertas formas gramaticales y discursivas” (Calvi et al., 2009, p. 15). Para que una variedad de la lengua se considere un LFE, debe cumplir los siguientes requisitos (Dudley-Evans, 1998, p. 4): contar con una metodología, habilidades, discurso y géneros propios de los usuarios del mismo y desempeñar la función práctica para la cual existe. En este sentido, se puede afirmar que el lenguaje de las notas de cata de aceite de oliva constituye un LFE por derecho propio, ya que su uso está restringido a unos usuarios concretos (catadores, ya sean profesionales o no) que siguen una metodología particular (las distintas fases de la cata) y que tienen unas habilidades concretas (análisis sensorial) con el propósito de expresar con palabras los atributos organolépticos de una muestra de aceite en una forma determinada (el género de la nota de cata).

Los LFE cuentan con dos dimensiones: la horizontal, que incluye el léxico y el componente temático o contenido, y la vertical, donde se diferencian distintos subsectores de especialización (Calvi et al., 2009, pp. 22-23). El LFE de la cata de aceite de oliva podría clasificarse, en la dimensión vertical, como un subsector del LFE del aceite de oliva en general (incluyendo aspectos agrícolas, de envasado y comercialización, etc.). En cuanto a los distintos grados de especialización (Calvi et al., 2009, p. 30), el LFE de la cata de aceite de oliva incluye situaciones comunicativas que abarcan casi todas las combinaciones posibles entre usuarios expertos, semi-expertos y no expertos: por ejemplo, un catador profesional (experto) puede producir notas de cata tanto para un concurso de aceites de oliva (experto) como para una página web orientada al consumidor (no-experto). Es de vital impor-

tancia considerar que, dentro de una comunidad discursiva, haya usuarios con distinto grado de especialización en la materia, ya que así se refleja más fielmente la realidad comunicativa del sector.

Dentro del eje horizontal de las LFE (Calvi et al., 2009), los elementos más relevantes y característicos son los términos y las unidades fraseológicas, así como el contenido semántico de los mismos. Un término en un LFE es un signo lingüístico distintivo y significativo que tiene aspectos sistemáticos (formal, semántico y funcional) y pragmáticos (Cabré, 1999, p. 80). El rasgo que diferencia un término de una palabra del lenguaje general es que el término se utiliza para designar conceptos que pertenecen a actividades o disciplinas especializadas (Cabré, 1999, p. 81). Esto incluye tanto términos *per se*, como *fruitiness* [afrutado] en nuestro corpus, como términos subtécnicos, como *body* [cuerpo]; un término subtécnico es un vocablo compartido por un LFE y el lenguaje general (Marín Pérez, 2016, p. 80).

Además de la terminología, en el eje horizontal del LFE (Calvi et al., 2009) también se enmarcan las unidades fraseológicas. Las unidades fraseológicas son grupos de palabras lexicalizados, fijos, reproductibles y sintácticamente estables (Gläser, 1994, p. 46). Además, se diferencian de un grupo de palabras cualesquiera por que suelen ser recurrentes y se producen dentro de un LFE particular (Méndez Cendón, 2009, p. 170) pero no están idiomatizadas ni tienen connotaciones estilísticas ni expresivas (Gläser, 1994, p. 51). Aunque hay una amplia literatura sobre la clasificación de las unidades fraseológicas, el presente estudio tiene un carácter eminentemente práctico y se decanta por la clasificación de Gläser (1994) y la posterior modificación de la misma de Roberts (1998) dada su funcionalidad: las unidades fraseológicas se clasifican en nominaciones, operadores y colocaciones. Las nominaciones son unidades fraseológicas que funcionan sintáctica y semánticamente como si fueran una sola palabra (Gläser, 1994, p. 48), como *virgin olive oil* [aceite de oliva virgen] o *beautiful gold color* [hermoso color oro]. Las unidades fraseológicas denominadas operadores son grupos funcionales que denotan relaciones entre fenómenos (Gläser, 1994, p. 48), como *with touches of* [con toques de] o *embellished by* [embellecido por]. Por último, las colocaciones son combinaciones habituales de palabras que no tienen por qué estar totalmente fijadas y que cuyos componentes siguen siendo elementos diferenciados, como *to open with* [abrir con] en nuestro corpus; aunque puedan llegar a lexicalizarse con el tiempo (Roberts, 1998, p. 65).

Dentro del eje horizontal estudiamos, además de la forma de términos y unidades fraseológicas, su componente temático o contenido. Un campo semántico se define como un constructo teórico que agrupa palabras que tienen en común algún tipo de concepto mental de forma más o menos específica (Garside, Leech & McEnery, 1997, p. 54). Además, aplicaremos un análisis de contenido (Garside,

Leech & McEnery, 1997, p. 64) a los resultados, entendido como la operacionalización de una teoría interpretativa, a saber, la teoría de Paradis (2005, 2010) basada en la lingüística cognitiva. En este sentido, nos centraremos en analizar la naturaleza ontológica del género y el LFE de la nota de cata, es decir, si son objetivamente producibles e interpretables o, por el contrario, son predominantemente inestables semánticamente entre individuos.

2.2. Género textual

Por otra parte, el eje vertical dispuesto por Calvi et al. (2009) es, como ella misma argumenta, equivalente al término género discursivo acuñado por Swales (1990) y Bhatia (1993, 2004), por lo que el presente trabajo estudiará este eje vertical siguiendo la terminología y teoría del análisis de género. Un género aúna cierta clase de eventos comunicativos, cuyos miembros comparten función comunicativa y son reconocidos por los miembros expertos de la comunidad discursiva en cuestión (Swales, 1990, p. 58). Estos rasgos definen la razón de ser fundamental de un género, que determina la estructura esquemática del discurso e influye en la elección de contenido y estilo (1990, p. 50). Además de compartir función comunicativa, los textos pertenecientes a un mismo género muestran patrones de semejanza en términos de estructura, estilo, contenido y audiencia meta (1990, p. 50). En resumen, si un texto cumple dichas expectativas, será aceptado como prototípico por parte de la comunidad discursiva implicada, por lo que será considerado parte de un género dado (Swales, 1990, p. 58). La importancia de la función comunicativa a la hora de definir un género textual es determinante, ya que una función diferente probablemente entrañe un género distinto (Bhatia, 1993, pp. 49-50). En esencia, podemos describir un género textual como un evento comunicativo reconocible caracterizado por sus funciones comunicativas, identificadas y mutuamente comprendidas por los miembros de la comunidad discursiva en la que ocurren con regularidad (Bhatia, 2004, p. 23). Los géneros están altamente estructurados y convencionalizados, son fácilmente reconocibles por los usuarios del mismo, y su integridad se basa en un conjunto de factores textuales, discursivos y contextuales (Bhatia, 2004, p. 23). En vista de estos hechos, podemos clasificar las notas de cata como un género especializado, ya que cumplen todas las características arriba mencionadas. Por eso, para poder desarrollar ayudas de redacción semi-automática de textos, es imprescindible estudiar el género en cuestión para identificar los patrones léxicos, sintáctico-gramaticales, semánticos y discursivos que lo caracterizan y distinguen, para, más tarde, aplicar esa información en el desarrollo del modelo lingüístico del Generador. Dentro de esta definición de género, el presente estudio se centra en identificar las expectativas de género, es decir, las características que, en conjunto, determinan que un texto dado pertenezca o no a un género particular. El objetivo es poder reproducirlos al diseñar el Generador 1.0, para lo que analizaremos un corpus textual (grupo de textos compilados de

acuerdo a una serie de criterios previamente establecidos – ver apartado 3.1) a todos los niveles lingüísticos. Así, los textos producidos con la herramienta de ayuda a la escritura cumplirán con los requisitos del género y serán válidos dentro de la comunidad discursiva de la cata de aceite de oliva.

Desde un punto de vista más práctico, el género puede estudiarse a partir de un análisis de movimientos retóricos (Swales, 1990) para identificar la estructura retórica y la organización discursiva del mismo. Los movimientos son secciones de texto que desempeñan una función comunicativa específica, pero también contribuyen a la función general del texto. Un movimiento puede contener distintas secciones de información, llamadas pasos retóricos, cuya función es contribuir a la función del movimiento al que pertenecen. Esas distintas funciones contenidas en el texto dan forma a las estructuras esquemáticas del discurso e influyen en su contenido y estilo (Swales, 1990, p. 58) de una forma reconocible por los miembros de la comunidad discursiva. En este contexto, los movimientos y los pasos pueden definirse como unidades funcionales de significado que, juntos, contribuyen a caracterizar los textos pertenecientes a un género. Para estudiarlos, realizaremos un análisis de contenido (Biber, Connor & Upton, 2007, p. 33), para el que no hay unas reglas fijas, dado que se trata de una tarea cognitiva y, por tanto, manual en mayor medida. Para comenzar, debe definirse el panorama general de las funciones de un texto, para luego ir identificando y evaluando las funciones locales de las distintas secciones del texto. El objetivo es hallar temas funcionales y semánticos que sean comunes entre textos a través de los distintos segmentos que desempeñan su función en ellos, de forma que, al agruparlos, se reflejen los pasos y el movimiento al que contribuyen (Biber, Connor & Upton, 2007, p. 33). Además, dado que no en todos los textos pertenecientes a un género se halla la estructura retórica al completo, aplicaremos la distinción de Suter (1993, p. 119) de los distintos tipos de información según su relevancia retórica de acuerdo a su frecuencia. De esta forma, no solo identificaremos la estructura retórica del género, sino que seremos capaces de formular recomendaciones al futuro redactor de notas de cata en cuanto a cómo producir su texto de la forma más prototípica posible.

2.3. Comunicación: aceite de oliva y lenguaje

El género textual y los lenguajes para fines específicos han sido ampliamente analizados, resultando en estudios clave para la disciplina como Swales (1990), Bhatia (1993; 2004), Dudley-Evans (1998) y Suter (1993), entre otros, centrados en la descripción de géneros textuales particulares, como es el caso de Suter (1993) y el *wedding report*. Sin embargo, no existe una referencia enfocada en el lenguaje para fines específicos relativa a la cata de aceite de oliva ni en su género textual más representativo: la nota de cata. Sí encontramos investigaciones sobre la lengua específica del aceite de oliva en general, incluyendo no el lenguaje para

describir el producto final, sino más bien el específico utilizado para hablar sobre las plantaciones de olivar, la cosecha de la aceituna, el prensado de la misma y el almacenamiento y transporte del aceite de oliva. Además de no explorar el lenguaje específico de la cata, ausente en la literatura, los estudios existentes se centran casi exclusivamente en la lengua española.

Los primeros intentos de describir desde una perspectiva académica la lengua utilizada para hablar de cualquier aspecto relativo al aceite de oliva incluyen “El Vocabulario del Olivar en el Sur de Córdoba” (Galeote López, 1992), “Sobre el Léxico del Olivo y la Almazara” (Moya Corral, 1994), y “El Léxico y la Realidad de la Almazara en Murcia” (González Blanco, 1999). Estos trabajos de investigación muestran el aspecto lingüístico del aceite de oliva como resultado marginal de estudios dialécticos, como es el caso de los artículos de Galeote López (1992) y González Blanco (1999). En cuanto al estudio realizado por Moya Corral (1994), se centró en analizar cuestionarios demográficos en los que preguntaba a la población de 24 localidades jiennenses por la producción de aceite de oliva. Igualmente localizados en regiones concretas se desarrollaron estudios más recientes, como “El Léxico del Olivo y la Almazara en Sierra Mágina” (Torres Quesada, 2011).

En la última década, sin embargo, el interés en el lenguaje para fines específicos del aceite de oliva ha crecido y se ha materializado en estudios más comprensivos de la materia: ejemplos de ello son los trabajos “Emergent Neologisms and Lexical Gaps in Specialised Languages” [Neologismos emergentes y vacíos léxicos en lenguajes especializados] (Roldán Vendrell & Fernández Domínguez, 2012), o “La Fraseología Especializada del Área Científica y Socioeconómica del Olivar y el Aceite de Oliva” (Montoro del Arco, 2012). Entre los más relevantes desde el punto de vista de la traducción y la internalización del sector, centrados en producir recursos de traducción multilingües, hallamos “The Terminology of Olive Oil Taste Testing: A Bilingual (Spanish-English and English-Spanish) Glossary” [La terminología de la cata de aceite de oliva: Un glosario bilingüe] (Santa María, 2013); “Las Denominaciones del Aceite de Oliva en los Reglamentos Oficiales: Evolución, Normalización y Problemas Terminológicos” (Montoro del Arco & Roldán Vendrell, 2013a); “Terminología, Normalización y Comunicación: las Categorías del Aceite de Oliva en Español, Inglés y Chino” (Montoro del Arco & Roldán Vendrell, 2013b); “Unidades de Conocimiento Especializado en Textos del Ámbito de la Oleicultura Antigua: Aspectos Sintagmáticos” (Roldán Vendrell, 2007); o *Bases para la Terminología Multilingüe del Aceite de Oliva* (Roldán Vendrell, 2010a) y “Olivaterm: Vocabulario Multilingüe del Aceite de Oliva” (Roldán Vendrell, 2010b); cuyo epítome es el *Diccionario de Términos del Aceite de Oliva (Español-Inglés-Chino)* (Roldán Vendrell, 2013).

Aun así, ninguno de estos avances investiga concretamente el lenguaje para fines específicos y el género de las notas de cata de aceite de oliva en lengua

inglesa; aunque Santa María (2013) trata el tema, lo hace desde una perspectiva puramente terminológica, mientras que el presente trabajo pretende aportar una descripción más comprensiva de la LFE y del género de las notas de cata de aceite de oliva. De hecho, la falta de investigación lingüística de esta materia contrasta radicalmente con la abundancia de estudios sobre las perspectivas químicas, biológicas y médicas, que, a menudo, tratan de explicar los beneficios del aceite de oliva para la salud o, si hablan de la cata, es mayormente desde el punto de vista de la fisiología de la actividad. Además, no hay trabajos sobre la lengua inglesa y el aceite de oliva, lo que añade interés a esta empresa. En resumen, la literatura de la que disponemos a día de hoy sobre este tema muestra un nicho de investigación que, ideal y parcialmente, cubriremos con el presente estudio.

La descripción lingüística del género y el lenguaje para fines específicos de la cata de aceite de oliva se utilizará para crear un modelo lingüístico que sirva de base a un generador semi-automático de notas de cata en inglés (Generador 1.0) destinado a facilitar la redacción de dichos textos a los profesionales hablantes españoles involucrados en el sector. Este esfuerzo por producir herramientas que contribuyan a la internacionalización del patrimonio tiene sus antecedentes en el grupo de investigación ACTRES (Análisis Contrastivo y Traducción Inglés-Español), pioneros en la ideación y materialización de la posibilidad de crear herramientas de este tipo a través de la lingüística, el análisis de corpus, y la ingeniería informática. ACTRES es un grupo de investigación consolidado conformado por miembros de las Universidades de León, Valladolid (CITTAC), País Vasco, Ottawa (Canadá), Bergen (Noruega) y la Universidad Miguel de Cervantes. Hasta la fecha, ACTRES ha creado generadores para géneros textuales como los *abstracts* científicos (GAC), descripciones de productos electrónicos (GDPE), actas de reunión (GARE), informes técnicos (GITEC), informes del director (GEDIRE), informes de auditoría (GInA), manuales de instrucciones para aparatos domésticos (GIT) y notas de cata de vino (FIVI) y fichas técnicas de vino (FITEVI). Estos dos últimos (ACTRES, 2008), de mayor interés para este estudio, se describen como una guía para hablantes nativos de español con un nivel intermedio de inglés que necesiten apoyo redactando las notas de cata de vino. Esta es la necesidad que el presente trabajo considera imperativo y se propone cubrir en el caso del aceite de oliva jiennense y español. El generador FIVI se basa en un corpus de notas de cata de vino etiquetadas retóricamente, y el sistema es capaz de proporcionar opciones lingüísticas y recomendaciones al usuario. Además, cuenta con un glosario bilingüe que permite al usuario personalizar la nota de cata a producir en función de las características de su propio producto. En este sentido, el Generador 1.0 propuesto en este estudio contará con características similares derivadas del análisis de un corpus de notas de cata de aceite de oliva.

3. Material y metodología

3.1. Material

El material empleado en el presente estudio es un corpus lingüístico. Un corpus es una colección de muestras lingüísticas que han sido seleccionadas y catalogadas siguiendo criterios lingüísticos explícitos, de forma que dicha colección pueda utilizarse como modelo de ese lenguaje (Sinclair, 1994, p. 14). Para ello, los criterios seguidos en el diseño del corpus lingüístico son de vital relevancia, y, en este caso, se delimitaron según el propósito del corpus: obtener la información lingüística necesaria para describir empíricamente el género y el lenguaje de la cata de aceite de oliva en lengua inglesa. Además, es necesario que el corpus compilado sea representativo de la variedad de la lengua en cuestión (Sinclair, 1994, p. 14) de forma que los resultados obtenidos de su análisis sean extrapolables como modelo de la totalidad de esa variedad. De esta forma, a partir del análisis de una muestra representativa del lenguaje de las notas de cata de aceite de oliva, se determinarán las características lingüísticas que ha de presentar un texto para ser considerado una nota de cata de aceite de oliva en lengua inglesa y ser reconocido como tal; y dichas características se clasificarán e integrarán en la mencionada aplicación de ayuda a la escritura.

En este sentido, Biber (1993) recalca la importancia de los criterios de diseño del corpus para que resulte representativo, y elabora un listado de decisiones a tomar en el muestreo que determinarán la consecución de la representatividad del corpus (Biber, 1993, p. 245): canal primario, formato, situación (público o personal), destinatario, emisor, nivel de factualidad, propósito y tema. Aplicando estos parámetros a nuestro estudio, podría definirse nuestro corpus como una compilación de textos escritos, públicos, y factuales sobre el tema de los atributos sensoriales del aceite de oliva con el propósito de describirlos y evaluarlos, dirigidos a una pluralidad de destinatarios y redactados por emisores pertenecientes a una demografía variada, como se describirá a continuación.

El factor relativo a los emisores y receptores de un género textual determina en gran medida las características del lenguaje utilizado. En un LFE, los

participantes implicados en la comunicación pueden tener distintos grados de dominio, de forma que podemos hallar comunicación no solo entre expertos, sino también con semi-expertos y no expertos. Por ello, el tipo de emisor se ha tenido en cuenta en el diseño del corpus; sin embargo, en cuanto al destinatario, es difícil determinar un perfil específico del receptor, dado que Internet está abierto a todo tipo de público, aunque los emisores pudiesen tener en mente un tipo de audiencia particular al redactar los textos. En cualquier caso, los receptores previstos serían consumidores de aceite de oliva, lo que, de nuevo, no reduce los receptores a un solo perfil. Por tanto, teniendo en cuenta a los emisores como participantes identificables en el proceso comunicativo en cuestión, este trabajo siguió la distinción de muestras por tipo de emisor según los resultados hallados por López Arroyo & Roberts (2016). Este estudio sobre las notas de cata de vino halló diferencias lingüísticas entre los textos escritos por tres tipos de emisor bien diferenciados (López Arroyo & Roberts, 2016, p. 373). Por ello, este estudio clasificó las muestras también de acuerdo al tipo de emisor que redactó cada texto: (1) productores (almazaras), (2) catadores o críticos de aceite de oliva profesionales, y (3) aficionados amateurs a la cata de aceite de oliva que, principalmente, comparten sus impresiones online. Este criterio, además de proporcionar una visión más realista de la comunidad discursiva implicada, imbrica los distintos grados de especialización de cada tipo de participante en el evento comunicativo, de acuerdo a la dimensión verticales de los LFE expuesta por Calvi et al. (2009).

Otro criterio que tomar en consideración es la fuente de las muestras que componen el corpus, ya que de ellas puede depender la fiabilidad de los resultados. En este caso, las notas de cata redactadas por amateurs de la cata de aceite de oliva se extrajeron de blogs online destinados al intercambio de opiniones entre entusiastas no formados y consumidores en general. Como ejemplos de este tipo de fuentes podemos nombrar el foro online *Olive Oil Online*, algunos hilos en el sitio *Wine Spectator* que están dedicados a la cata de aceite de oliva, o páginas de consejos escritas por aficionados, como *Epicurious* o *Good Housekeeping*. En segundo lugar, los textos redactados por críticos profesionales se seleccionaron de informes de concursos de aceite de oliva (como el *International Olive Oil Contest* o la *International Olive Oil Competition*), así como de guías especializadas en aceite de oliva, como la *Guía Flos Olei* o la *Guía EVOOLEUM*, que recogen las notas de cata escritas por los paneles de expertos que ejercen de jurado. Además, también pudimos recabar muestras de este tipo a partir del contacto directo con catadores como Kathryn Tomajan, que muy amablemente proporcionaron notas de cata de su propia elaboración para este trabajo. Finalmente, las muestras redactadas por almazaras se compilaron de las páginas web de almazaras ubicadas en países anglófonos, como, por ejemplo, Alta Cresta Olive Oil, Inc., Moonshadow Grove, o Rio Bravo Rach, ubicadas en California (Estados Unidos). Hallamos estas alma-

zaras en las páginas web oficiales de instituciones oleícolas como el *International Olive Council* o el *California Olive Oil Council*.

Una vez aplicados todos los criterios de compilación, las muestras resultaron en un corpus de 620 notas de cata (20.855 palabras) escritas originalmente en lengua inglesa. El corpus general está compuesto por tres subcorpus lo más equilibrados posible, en los que las muestras se organizan de acuerdo al tipo de emisor: 250 muestras pertenecen al subcorpus de almazaras, con un total de 8.850 palabras; 230 muestras pertenecen al subcorpus de críticos profesionales, con 9.715 palabras; y 140 muestras pertenecen al subcorpus de catadores aficionados, con 2.290 palabras. Las muestras fueron etiquetadas con información identificatoria, incluyendo un número de ID, el ámbito y género al que pertenecen (notas de cata de aceite de oliva, en todos los casos); emisor (PR para productores, CR para críticos y BG para amateurs o “blogueros”); el nombre de la almazara de origen de la muestra de aceite de cada nota de cata; la variedad de aceituna de la que se extrajo el aceite; fecha y lengua original (inglés). Un ejemplo sería la etiqueta “0012_OT_PR_CW_MI_18_EN”, que indicaría que se trata de la muestra número 12 del corpus de cata de aceite de oliva (OT, por las siglas en inglés de *olive oil tasting*), perteneciente al subcorpus de productores (PR), escrita en el año 2018 en inglés (EN) en la almazara Chacewater y referida a un aceite elaborado con la variedad de aceituna mission (MI).

Sin embargo, antes de analizar el corpus, se debe comprobar si es representativo cuantitativa y cualitativamente del LFE en cuestión. Para evaluar la efectividad en la aplicación de los criterios de diseño del corpus, se realizaron dos test: el primero, para averiguar si el corpus es representativo cuantitativamente y, el segundo, para determinar su representatividad cualitativa. El primero consiste en la aplicación de fórmulas estadísticas y de muestreo: el intervalo de confianza (IC) y el error estándar de la media (SEM por sus siglas en inglés). Este método fue propuesto por Biber (1993, p. 248) para determinar si un corpus resulta representativo desde el punto de vista cuantitativo. El IC del corpus se calculó mediante la fórmula $\bar{X} \pm Z \frac{s}{\sqrt{n}}$, donde \bar{X} es la media (33,47191 en nuestro corpus), Z es el valor que se le da al IC elegido (entre 80% y 99,90%), s es la desviación estándar de acuerdo a la extensión de cada muestra (18,39156 en nuestro corpus), y n es el número de observaciones (620 muestras). El valor resultante de aplicar los valores de nuestro corpus a la ecuación anterior es $33,47 \pm 2,43$ para un IC de 99,90%, de donde puede extraerse que el error estándar de la media es $SEM = 0,737$. La implicación de este resultado relativo al tamaño de la muestra es que el corpus es altamente representativo en términos cuantitativos, dado que el SEM indica el parecido de la muestra con la media real de la población muestreada. De hecho, a menor IC (Biber ejemplifica $SEM = 10$ como valor bajo), más confianza puede tener el investigador en que su muestra es un modelo preciso de la media de la población a estudiar (Biber, 1993, p. 248).

El segundo test, destinado a evaluar la representatividad cuantitativa y cualitativa del corpus, se realizó utilizando Recor, un software desarrollado por Corpas et al. (2007) dentro del grupo de investigación en traducción y lexicografía Lexytrad (Universidad de Málaga). Esta herramienta realiza un análisis cuantitativo basado en el algoritmo N-Cor utilizando la ratio tipo/token para determinar la representatividad del corpus en términos de densidad léxica (Seghiri, 2016, p. 386). Recor genera dos gráficos (gráficos 1 y 2): el gráfico 1 representa el número de muestras en el eje horizontal, y la ratio tipo/token en el vertical, determinando cuántas muestras necesitaría el corpus para ser representativo. El gráfico 2 representa el número de tokens en el eje horizontal y determina el número mínimo de palabras que necesitaría el corpus.

Gráfico 1.

Muestras versus ratio type/token

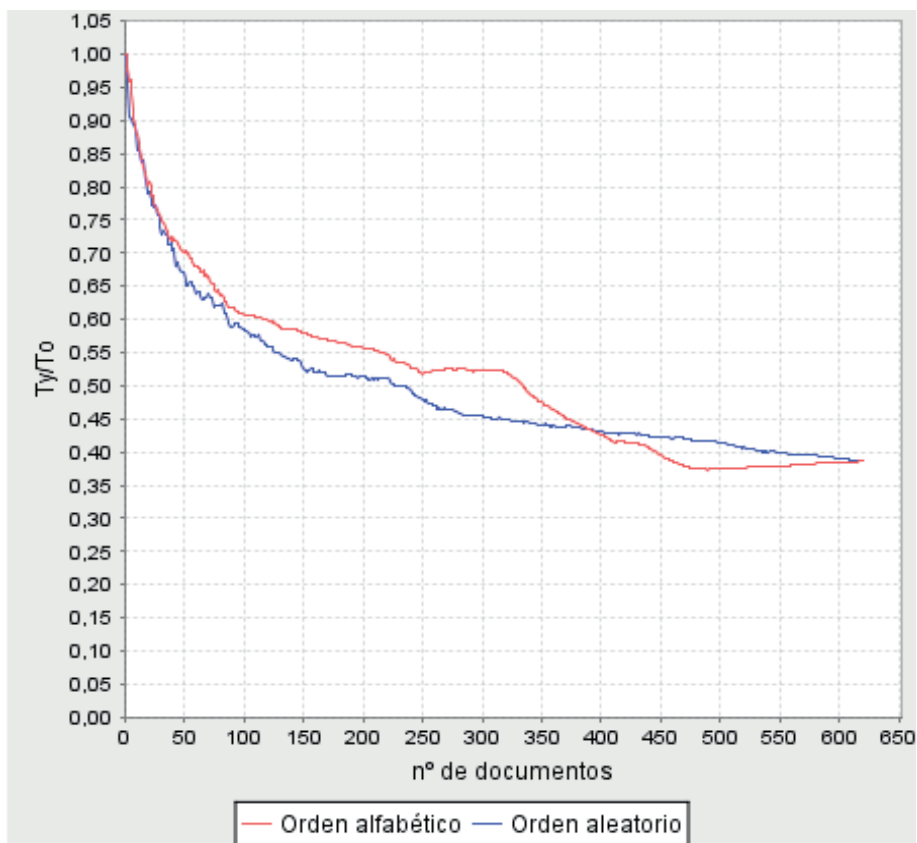
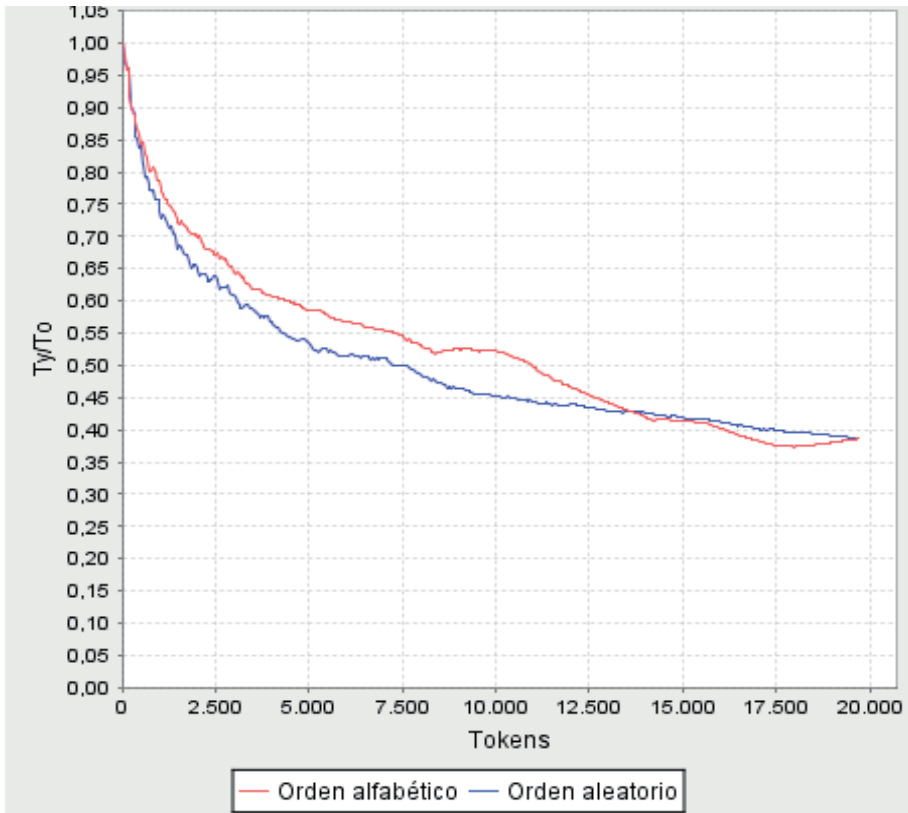


Gráfico 2.

Palabras versus ratio type/token



Ambos gráficos muestran un descenso exponencial según aumenta el número de textos (Seghiri, 2015, p. 127). La representatividad se alcanza con la estabilización de las líneas roja y azul, punto en el que puede afirmarse que el corpus es representativo (Corpas & Seghiri, 2009).

Como puede observarse en el gráfico 1, el corpus resultaría representativo con alrededor de 50 muestras. El gráfico 2 muestra un valor cercano a las 13.500 palabras como punto en el que el corpus comenzaría a ser representativo. En vista de estos resultados, puede afirmarse que el corpus utilizado en el presente estudio es representativo y, por tanto, los resultados de su posterior análisis son fiables y extrapolables. Así, la información lingüística extraída del corpus puede tomarse como modelo del conjunto del género y del LFE de las notas de cata de aceite de oliva en lengua inglesa, de forma que se garantiza que las notas de cata producidas

por la herramienta de ayuda a la escritura serán aceptables dentro de la comunidad discursiva meta.

3.2. Metodología

Hasta este punto, el corpus lingüístico no es sino una compilación de textos seleccionados siguiendo unos criterios particulares según el objetivo del estudio. Para crear una herramienta de ayuda a la escritura de notas de cata, es necesario identificar en el corpus las características lingüísticas del género y su LFE. En palabras de Garside, Leech & McEnery (1997), para extraer conocimiento de un corpus, a menudo hay que comenzar construyendo información en él (1997, p. 4); en otras palabras, anotar o dotar al corpus de información lingüística identificativa. En este caso, se realizaron distintas formas de anotación y análisis del corpus según el tipo de características lingüísticas de interés: léxico, sintáctico-gramatical, semántico y retórico. En esencia, se anotó el corpus y se analizó con varios softwares procesadores de corpus.

Más concretamente, para analizar el plano léxico, el estudio se centró en términos y unidades fraseológicas, ya que la terminología es un rasgo determinante de los LFE. Por ello, el Generador 1.0 debe incluir términos y unidades fraseológicas propios del LFE objeto de estudio. Por otra parte, en el nivel sintáctico-gramatical se analizaron las estructuras gramaticales de uso más frecuente, así como la preferencia por ciertas formas gramaticales y la disposición sintáctica de los elementos de cara a reproducir esos patrones en los textos producidos por el generador. En cuanto al plano semántico, se estudió el significado léxico identificando primero los campos semánticos más relevantes en el corpus, para después estudiar la naturaleza expresiva e interpretativa de las notas de cata de aceite de oliva y obtener una descripción más comprensiva del género. En cuanto al nivel discursivo, se trató de definir la organización interna del contenido del género de forma que se viera representada en el diseño de la herramienta de ayuda a la escritura, ya que, para que los textos producidos por el Generador 1.0 sean aceptados como ejemplos del género, deben compartir también la estructura retórica. La metodología expuesta responde al objetivo del trabajo: diseñar una herramienta de ayuda a la escritura (Generador 1.0) capaz de producir textos que los expertos y profesionales que usan este LFE reconozcan como tal a partir de la identificación empírica, representativa y fiable de las características lingüísticas propias del género de las notas de cata de aceite de oliva en inglés.

3.2.1. Análisis léxico

Para estudiar el nivel léxico, se procedió a identificar términos en el corpus, ya que la terminología es una característica determinante en los lenguajes para fines específicos (Calvi et al., 2009, p. 23). Identificamos los candidatos a térmi-

no utilizando el extractor automático TermoStat. TermoStat (Drouin, 2010) puede comparar un corpus especializado contra uno predefinido de lenguaje general, de 8 millones de palabras, compuesto por muestras del British National Corpus y de artículos periodísticos del diario de Montreal *La Gazette*. El proceso de extracción de términos comienza con un etiquetado POS (etiquetado gramatical, POS por las siglas *Part-Of-Speech* en inglés). En el corpus etiquetado gramaticalmente, el software aplica un filtro para detectar diferentes matrices sintácticas predefinidas, como sustantivo + sustantivo o adjetivo + sustantivo. A partir de la frecuencia observable de estos elementos, se obtiene la especificidad de cada candidato calculando X^2 y la probabilidad logarítmica. Al final de este proceso, el software permite recuperar los candidatos a término, además de datos sobre su frecuencia, especificidad, categoría gramatical, y líneas de concordancias, entre otros (Drouin, 2010).

Este resultado automático se examinó manualmente para descartar el ruido, es decir, candidatos a término inválidos detectados por el software (Muegge, 2012, p. 24). De esta forma, obtuvimos una lista de términos, de la cual descartamos también aquellos que presentasen menos de 5 ocurrencias en todo el corpus. Consideramos términos sustantivos y verbos ya que, para identificar unidades fraseológicas especializadas más adelante, se comienza por identificar el núcleo de la unidad, que debe ser un término (L'Homme, 2009, p. 239) que se asociará con el resto de las categorías gramaticales. Una vez identificados los términos (núcleos de unidades fraseológicas potenciales), se volvió a examinar el corpus en busca de sustantivos y adjetivos que ocurren frecuentemente para formar unidades fraseológicas como colocaciones o descriptores. Para identificar unidades fraseológicas, utilizamos dos herramientas distintas: por un lado, un software especialmente preparado para los análisis de los corpus del grupo ACTRES, que proporciona líneas de concordancia y estadísticas para corpus monolingües y corpus bilingües comparables. Por otro lado, utilizamos Lancsbox, un software desarrollado en la Universidad de Lancaster para el análisis de corpus. Esta última herramienta nos permitió no solo identificar elementos que ocurrían frecuentemente en el contexto de un término dado, sino también calcular la frecuencia de esa asociación y localizar en qué punto de cada muestra del corpus se da. Tras analizar todo el corpus con estas herramientas, conservamos solamente aquellas unidades fraseológicas con al menos 10 ocurrencias.

3.2.2. Análisis sintáctico-gramatical

Para analizar en nivel sintáctico y gramatical del corpus, se anotó gramaticalmente (POS) utilizando el etiquetador TagAnt (Anthony, 2018) y sintácticamente con el Stanford Parser (modelo factorial, Klein & Manning, 2003). La anotación sintáctica consiste en añadir información sobre la sintaxis del texto en un corpus incorporando indicadores textuales de la estructura sintáctica (Garside, Leech & McEnery, 1997, p. 34). Además, el corpus fue examinado manualmente a

fin de determinar si las muestras se componen principalmente de oraciones finitas o no finitas (con verbos conjugados o sin ellos), y en qué proporción, para añadir más datos a los análisis automáticos realizados con TagAnt y Standford Parser.

3.2.3. Análisis semántico

Aunque este aspecto de la lingüística no está directamente aplicado en el diseño de la herramienta de ayuda a la escritura, el estudio del nivel semántico de un LFE es necesario para obtener una perspectiva más completa del mismo. Además, el presente trabajo también tiene por objetivo poder proporcionar una descripción integral del género y el LFE que, si bien no tiene aplicación directa en la herramienta de ayuda a la escritura objeto de este estudio (el Generador 1.0), puede servir en el futuro para otro tipo de fines similares.

El corpus se anotó semánticamente, es decir, se clasificaron todas las palabras en campos semánticos (Bowker & Pearson, 2002, p. 84). Esto se llevó a cabo utilizando el sistema de análisis semántico (USAS, por sus siglas en inglés) desarrollado por el grupo de investigación UCREL (University Centre for Computer Corpus Research on Language) de la Universidad de Lancaster (Piao et al., 2003). El software consiste en un etiquetador que asigna etiquetas a las palabras de acuerdo a tres niveles jerárquicos de campos semánticos codificados con letras y números. Por ejemplo, en *Its_Z8 flavour_X3.1 is_A3+ incredibly_A13.3 smooth_O4.5* [su sabor es increíblemente suave], X significa “Procesos, acciones y estados psicológicos”, dentro de la cual, X3 significa “Sensorial” y, X3.1, “Sensorial: gusto”. El etiquetado automático fue revisado manualmente para descartar ruido. Los resultados se cuantificaron y se analizaron desde la lingüística cognitiva para tratar de esclarecer la relación entre la cata de aceite de oliva y el lenguaje utilizado para hablar de ella.

3.2.4. Análisis retórico

Las muestras se examinaron y se etiquetaron sus movimientos retóricos, pasos y subpasos (Swales, 1990; 2002). Un movimiento es una unidad retórica o discursiva que realiza una función comunicativa coherente en un discurso escrito u oral (Swales, 2002, pp. 228-229) y, un paso y un subpaso, subsiguientes divisiones del movimiento. Un análisis de movimientos retóricos no es un proceso automatizable, ya que es una tarea cognitiva que ha de realizarse manualmente. Para ello, debe delimitarse el panorama general a través de la identificación de la función del texto y de los fragmentos textuales que lo componen. Después, hay que examinar la funcionalidad de cada fragmento para determinar su objetivo local dentro del texto general (Biber, Connor & Upton, 2007, p. 33). Para trazar ese panorama general, identificamos los objetivos funcionales básicos de las notas de cata y los agrupamos de forma que reflejasen los movimientos y los pasos dentro de cada movimiento: tomamos las tres grandes fases de la cata (apariciencia, aroma y gusto)

y sus funciones como punto de partida del análisis retórico (Caballero, 2017, p. 69). Para continuar, recurrimos a trabajos anteriores publicados sobre la retórica de las notas de cata de vino (López Arroyo & Roberts, 2016), en los cuales se identificaba una estructura retórica. Esta estructura fue comparada con la que estábamos analizando en nuestro corpus a través del análisis de movimientos retóricos para poder acotar los movimientos, pasos y subpasos a identificar.

Una vez identificada la estructura retórica (ver sección 4. Resultados), etiquetamos el corpus retóricamente utilizando un software diseñado en ACTRES para esta tarea, a fin de clasificar todos los fragmentos de los textos dentro de la estructura retórica hallada y poder así acceder a esa información y cuantificarla usando las etiquetas.

No todos los movimientos y pasos ocurren en el corpus con la misma frecuencia, de modo que decidimos aplicar la distinción realizada por Suter (1993) entre información obligatoria y opcional. Suter divide las unidades de información en cinco categorías según su frecuencia en el corpus (1993, p. 119): información obligatoria (si ocurre en el 80-100% de las muestras); información opcional de alta prioridad (60-80%); información opcional de prioridad media (40-60%); información opcional de baja prioridad (20-40%); e información ocasional (0-20%). Los porcentajes se calcularon tras cuantificar todos los movimientos, pasos y subpasos retóricos, como en el siguiente extracto (i) del corpus etiquetado retóricamente:

- (i) [*It is a strong and robust oil* <Intensity>] [*that is characterized by a soft green entry* <Entry>], [*aromas of freshly cut grass* <Aroma>], [*slight bitterness* <Bitterness>] [*and is mildly pungent* <Pungency>]
[(*Es un aceite fuerte y robusto* <Intensidad>) (*que se caracteriza por una entrada verde y suave* <Entrada>), (*aromas de hierba recién cortada*<Aroma>), (*leve amargor*<Amargor>) (*y suave picor*<Picor>)].

Las etiquetas de todo el corpus se contaron, sumaron y normalizaron en base 100 para obtener el porcentaje de ocurrencia de cada unidad retórica.

4. Resultados

4.1. Nivel léxico

Una vez que se aplicó la metodología anteriormente descrita, se pudieron identificar los términos y unidades fraseológicas propios del género y el LFE de la cata de aceite de oliva. En concreto, se hallaron 69 términos, de los cuales 54 son sustantivos (*fruitiness, pungency, undertone*) y 15 son verbos (*to boast, to enhance*). El listado total de la terminología identificada en el corpus puede encontrarse en la tabla 1 a continuación:

Tabla 1.

Términos identificados en el corpus.

Sustantivos			Verbos
<i>acidity</i>	<i>fruit</i>	<i>phenol</i>	<i>to boast</i>
<i>aftertaste</i>	<i>fruitiness</i>	<i>polyphenols</i>	<i>to characterise</i>
<i>antioxidant</i>	<i>grassiness</i>	<i>profile</i>	<i>to display</i>
<i>aroma</i>	<i>harmony</i>	<i>pungency</i>	<i>to enhance</i>
<i>balance</i>	<i>harvest</i>	<i>quality</i>	<i>to evoke</i>
<i>bite</i>	<i>hint</i>	<i>sensation</i>	<i>to exhibit</i>
<i>bitterness</i>	<i>hue</i>	<i>smell</i>	<i>to finalize</i>
<i>blend</i>	<i>intensity</i>	<i>spice</i>	<i>to finish</i>
<i>bouquet</i>	<i>kick</i>	<i>sweetness</i>	<i>to lack</i>
<i>character</i>	<i>mouth</i>	<i>taste</i>	<i>to offer</i>
<i>characteristic</i>	<i>nose</i>	<i>texture</i>	<i>to open</i>
<i>colour</i>	<i>note</i>	<i>tone</i>	<i>to possess</i>
<i>complexity</i>	<i>nuance</i>	<i>touch</i>	<i>to ripen</i>
<i>entry</i>	<i>odour</i>	<i>trace</i>	<i>to start</i>
<i>finish</i>	<i>oil</i>	<i>transfer</i>	<i>to taste</i>
<i>flavour</i>	<i>olive</i>	<i>undertone</i>	
<i>fragrance</i>	<i>palate</i>	<i>varietal</i>	
<i>freshness</i>	<i>persistence</i>	<i>variety</i>	

Respecto a las unidades fraseológicas, se hallaron por la co-ocurrencia frecuente de los términos identificados con otros elementos léxicos. Un ejemplo es el término *fruitiness* [afrutado] y su asociación con un adjetivo para dar lugar a la unidad fraseológica *green fruitiness* [afrutado verde]; o el término *color* [color], que da lugar a la nominación *limpid gold color* [color oro limpio] o la colocación *yellow color* [color amarillo]. De la misma forma, se hallaron operadores como *with notes of* [con notas de], *with a touch of* [con un toque de] o *enriched by* [enriquecido por], que desempeñan un rol puramente funcional en la nota de cata. En la tabla 2 se muestra un conjunto de unidades fraseológicas esperables en una nota de cata en lengua inglesa prototípica.

Tabla 2.

Unidades fraseológicas identificadas en el corpus

Nominaciones	Operadores	Colocaciones
<i>olive oil</i>	<i>notes of</i>	<i>golden yellow</i>
<i>virgin olive oil</i>	<i>hints of</i>	<i>grass green</i>
<i>extra virgin olive oil</i>	<i>with notes of</i>	<i>green apple</i>
<i>well balanced</i>	<i>with hints of</i>	<i>ripe tomato</i>
<i>early harvested</i>	<i>a touch of</i>	<i>medium intensity</i>
<i>intensity of bitter</i>	<i>with a hint of</i>	<i>green banana</i>
<i>intensity of pungency</i>	<i>intensity of</i>	<i>tomato leaf</i>
<i>olive fruitiness</i>	<i>level of</i>	<i>green tomato</i>
<i>bitter taste</i>	<i>together with</i>	<i>green olives</i>
<i>stinging sensation</i>	<i>followed by</i>	<i>black pepper</i>
<i>green fruit</i>	<i>rich in</i>	<i>fresh almond</i>
<i>ripe fruitiness</i>	<i>enriched by</i>	<i>olive leaf</i>
<i>bitter and pungent</i>	<i>characterised by</i>	<i>yellow colour</i>
<i>well balanced with</i>	<i>reminiscent of</i>	<i>ripe fruits</i>
<i>bitterness and pungency</i>	<i>on the palate</i>	<i>bitter almond</i>
<i>limpid golden yellow</i>	<i>in the mouth</i>	<i>delicate intensity</i>
<i>golden yellow colour</i>	<i>in the nose</i>	<i>pink pepper</i>
<i>freshly mown grass</i>	<i>in the throat</i>	<i>fig leaf</i>
<i>freshly cut grass</i>		<i>medium harmonious</i>
<i>tasting notes of</i>		<i>slightly bitter</i>
<i>a flavour of</i>		<i>green and ripe</i>
<i>blend of</i>		<i>clean and persistent</i>
<i>flavours of</i>		<i>to open with</i>
<i>present and well balanced</i>		<i>to start with</i>

Estos resultados del análisis léxico contribuyen no solo a la creación del Generador de notas de cata, sino que también reflejan la conceptualización

lingüística de la realidad de la cata de aceite de oliva a nivel léxico. De hecho, la estandarización terminológica es de gran relevancia en los LFE a fin de reducir ambigüedades lingüísticas, lo cual se vuelve vital para la tarea de teóricos y practicantes de labores en el sector, como el diseño de etiquetado, entre otras. La estandarización de la terminología hallada, sin embargo, está fuera de los objetivos del presente estudio, si bien supone un primer paso hacia una estandarización más rica y comprensiva en el futuro. En cualquier caso, los datos recabados suplen las necesidades del presente trabajo de cara al diseño un Generador que produzca textos que se adhieran a los requisitos del género de la nota de cata y muestren las características propias del LFE en cuestión.

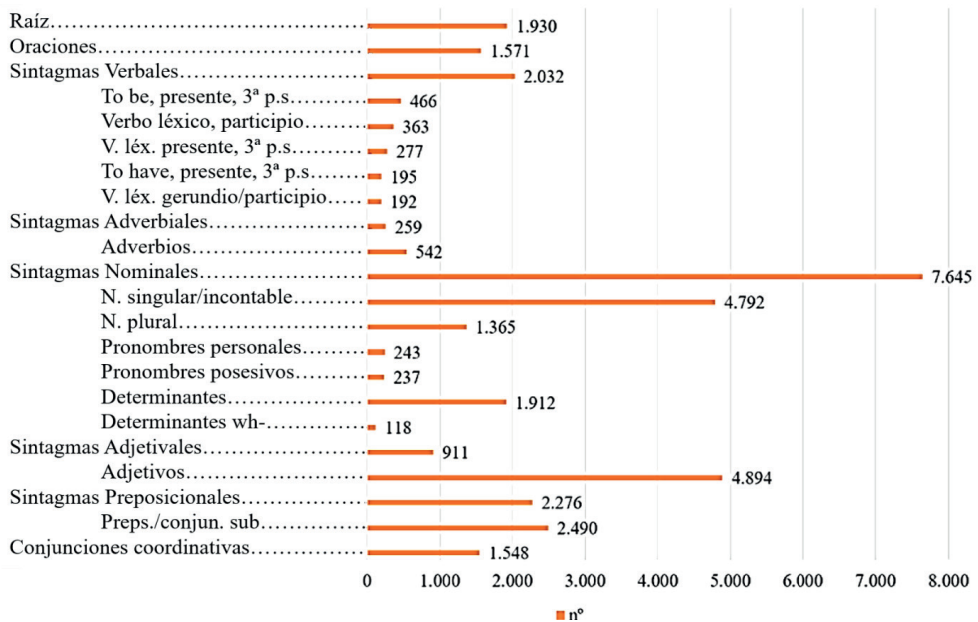
4.2. Nivel sintáctico-gramatical

Los resultados del etiquetado gramatical del corpus se examinaron, cuantificaron y resumieron (gráfico 3).

Aunque los resultados cuantitativos puedan parecer superficiales, ya que el análisis no comprende las funciones sintáctico-gramaticales, de los datos del gráfico 3 se puede extraer información cualitativa de gran relevancia.

Gráfico 3.

Elementos sintáctico-gramaticales en el corpus



Por ejemplo, puede observarse que son más frecuentes los sustantivos y los adjetivos que los sintagmas nominales y adjetivales. Sin embargo, no hay más verbos que sintagmas verbales, lo que indica que las notas de cata de aceite de oliva en inglés poseen alta densidad léxica y carecen de elementos narrativos. De hecho, la mayoría de los verbos aparecen conjugados en tercera persona singular y en formas no finitas (no conjugadas), por lo que observamos que los verbos se utilizan para describir al aceite de oliva en sí prácticamente en exclusividad. En esta misma línea, el corpus es más abundante en adverbios que en sintagmas adverbiales, lo que solo puede significar que los verbos están generosamente modificados por adverbios. De nuevo, esto resulta en un lenguaje rico y condensado, dominado por palabras de contenido y no funcionales. En esencia, se podría decir que solo se utilizan los verbos estrictamente necesarios para comunicar acciones.

Esta perspectiva se completa teniendo en cuenta los resultados del análisis del corpus en función de las cláusulas finitas y no finitas que contiene (tabla 3).

Tabla 3.

Cláusulas finitas y no finitas en el corpus

Notas de cata que contienen...	Probabilidad %
... al menos 1 cláusula finita	66,51
... al menos 1 cláusula no finita	33,49
... 2 cláusulas finitas	28,71
... 1 cláusula finita	26,77
... 0 cláusulas finitas	21,45
... 2 cláusulas no finitas	17,74
... 1 cláusula no finita	29,19
... 0 cláusulas no finitas	41,62

Esta parte del análisis añadió la siguiente perspectiva al análisis sintáctico-gramatical anterior: de media, una nota de cata de aceite de oliva contiene 1,63 cláusulas finitas y 1,03 no finitas. De hecho, hay el doble de probabilidades de que una nota de cata esté totalmente compuesta por cláusulas finitas (41,62% de las muestras del corpus) que de que lo esté por no finitas (21,45%). Aun así, de las 620 muestras en total que contiene el corpus, hay al menos una cláusula finita en el 66,51% de ellas. Al mismo tiempo, el 33,49% de las notas de cata de aceite de oliva contienen al menos una cláusula no finita. En este sentido, una nota de cata tiene un 28,71% de probabilidad de contener dos cláusulas finitas; 26,77% de contener una, y 21,45% de no contener ninguna. Por otra parte, tienen un 17,74% de probabilidades de contener dos cláusulas no finitas; 29,19% de contener una, y 41,62% de no contener ninguna. Aunque pueda parecer contradictorio que los elementos

finitos sean menos frecuentes que los no finitos en relación al número de cláusulas y muestras en las que aparecen, esto puede explicarse con el hecho de que una sola cláusula finita se utiliza para incluir proporcionalmente más contenido descriptivo y estático que acciones, como puede observarse en el siguiente ejemplo (ii) (muestra 0010_OT_PR_SDM_MR_18_EN del corpus):

- (ii) *Volcanic soil, early harvesting, not stressing the fruit when picked, and a unique method of extraction and storage give this extra virgin olive oil a high phenolic bio-active content, expressed in an intense, green tomato fruity flavor that is sweetly bitter and pungent with a decidedly long lasting, grassy taste.*

[El suelo volcánico, la recolección temprana, no estresar la fruta al recolectarla y un método de extracción y almacenado único otorgan a este aceite de oliva virgen extra un alto contenido fenólico bioactivo, expresado en un sabor afrutado, intenso, a tomate verde que es dulcemente amargo y pungente con un sabor herbáceo decididamente alargado].

Como se puede observar, la nota de cata solamente contiene dos elementos finitos, *give* [otorgar] e *is* [es], mientras que contiene numerosos sustantivos, adjetivos y adverbios, ninguno de los cuales modifica a los verbos finitos. En esencia, los campos semánticos a los que pertenecen los escasos verbos hallados poseen baja carga semántica (ser, dar, hacer), por lo que la información la transmiten los sustantivos, adjetivos y adverbios, lo que da lugar a una alta densidad léxica. Esta muestra del corpus ejemplifica la tendencia explorada anteriormente: aunque la mayoría de las notas de cata sí contienen cláusulas finitas, estas no son predominantes en el género, sino que se usan principalmente como base en la que incorporar rica información descriptiva y estática en forma de sustantivos, adjetivos y adverbios. En resumen, el análisis realizado en este nivel lingüístico retrata el género y el LFE de la nota de cata de aceite de oliva como altamente denso léxicamente y escaso en recursos narrativos que comuniquen información relativa a acciones, aunque los resultados prueban que no se obvian por completo los elementos finitos. Aun así, esto se combinará con el hecho de que un 21,45% de las muestras no contienen ningún elemento finito a la hora de diseñar el Generador de notas de cata, como se detalla más adelante.

4.3. Nivel semántico

El análisis de contenido (Garside, Leech & McEnery, 1997) llevado a cabo a partir del etiquetado semántico automático del corpus (USAS) reveló los principales campos semánticos alrededor de los cuales orbita la actividad de la cata y su plasmación en lengua inglesa. Los resultados del etiquetado automático

tuvieron que ser revisados manualmente dado que el sistema etiquetó ciertas palabras incorrectamente. Por ejemplo, *rich* [rico] recibió la etiqueta de “Dinero y comercio”, en el campo de la “Industria” (I1), cuando el uso de *rich* en este LFE es indicar la abundancia de cualidades organolépticas y no de dinero. Otro ejemplo es la palabra *virgin* [virgen], que el sistema USAS etiquetó en el campo semántico de “Relaciones íntimas y sexuales” (S3.2). En la fase de post-edición de los resultados automáticos también se descartaron las ocurrencias del verbo *to be* [ser/estar] (A13), del que se hallaron 452 ocurrencias, y de otras “Palabras gramaticales y funcionales” (Z5) (5.743 ocurrencias), así como los elementos categorizados como “Sin categoría” (Z99), con 1.192 ocurrencias. Los resultados emitidos por el sistema USAS, una vez limpios, se presentan en la tabla 4 a continuación.

Tabla 4.

Campos semánticos predominantes en el corpus

Campo semántico	Nº elementos	Ejemplos
A – Términos generales y abstractos		
A5 – Evaluación	317	<i>perfect, quality</i> [perfecto, cualidad]
A6 – Comparación	268	<i>odd, reminiscent</i> [extraño, reminiscete]
A9 – Dar y recibir; posesión	332	<i>has, lacks</i> [tiene, carece]
A13 – Grado	280	<i>slightly, awfully</i> [levemente, horriblemente]
F – Alimentación y agricultura		
F1 – Alimentación	2.066	<i>nutty, pepper</i> [a frutos secos, pimienta]
L – Vida y seres vivos		
L3 – Plantas	1.129	<i>ripe, grassy</i> [maduro, herbáceo]
N – Números y medidas		
N3 – Medidas	289	<i>length, medium</i> [longitud, medio/a]
N5 – Cantidades	540	<i>moderate, ample</i> [moderado, amplio]
O – Sustancias, materiales, objetos y equipamiento		
O4 – Atributos físicos	1.872	<i>lush, pleasant</i> [jugoso, agradable]
O4.1 – Apariencia general y propiedades físicas	377	<i>nuances, structured</i> [matices, estructurado]

Tabla 4 (continuación)

Campos semánticos predominantes en el corpus

Campo semántico	Nº elementos	Ejemplos
O4.2 – Juicios sobre apariencia	351	<i>elegant, impressive</i> [elegante, impresionante]
O4.3 – Color y patrones de color	1.014	<i>vivid, green</i> [vívido, verde]
O4.5 – Textura	93	<i>soft, smooth</i> [suave, suave/fluido]
S – Acciones, estados y procesos sociales		
S1 – Acciones, estados y procesos sociales	266	<i>approachable, to boast</i> [accesible, presumir]
S1.2 – Rasgos de la personalidad	258	<i>strong, intimate</i> [fuerte, íntimo]
S1.2.5 – Dureza; fuerte/débil	208	<i>robust, mild</i> [robusto, leve]
T – Tiempo		
T2 – Tiempo: principio y final	247	<i>long-lasting, short</i> [de larga duración, corto]
T3 – Tiempo: Viejo, nuevo y edad	262	<i>freshness, young</i> [frescura, joven]
X – Acciones, estados y procesos psicológicos		
X3 – Sensorial	1.494	<i>sensation, to sense</i> [sensación, percibir]

Los datos muestran que las notas de cata de aceite de oliva giran en torno a varios campos semánticos interrelacionados entre ellos: términos abstractos (evaluación, comparación, posesión y graduación); alimentación y vida (alimentación y plantas); medidas y tiempo (cantidades, longitud, edad); atributos físicos (apariencia general y juicio de la apariencia, incluyendo color y textura); y atributos y procesos humanos, incluyendo rasgos sociales, de personalidad y sensoriales. Algunos de estos campos semánticos hacen referencia a entidades reales (alimentos, plantas, color) que pueden percibirse y comunicarse de forma objetiva y empírica. Sin embargo, el resto de los campos semánticos (términos abstractos, medidas y tiempo, apariencia general y su juicio) transmiten usos más subjetivos del lenguaje con el objetivo de formular información sobre los atributos del aceite. Dado que los campos semánticos agrupan las palabras cuyo concepto mental es común en mayor o menor grado (Garside, Leech y McEnery, 1997, p. 54), y que los conceptos son la base ontológica del léxico (Paradis, 2005, p. 542), podrían usarse

los resultados relativos a los campos semánticos de las notas de cata para explicar la naturaleza ontológica del LFE de la cata de aceite de oliva.

En este sentido, entran en juego las cuestiones de la percepción, la comunicación de las percepciones propias y su interpretación por parte del receptor. Para estudiar la naturaleza de la construcción del significado en este LFE, tomamos un enfoque basado en la lingüística cognitiva (Paradis, 2010). Dado que nuestro acceso principal y relativamente objetivo al mundo que nos rodea es a través de la vista, las percepciones olfativas y gustativas pasan a ser conceptualizados como efectos de las entidades visualmente perceptibles (Dubois, 2007, p. 170). Este hecho afecta al lenguaje usado para la comunicación de percepciones sensoriales, conllevando diferencias ontológicas entre descriptores lingüísticos (Paradis, 2010, p. 7), que, a su vez, causan que el lenguaje utilizado esté altamente ligado al perceptor y sea menos autónomo (Dubois, 2007, p. 170). En esencia, esto se plasma en una mayor objetividad a la hora de comunicar sobre percepciones visuales que sobre las olfativas y gustativas, que se conceptualizan en función de la información proporcionada por el sentido de la vista. De esta forma, estas últimas modalidades sensoriales resultan más subjetivas y heterogéneas entre distintos perceptores a consecuencia de esa ruta cognitiva indirecta que tomamos al conceptualizar la realidad de este tipo. Un ejemplo son las siguientes notas de cata (muestras 0018_OT_PR_APH_KK_18_EN (iii) y 0275_OT_CR_FRA_FR_18_EN (iv) del corpus):

(iii) *The oil has a sparkling colour with shades of bright green.*

[El aceite tiene un color centelleante con sombras de verde brillante.]

(iv) *Soft and beautifully balanced on the palate with a complex mixture of green and ripe fruity notes (nutty, ... straw). The finish is long—pleasantly peppery and slightly bitter with a lingering nutty sweetness*

[Suave y hermosamente equilibrado en boca con una compleja mezcla de notas afrutadas verdes y maduras (a frutos secos, ... paja). El final es largo; agradablemente picante y levemente amargo con un dulzor persistente a frutos secos.]

La nota de cata (iii) contiene información acerca de percepciones visuales, sobre las que se puede llegar a cierto consenso entre distintos perceptores, a través de parámetros contrastables como el código de color RGB, por ejemplo. Por el contrario, la nota de cata (iv) muestra cuán ligadas al perceptor (Dubois, 2007, p. 170) están las percepciones olfativas y gustativas, ya que es difícil alcanzar un consenso sobre qué significan exactamente expresiones como *beautifully balanced* [hermosamente equilibrado], *pleasantly peppery* [agradablemente picante] o *lingering nutty sweetness* [dulzor a frutos secos persistente].

Teniendo en cuenta esta información, tratamos de determinar el grado de objetividad del lenguaje utilizado en las muestras de nuestro corpus, dado que la nota de cata consiste en la plasmación lingüística de las percepciones sensoriales del catador respecto a una muestra de aceite. A partir del etiquetado retórico realizado en el corpus, se calculó el porcentaje de lenguaje que alude a percepciones visuales, por un lado, y olfativas y gustativas, por el otro. Los resultados muestran que, en nuestro corpus, el 11,29% de muestras contienen lenguaje que alude a percepciones visuales en oposición al 58,55% y 99,52% de muestras que contienen alusión al aroma y gusto del aceite de oliva. A la luz de estos datos, podría afirmarse que el LFE de las notas de cata de aceite de oliva tiene una función principalmente descriptiva y evaluativa, pero predominantemente subjetiva en cuanto a la naturaleza de la información sensorial que se comunica.

4.4. Nivel retórico

El estudio retórico del género de la nota de cata de aceite de oliva se realizó siguiendo un análisis de contenido (Biber, Connor, & Upton, 2007) consistente en adquirir una visión general de la función textual de la nota de cata y continuar identificando las funciones locales dentro del texto de cada uno de sus segmentos. El objetivo es detectar patrones semánticos y funcionales comunes (2007, p. 33), que se clasificarán jerárquicamente en movimientos, pasos y subpasos (Swales, 1990, p. 58). En este sentido, la función general de las notas de cata puede formularse como describir y evaluar los atributos positivos y negativos de una muestra de aceite de oliva por parte de un catador particular. Partiendo de esta función general, se examinaron las muestras del corpus hasta agrupar las funciones locales, resultando en la siguiente estructura retórica:

1. Variedad de aceituna + D.O.
2. Localización
 - 2.1. Clima
 - 2.2. Terreno
 - 2.3. Geografía
3. Elaboración
 - 3.1. Método de recogida
 - 3.2. Método de prensado
 - 3.3. Fecha
 - 3.4. Temperatura
4. Nota de cata
 - 4.1. Aroma
 - 4.1.1. Aroma
 - 4.1.2. Madurez

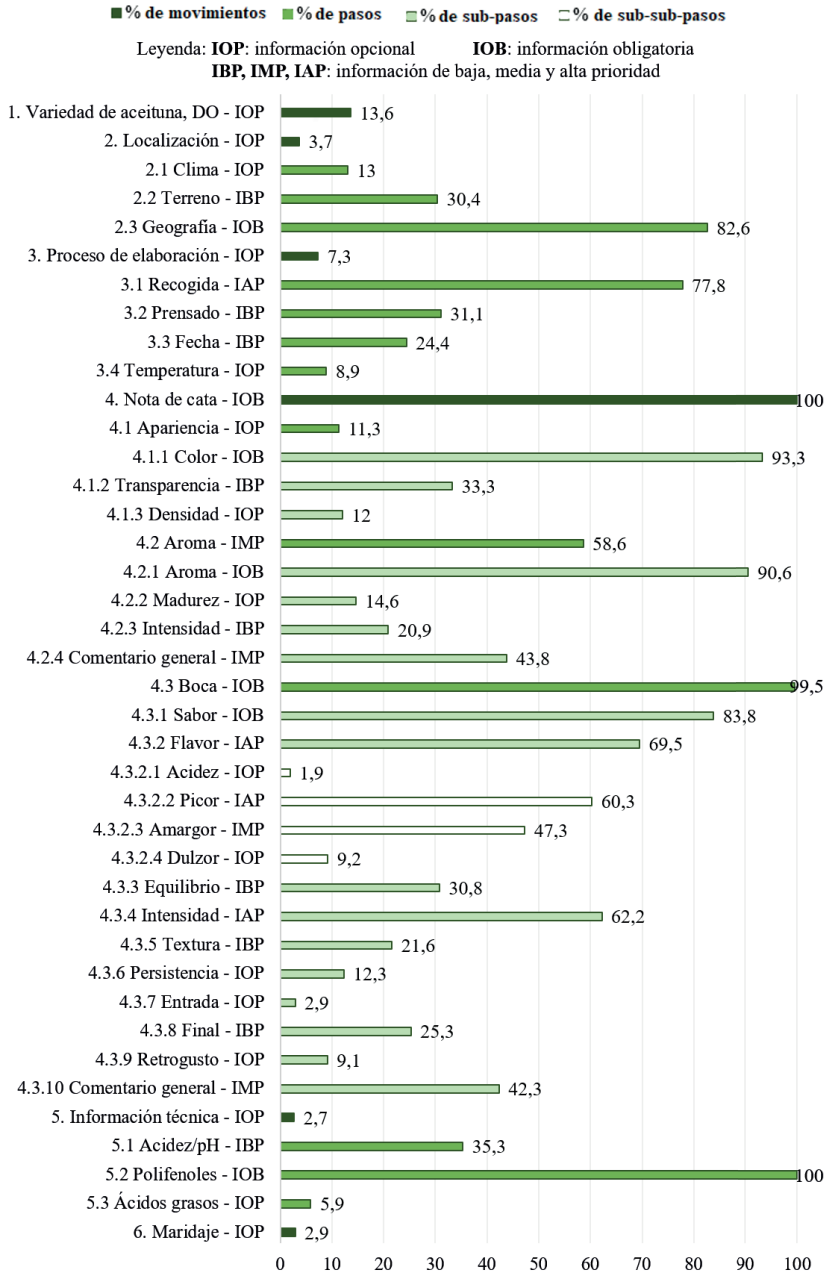
- 4.1.3. Intensidad
- 4.1.4. Comentario general
- 4.2. Apariencia
 - 4.2.1. Color
 - 4.2.2. Densidad
 - 4.2.3. Transparencia
- 4.3. Boca
 - 4.3.1. Sabor
 - 4.3.2. Flavor
 - 4.3.2.1. Acidez
 - 4.3.2.2. Picor
 - 4.3.2.3. Amargor
 - 4.3.2.4. Dulzor
 - 4.3.3. Equilibrio
 - 4.3.4. Intensidad
 - 4.3.5. Textura
 - 4.3.6. Persistencia
 - 4.3.7. Entrada
 - 4.3.8. Final
 - 4.3.9. Retrogusto
 - 4.3.10. Comentario general
- 5. Información técnica
 - 5.1. Acidez/nivel de pH
 - 5.2. Contenido de polifenoles
 - 5.3. Contenido en ácidos grasos Omega-3 y -6
- 6. Maridaje

Esta estructura prototípica del género de la nota de cata de aceite de oliva en lengua inglesa no aparece reflejada en su totalidad en cada nota de cata. Por ello, se aplicó la clasificación de Suter (1993, p. 119) sobre los diferentes tipos de información según su nivel de obligatoriedad. En el gráfico 4 se muestra la distribución de obligatoriedad en cada uno de los movimientos, pasos y subpasos retóricos identificados.

En resumen, los resultados del análisis retórico muestran que una nota de cata de aceite de oliva prototípica contendría 6 movimientos principales, 13 pasos, 17 subpasos y 4 subsubpasos. Las unidades de información principales son Variedad de aceituna, Localización, Proceso de elaboración, Nota de cata, Información técnica y Maridaje, todas ellas opcionales a excepción de la Nota de cata en sí, que es la única unidad de inclusión obligatoria y que aparece en el 100% de las

Gráfico 4.

Distribución y relevancia de movimientos, pasos y subpasos retóricos en el género de la nota de cata



muestras del corpus. Dentro de cada movimiento hay pasos y subpasos retóricos, algunos de los cuales son unidades de información obligatoria siempre que la unidad superior en la jerarquía esté incluida en el texto. Un ejemplo es Color: siempre que una nota de cata describa y evalúe la Apariencia del aceite, el subpaso Color debe incluirse obligatoriamente. Como se muestra en el gráfico 4, el género solo presenta una unidad de información funcional obligatoria, la Nota de cata, dentro de la cual Boca es el único paso obligatorio, al igual que Sabor dentro de Boca. Para que un texto en lengua inglesa sea considerado una nota de cata aceptable por la comunidad discursiva de la cata de aceite de oliva, este no tiene por qué incluir la totalidad de unidades de información identificadas en la estructura retórica anterior. De hecho, ninguna de las muestras de nuestro corpus contiene todos los movimientos y pasos retóricos; la muestra (v) 0001_OT_PR_SDM_BL_18_EN es un ejemplo de nota de cata que cumple con los requisitos retóricos del género, ya que hace alusión al único paso de inclusión obligatoria:

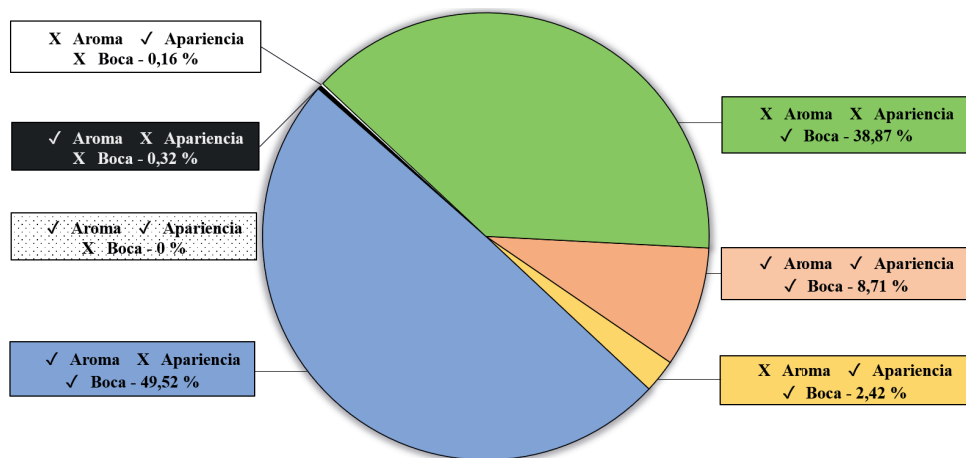
- (v) [*Grassy and fruity, [the flavor of this oil [has pungent]<Picor> [and bitter sensations]<Amargor> accompanied by [long, aromatic notes]<Final>]<Sabor>]<Boca>*

[Herbáceo y afrutado, el sabor de este aceite tiene sensaciones pungentes y amargas acompañadas de notas largas y aromáticas.]

Para ahondar en la estructura retórica más común en las notas de cata en lengua inglesa, se estudió la co-ocurrencia de los pasos del único movimiento obligatorio, Nota de cata, a saber, Aroma, Apariencia y Boca. En el gráfico 5 se

Gráfico 5.

Co-ocurrencia de los pasos del movimiento obligatorio Nota de cata



muestra la frecuencia en la que coinciden estos tres pasos retóricos principales del género incluyendo todas las posibilidades combinatorias, esto es, el porcentaje de co-ocurrencia en el corpus los tres pasos, dos de ellos, o uno solo.

Los resultados muestran que la estructura retórica más frecuente en nuestro corpus es aquella que incluye los pasos Aroma y Boca sin Apariencia (49,52% de las muestras del corpus), seguida de aquella que incluye Boca sin Aroma ni Apariencia (38,87%). Estos datos parecen indicar que, para que se considere que un texto cumple con los requisitos retóricos del género de la nota de cata de aceite de oliva, la opción más adecuada es incluir, obligatoriamente, una unidad de información funcional que aluda las sensaciones en boca del aceite de oliva. También es altamente recomendable incluir referencia a su aroma, si bien no es obligatorio. Complementariamente, puede incluirse cualquiera de las otras unidades retóricas identificadas, comenzando con las más frecuentes. En resumen, el estudio del nivel retórico del corpus ha proporcionado información cualitativa en cuanto a la organización interna de las notas de cata de aceite de oliva, y cuantitativa en cuanto a cuáles es más frecuente incluir, cuáles son obligatorias y cuáles totalmente opcionales.

4.5. Aplicación de los resultados a la creación del Generador 1.0

Una vez obtenidos los resultados, se procedió al diseño de la herramienta de ayuda a la escritura de forma que ésta reproduzca en los textos las características lingüísticas de las notas de cata de aceite de oliva. Para ello, se decidió dotar al Generador 1.0 de diferentes opcionalidades de funcionamiento, recursos de apoyo para su uso eficiente, y estructuras retóricas y gramaticales ya incluidas en base a los resultados y conclusiones extraídos del análisis lingüístico del corpus. Este proceso de diseño garantiza que los textos producidos por esta ayuda a la escritura puedan ser identificados como notas de cata de aceite de oliva por la comunidad discursiva meta. Dado que para conseguir este objetivo ha sido necesario el análisis multidimensional del corpus, a continuación se detalla cómo se han tratado e interpretado los resultados de cada uno de los niveles lingüísticos (léxico, sintáctico-gramatical, semántico y discursivo) y cómo se han aplicado al diseño de la herramienta de ayuda a la escritura.

Desde el punto de vista léxico, los términos y unidades fraseológicas identificadas abren camino hacia una estandarización más detallada de las expresiones léxicas ligadas a la actividad profesional de la cata de aceite. Aun así, se decidió incluir en el Generador 1.0 el mayor número de unidades que pudieran ser de utilidad para el usuario. Por eso, se incluyeron también las palabras clave presentes en el corpus en un glosario de equivalencias español-inglés al que se accede a través del botón “Glosario”. De esta forma, el usuario puede encontrar la expresión en español cuyo equivalente inglés desconozca para incluirlo en su ficha de cata y así adaptarla a los atributos de su aceite de oliva particular.

Los resultados del análisis sintáctico-gramatical parecen apuntar a que el género de las fichas de cata es un género denso léxicamente, con sustantivos modificados profusamente por adjetivos y adverbios, como describimos en la sección anterior. Además, puede observarse escasa narración, ya que la mayoría de los verbos son aquellos en 3ª persona singular o en formas no finitas destinadas a la descripción del aceite. En vista de estas características sintáctico-gramaticales, se incluyó en el Generador 1.0 la opción de que el usuario elija una opción “Básica”, en la cual las líneas modelo (estructuras léxico-gramaticales prototípicas con huecos a rellenar por el usuario donde se esperaría variación de la información) no presentan verbo finito, o una opción “Avanzada”, en la cual las líneas modelo contienen un verbo conjugado. Las líneas modelo incluidas en el Generador 1.0 son las dos más frecuentes de uso en nuestro corpus en cada uno de los movimientos y pasos retóricos.

Los resultados del plano semántico no tienen una aplicación directa en el diseño de la herramienta de ayuda a la escritura; sin embargo, aportan información relevante en cuanto a la descripción integral del género de las notas de cata de aceite de oliva en lengua inglesa. Los campos semánticos identificados en el corpus revelan los principales conceptos mentales que subyacen al LFE de la cata de aceite de oliva (alimentos, plantas, atributos físicos y su evaluación, términos abstractos de evaluación y comparación, etc.). Además, aplicamos un análisis centrado en la existencia de sinestesia en las percepciones expresadas a través del lenguaje. Es decir: las proposiciones lingüísticas que describen percepciones visuales son susceptibles de expresión objetiva en mayor medida que aquellas percepciones olfativas y gustativas, que se conceptualizan en referencia a la información visual, y son por tanto más inestables entre individuos. En vista de estos hechos, parece factible tipificar el género como uno predominantemente subjetivo por naturaleza y en relación con los campos semánticos identificados.

Los resultados del análisis del nivel discursivo se integraron en la herramienta de ayuda a la escritura como base en la cual el usuario puede seleccionar qué unidades de información desea incluir o no. Además, gracias a la clasificación de los movimientos según su prioridad y al análisis de coincidencia entre los pasos del único movimiento obligatorio de la estructura retórica, se aconseja al usuario que incluya siempre el movimiento Nota de cata, dentro del cual lo más apropiado es incluir Boca o Boca y Aroma juntos.

En esencia, el diseño del Generador 1.0 se ha beneficiado de los rasgos lingüísticos integrados en su modelo lingüístico, ya que los textos que se producen presentan las características propias de un ejemplar prototípico del género de las notas de cata de aceite de oliva en lengua inglesa. La estructura retórica identificada sirve como menú en el cual el usuario elige la información a incluir en su nota de cata.

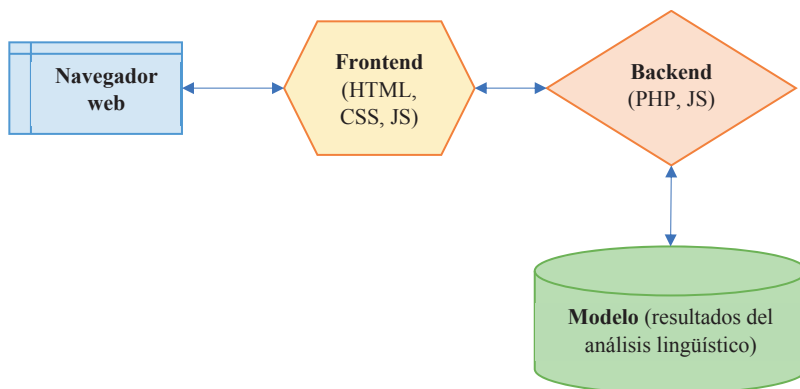
5. Generador de Notas de Cata de Aceite de Oliva

Desde el punto de vista técnico, el Generador 1.0 está basado en una web interactiva disponible en los exploradores de internet de uso común, incluyendo Google Chrome, Microsoft Edge, Mozilla Firefox e Internet Explorer. La interfaz web del Generador 1.0, producto final de este trabajo de investigación, además de cubrir una necesidad socioeconómica, cuenta con un diseño intuitivo para el usuario. También es una herramienta fiable, ya que los textos producidos serán gramaticalmente correctos, semánticamente comprensibles y textualmente aceptables.

Esta versión 1.0 del Generador de Notas de Cata de Aceite de Oliva fue desarrollada en forma de aplicación informática por el Dr. Francisco José Palacios Burgos. Se dotó a la aplicación con una arquitectura web, lo que hace posible su uso desde cualquier ubicación y dispositivo. El Generador está construido en base a los siguientes componentes arquitectónicos (el gráfico 6 contiene un esquema de la arquitectura): el modelo lingüístico, el *backend*, el *frontend* y el navegador web.

Gráfico 6.

Arquitectura de la aplicación web con constituye el Generador



El modelo lingüístico es la base de conocimiento que abarca el conjunto de rasgos y elementos identificados y clasificados hasta el momento a partir del análisis lingüístico del corpus. El *backend* es el bloque tecnológico que procesa el modelo, y está formado por varios elementos desarrollados con tecnologías de programación PHP y JavaScript. Una vez que el *backend* procesa el modelo, el resultado del procesamiento se traslada al *frontend*, cuya misión es doble: por un lado, el *frontend* recibe las peticiones del usuario (tipo de línea modelo, sección de información a incluir, apertura del glosario, etc.); y, por otra parte, el *frontend* recoge el resultado del procesamiento del *backend* para construir una respuesta web que el usuario visualizará en el navegador utilizando HTML, CSS (hojas de estilo en cascada) y JavaScript. Adicionalmente, el *backend* es capaz de generar versiones de notas de cata en formatos descargables, editables e imprimibles (Word).

Las ventajas del Generador 1.0 frente a otras alternativas de internacionalización de la cata de aceite de oliva jiennense, como la traducción, son varias. Posibilitar la redacción directa y semi-automática de las notas de cata es la más significativa. Para empezar, una traducción automática sería una solución demasiado arriesgada, ya que, si estos sistemas producen errores en traducciones de lenguaje general, una traducción de un LFE no hace sino aumentar la posibilidad de error dada la especificidad del texto. Por otro lado, el coste de una traducción especializada por parte de un traductor humano es alto en relación con el coste de adoptar el uso de una sola aplicación web. Además, la traducción de una nota de cata, donde los matices culturales tienen gran peso, puede desembocar en una pérdida en la calidad del texto en la lengua meta y de la imagen del producto. Sin embargo, el Generador, al basarse en un modelo lingüístico diseñado según las características reales del LFE de la cata de oliva en inglés, garantiza la reproducción de los rasgos propios de ese lenguaje, sin distorsión ni intermediación lingüística. No se trata de un software que traduce textos ni que proporciona plantillas al usuario para que éste las tome como modelo a la hora de redactar su nota de cata en inglés; se trata de una herramienta de composición asistida de textos especializados basado en un modelo léxico-gramatical del LFE en cuestión establecido en base a parámetros de frecuencia y relevancia, tal y como se ha detallado en el presente trabajo.

La aplicabilidad del Generador es versátil, ya que la exportación de producto es una tarea que cobra cada vez más importancia, como evidencian los datos aportados en la sección 1 del trabajo. El Generador es útil para profesionales que desempeñen su labor en departamentos de marketing o de comercio exterior de grandes cooperativas, pero también para aquellos pequeños productores que deseen internacionalizar su aceite de oliva. Desde notas de cata para etiquetas o páginas web de las empresas en versión bilingüe, hasta notas de cata aplicables a la preparación de catas guiadas a turistas extranjeros en almazaras locales, el Generador de Notas de Cata de Aceite de Oliva 1.0 supone una solución sencilla, rápida, fiable y económica a los desafíos que enfrenta actualmente el sector olivarero.

5.1. Uso del Generador de Notas de Cata de Aceite de Oliva 1.0

Al abrir el Generador, el usuario visualiza la web en la que se basa (ilustración 1).

Ilustración 1.

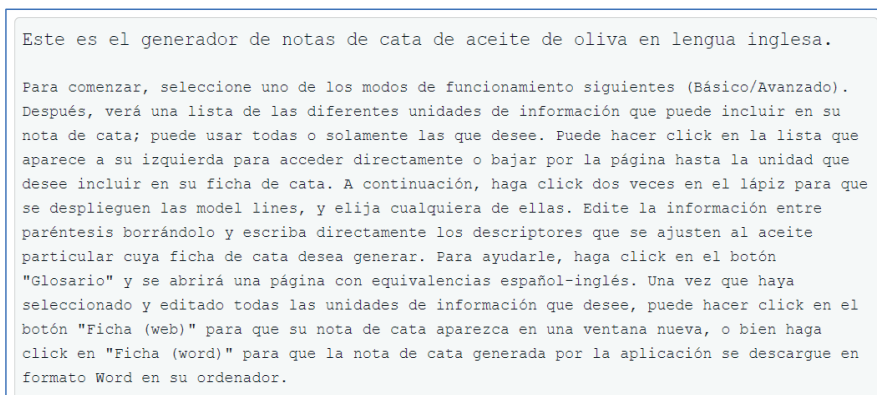
Interfaz



En primer lugar, aparece un resumen de las instrucciones de uso del Generador (ilustración 2) para que el usuario pueda seguir las durante el uso de la herramienta.

Ilustración 2.

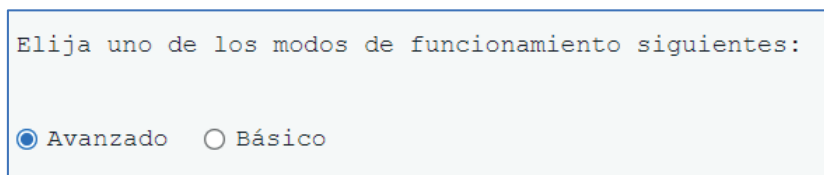
Instrucciones de uso



El primer paso para comenzar a utilizar el Generador es seleccionar uno de los modos de uso, Básico o Avanzado, basados en los resultados del estudio lingüístico (ilustración 3).

Ilustración 3.

Primer paso: selección del nivel de complejidad lingüística



Elija uno de los modos de funcionamiento siguientes:

Avanzado Básico

A continuación, en la parte izquierda de la pantalla, aparece la estructura retórica de las notas de cata de aceite de oliva en lengua inglesa (ilustración 4). El usuario puede utilizar esta estructura para acceder a las unidades de información que desea incluir en su nota de cata, aunque también puede desplazarse arriba y abajo en la página web para acceder a ellas.

Al hacer click en una de las unidades de información, la pantalla traslada al usuario directamente a dicha sección (ilustración 5). Aun así, el resto de las unidades siguen visibles y disponibles en la web en caso de que el usuario quiera incluirlas en su nota de cata, aunque no es necesario incluir todas ellas.

Al seleccionar una unidad de información en el modo Básico, por ejemplo, Aroma, el usuario debe hacer click en el icono del lápiz para editar dicha unidad e incluir la información relativa a su muestra de aceite en concreto. Al tratarse del modo Básico, la complejidad lingüística es menor y se muestran líneas modelo sin verbo finito (ilustración 6).

Si el usuario selecciona una unidad de información en el modo Avanzado, se desplegarán líneas modelo más complejas lingüísticamente, incluyendo verbos finitos (ilustración 7).

Una vez seleccionado el modo de funcionamiento y la primera unidad de información que el usuario desea incluir en su nota de cata, debe procederse a editar la línea modelo. Para ello, el usuario debe seleccionar una línea modelo y sustituir la palabra o palabras entre paréntesis con descriptores que se ajusten a las propiedades particulares del aceite de oliva que se esté evaluando (ilustración 8).

En caso de no conocer el término en lengua inglesa que el usuario desee utilizar para personalizar una línea modelo, puede desplazarse hasta la parte superior de la web y hacer click en el botón Glosario (ilustración 9).

Ilustración 4.

Segundo paso: selección de unidad de información a incluir

- Variedad + D.O.
- Localización
 - Clima
 - Suelo
 - Geografía
- Proceso de elaboración
 - Cosecha
 - Prensado
 - Fecha
 - Temperatura
- Nota de cata
 - Aroma
 - Aroma
 - Madurez
 - Intensidad
 - Comentario general
 - Apariencia
 - Color
 - Densidad
 - Transparencia
 - En boca
 - Flavor
 - Acidez
 - Picor
 - Amargor
 - Dulzor
 - Sabor
 - Equilibrio
 - Intensidad
 - Textura
 - Persistencia
 - Entrada
 - Final
 - Retrogusto
 - Comentario general
- Información técnica
 - Acidez / pH
 - Polifenoles
 - Ácidos grasos Omega-3 y -6
- Maridaje


Ilustración 5.

Selección y vista de las unidades de información


Nota de cata

Aroma

Aroma



Madurez



Intensidad




Ilustración 6.


Modo básico

Nota de cata


Aroma

Aroma

(Aroma) aroma.
Hints of (aroma).



Madurez



Intensidad




Ilustración 7.

Modo avanzado

Nota de cata

Aroma

Aroma

The oil has aromas of (aroma).

Its aroma is (aroma).

Madurez

Intensidad

Ilustración 8.

Edición de una línea modelo

Apariencia

Color

The oil has a golden green color.

Densidad

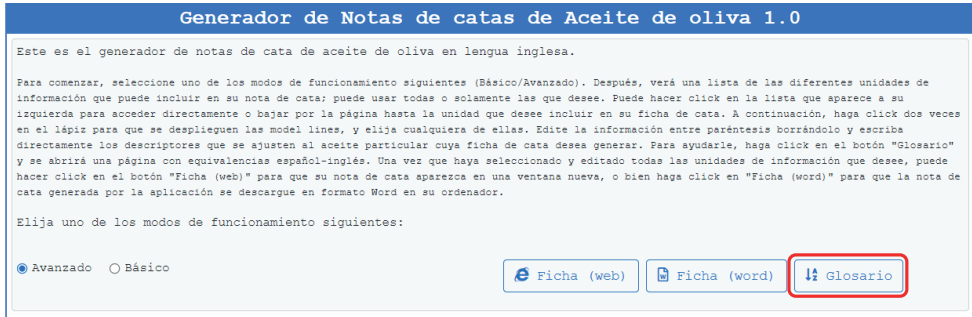
It has a fluid consistency.

Transparencia

Color is (transparencia).

Ilustración 9.

Botón glosario



Al hacer click en el botón Glosario, se abrirá una nueva pestaña que no altera el trabajo realizado hasta el momento, que permanece tal y como se dejó en la pestaña principal. Tras acceder al glosario (ilustración 10) aparecen dos columnas con diversos descriptores. En la columna de la izquierda aparecen los términos en español en orden alfabético. En la columna de la derecha aparecen los equivalentes en lengua inglesa.

Ilustración 10.

Vista del glosario

Glosario	
Español	Equivalente en inglés
aceituna	olive
achicoria	chicory
ácido	acid, sour
ácido, fuerte	tangy
agradable	pleasant
aguacate	avocado
albahaca	basil
alcachofa	artichoke

Para encontrar el equivalente en inglés del descriptor que el usuario desea incluir, puede bajar por la página hasta encontrarlo en la columna izquierda. También puede utilizar el buscador de su navegador, introducir el descriptor en español y pulsar Buscar. Así, el navegador le dirigirá directamente al descriptor, y a su derecha, el usuario encontrará el equivalente en inglés (ilustración 11). El usuario deberá seleccionar y copiar dicho equivalente antes de cerrar la pestaña del Glosario.

Ilustración 11.

Búsqueda de equivalentes en el Glosario

Descriptor (Español)	Equivalente (Inglés)
sucia	subtira
textura	texture
tomate	tomato
tomillo	thyme
tropical	tropical
turbio	cloudy
vainilla	vanilla
variedad	variety, varietal
verde	green
verde claro	light green

Al volver a la página principal, el usuario debe regresar a la unidad de información que estaba editando (a través del índice en la izquierda de la web o bajando por la página hasta encontrarla). Una vez en la unidad de información en la que estaba trabajando, deberá seleccionar la palabra entre paréntesis y pegar el descriptor que copió del glosario (ilustración 12). Así, la línea modelo en cuestión quedaría completa.

Este proceso de selección de unidades de información y edición de líneas modelo debe repetirse hasta que el usuario esté satisfecho con la información incluida. Una vez que ya no haya más información sobre la cata de la muestra de aceite del usuario, este debe proceder a completar la nota de cata final personalizada para su producto. Para ello, el usuario debe acceder a la parte superior de la web, donde encontrará dos botones (ilustración 13): Ficha (web) y Ficha (Word).

Si el usuario hace click en Ficha (web), se abrirá una nueva pestaña que mostrará la nota de cata redactada (ilustración 14) y dividida en tres secciones principales: *About*, con información sobre los tres primeros y el penúltimo movimientos retóricos del género (Variedad de aceituna + D.O., Localización; Proceso de elabora-

Ilustración 12.

Inclusión de un descriptor del Glosario en una unidad de información

Apariencia

Color

The oil has a golden green color.

Densidad

It has a fluid consistency.

Transparencia

Color is cloudy.

Ilustración 13.

Opciones para obtener de la nota de la cata finalizada

Generador de Notas de catas de Aceite de oliva 1.0

Este es el generador de notas de cata de aceite de oliva en lengua inglesa.

Para comenzar, seleccione uno de los modos de funcionamiento siguientes (Básico/Avanzado). Después, verá una lista de las diferentes unidades de información que puede incluir en su nota de cata; puede usar todas o solamente las que desee. Puede hacer click en la lista que aparece a su izquierda para acceder directamente o bajar por la página hasta la unidad que desee incluir en su ficha de cata. A continuación, haga click dos veces en el lápiz para que se desplieguen las model lines, y elija cualquiera de ellas. Edite la información entre paréntesis borrándolo y escriba directamente los descriptores que se ajusten al aceite particular cuya ficha de cata desea generar. Para ayudarle, haga click en el botón "Glosario" y se abrirá una página con equivalencias español-inglés. Una vez que haya seleccionado y editado todas las unidades de información que desee, puede hacer click en el botón "Ficha (web)" para que su nota de cata aparezca en una ventana nueva, o bien haga click en "Ficha (word)" para que la nota de cata generada por la aplicación se descargue en formato Word en su ordenador.

Elija uno de los modos de funcionamiento siguientes:

Avanzado Básico

Ficha (web) Ficha (word) Glosario

ción e Información técnica); *Tasting Note*, con el movimiento Nota de cata; y, finalmente *Pairing*, con Maridaje, el último movimiento retórico identificado.

El usuario también tiene la opción de obtener su nota de cata en formato Microsoft Word pinchando en el botón Ficha (Word). En este caso, se descargará en el ordenador del usuario un documento Word, que el usuario podrá abrir (ilustración 15) y seguir editando en caso de que quiera cambiar el formato, los colores, o cualquier otro aspecto de la nota de cata.

Ilustración 14.

Vista de la nota de cata completada a través del botón Ficha (web)


About

This oil is a blend of Picual and Mission olives.

The warm climate and rich soil of our area are responsible for its unique taste.

The olives are handpicked by October and cold pressed within hours at our mill.

It is an olive oil rich in phenols. It is high in phenolic antioxidants, vitamins E and K as well as Omega 3 and 6.



Tasting Note

The oil has a golden green color. It has a fluid consistency. Color is cloudy.

The acidity is well balanced. Pungency is distinct.

The oil displays flavors of tomato and figs. The flavour is beautifully balanced. It is intense. It has a long-lasting persistence. The mouth opens with tomato and vanilla. The oil has a spicy aftertaste. The taste is mouth-watering.

Pairing

It pairs perfectly with green salads, roasted vegetables and even desserts like ice cream.


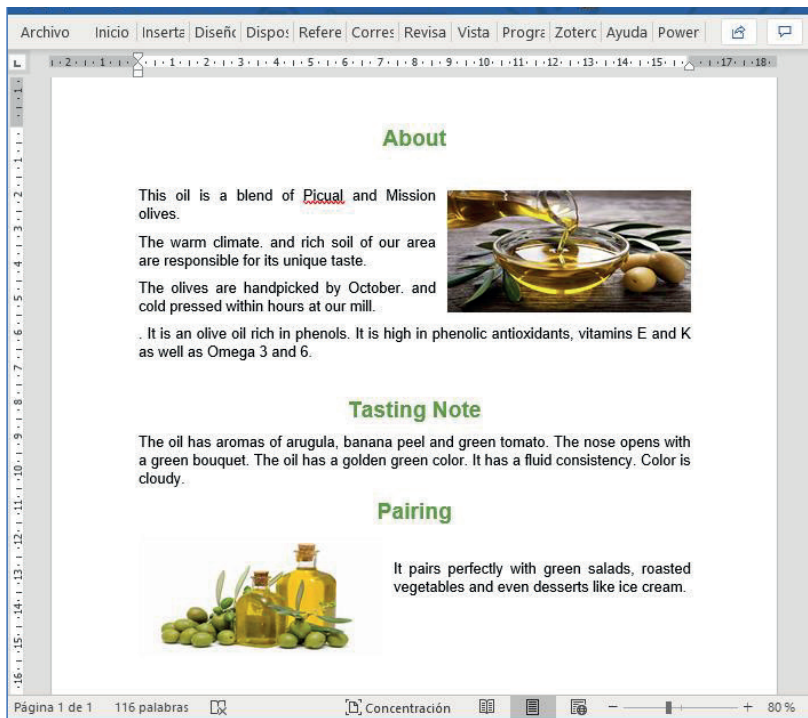


Ilustración 15.

Vista de la nota de cata completada a través del botón Ficha (Word)



Archivo Inicio Insertar Diseñar Disposición Referencias Corrección Revisión Vista Programación Zotero Ayuda Power


About

This oil is a blend of Picual and Mission olives.

The warm climate and rich soil of our area are responsible for its unique taste.

The olives are handpicked by October and cold pressed within hours at our mill.

It is an olive oil rich in phenols. It is high in phenolic antioxidants, vitamins E and K as well as Omega 3 and 6.




Tasting Note

The oil has aromas of arugula, banana peel and green tomato. The nose opens with a green bouquet. The oil has a golden green color. It has a fluid consistency. Color is cloudy.

Pairing

It pairs perfectly with green salads, roasted vegetables and even desserts like ice cream.



Página 1 de 1 116 palabras Concentración 80%

Conclusión

A través de los resultados obtenidos en este trabajo de investigación, las notas de cata de aceite de oliva en lengua inglesa pueden tipificarse como un género y una LFE que incluyen los términos y unidades fraseológicas aquí identificados y pertenecientes a los campos semánticos clasificados. Además, dichos términos y unidades fraseológicas forman parte de oraciones léxicamente densas con un número reducido de verbos, normalmente en 3ª persona singular o en forma no finitas y seguidos de sintagmas nominales, adjetivales y adverbiales muy condensados. Un texto de estas características además deberá, para ser considerado parte del género, hacer alusión al menos al sabor e incluso al aroma del aceite, aparte de incluir cualquiera de las otras unidades de información retórica. Asimismo, las notas de cata constituyen un género evaluativo y descriptivo predominantemente subjetivo cuya finalidad es expresar las percepciones sensoriales del catador sobre una muestra particular de aceite, lo que liga de forma indisoluble el género con los campos semánticos aquí identificados y la realidad que representan. Por tanto, este es un género orientado al contenido y cuyo peso funcional recae sobre la manera en la que se organiza la información a través de su estructura retórica, que es flexible pero bien definida.

Esta descripción multidimensional del género y su LFE a todos los niveles lingüísticos no sólo tiene importancia académica, ya que no se había analizado empíricamente hasta ahora; también ha permitido la identificación y clasificación de los patrones lingüísticos que definen al género y su LFE, y su aplicación al desarrollo del Generador 1.0 que ayuda a su redacción, de forma que los textos resultantes son ejemplares válidos del género al servicio de los profesionales españoles que necesiten poner en valor su producto en lengua inglesa sin tener la suficiente competencia en inglés. Además de eficaz, se ha conseguido desarrollar como una opción fiable, empírica, económica y sencilla para aquellas empresas interesadas en abrir su producción al mercado internacional. Esto es de especial relevancia en sectores con el potencial internacional y la necesidad de desarrollarlo que tiene el del aceite de oliva jiennense. En el futuro, el Generador deberá ser testado en

almazaras y empresas del sector oleícola para recibir retroalimentación sobre su uso en entornos reales y así poder desarrollar mejoras en secciones localizadas del mismo de acuerdo a la experiencia de los usuarios. En todo caso, este trabajo ha demostrado la adecuación de la lingüística aplicada y de corpus al desarrollo de herramientas que, en combinación con las nuevas tecnologías, ayuden a mejorar el rendimiento profesional de los usuarios de las LFE de cara a la promoción internacional de un producto y su sector.

Referencias

- ¿Qué Es El 'Oleoturismo'? (2017). *Aceites De Oliva De España*. www.aceitesdeolivadeespana.com/diario-del-aceite/que-es-el-oleoturismo.
- ACTRES - *Generator of Wine Tasting Notes*, (2008). <http://contraste2.unileon.es/web/es/index.html>
- Anthony, L. (2018). AntConc Homepage. *Laurence Anthony's AntConc*. www.laurenceanthony.net/software/antconc/.
- Bhatia, V. K. (1993). *Analysing Genre. Language Use in Professional Settings*. New York: Routledge.
- Bhatia, V. K. (2004). *Worlds of Written Discourse, A Genre-Based View*. New York: Continuum.
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243-257.
- Biber, D., Connor U., & Upton T.A. (2007). *Discourse on the Move. Using Corpus Analysis to Describe Discourse Structure*. Amsterdam: John Benjamins Publishing Company.
- Bowker, L. & Pearson, J. (2002). *Working with Specialized Language. A Practical Guide to Using Corpora*. London: Routledge.
- Caballero, R. (2017). From the glass through the nose and the mouth: Motion in the description of sensory data about wine in English and Spanish. *Terminology*, 23(1), 66-88.
- Cabré, M. T. (1999). *Terminology. Theory, Methods and Applications*. Amsterdam: John Benjamins Publishing Company.
- Calvi, M. V., Bordonaba Zabalza, C., Mapelli, G., & Santos López, J. (2009). *Las Lenguas de Especialidad en Español*. Roma: Carocci editore.
- Cifras Aceites De Oliva. (2019). *International Olive Council*. www.internationaloliveoil.org/estaticos/view/131-world-olive-oil-figures?lang=es_ES.
- Consejo Económico y Social de la Provincia de Jaén (CES). (2020). *Memoria 2019 sobre la situación socioeconómica y laboral de la provincia de Jaén*. <https://www.dipujaen.es/export/sites/default/galerias/galeriaDescargas/diputacion/dipujaen/CES/otras-image->

nes/RESUMEN_CES2019_-_memoria_situacion_socioeconomica_y_laboral_-_completa.pdf

- Corpas Pastor, G., & Seghiri Dominguez, M. (2009). Virtual corpora as documentation resources: translating travel insurance documents. In: A. Beeby, P. Rodríguez., & P. Sánchez Gijón (Eds.), *Corpus use and translating* (pp. 75-107). Antwerp: John Benjamins.
- Corpas Pastor, G., Seghiri Dominguez, M., & Maggi, M. (2007). Recor. Software. <http://www.lexytrad.es/es/recursos/recor-2/>.
- Drouin, P. (2010). TermostaT Users Guide. http://termostat.ling.umontreal.ca/doc_termostat/doc_termostat.html
- Dubois, D. (2007). From Psychophysics to Semiophysics: Categories as Acts of Meaning. A Case Study from Olfaction and Audition back to Colors. In John Benjamins Publishing Company's *Speaking of colors and odors*, 167-185.
- Dudley-Evans, T. & St John, M. J. (1998). *Developments in English for Specific Purposes. A Multi-Disciplinary Approach*. Cambridge: Cambridge University Press.
- Extra Virgin Olive Oil. (2016). *California Olive Oil Council*. www.cooc.com/.
- Free USAS English Web Tagger. (2016). *USAS Online English Tagger*. ucrel-api.lancaster.ac.uk/usas/tagger.html.
- Fundación Estrategias para el Desarrollo de la Provincia de Jaén. (2020). *II Plan Estratégico de la Provincia de Jaén*. https://www.planestrajaen.org/export/sites/default/galerias/galeriaDescargas/plan-estrategico/libros/II_Plan_Estratxgico_de_la_provincia_de_Jaxn.pdf
- Fundación Estrategias para el Desarrollo de la Provincia de Jaén. (2021). *Observatorio Económico de la Provincia de Jaén*. https://planestrajaen.org/export/sites/default/galerias/galeriaDescargas/plan-estrategico/otras-publicaciones/OBSERVATORIO_ECONOMICO_298.pdf
- Galeote López, M. (1992). El Vocabulario del Olivar en el Sur de Córdoba. *Boletín de la Real Academia de Córdoba de Ciencias, Bellas Letras y Nobles Artes*, (123), 277-300.
- Garside, R., Leech G., & McEnery, T. (1997). *Corpus Annotation. Linguistic Information from Computer Text Corpora*. New York: Addison Wesley Longman Limited.
- Gläser, R. (1994). Relations between Phraseology and Terminology with Special Reference to English. *ALFA*, (7/8), 41-60.
- Gobierno de España. (2019). *Anuario de Estadísticas Agrarias y Pesqueras*. https://www.mapa.gob.es/estadistica/pags/anuario/2019/CAPITULOS_TOTALES/AE19-C07.pdf
- González Blanco, A. (1999). Léxico y Realidad en el Mundo de las Almazaras de la Región Murciana. *Revista Murciana de Antropología*, (1), 91-97.
- Junta de Andalucía. (2019). *Anuario de Estadísticas Agrarias y Pesqueras*. <https://www.juntadeandalucia.es/organismos/agriculturaganaderiapescaydesarrollosostenible/servicios/estadistica-cartografia/anuarios.html>
- Klein, D. & Manning, C.D. (2003). Fast Exact Inference with a Factored Model for Natural Language Parsing. *Advances in Neural Information Processing Systems 15 (NIPS 2002)*, 3-10.

- L'Homme, Marie-Claude (2009). "A Methodology for Describing Collocations in a Specialized dictionary." In *Lexicography in the 21st century* ed. By Sandro Nielsen and Sven Tarp, 237-256. Amsterdam/Philadelphia: John Benjamins.
- López Arroyo, B. & Roberts R.P. (2016). Differences in Wine Tasting Notes in English and Spanish. *Babel*, 62(3), 370-401.
- Marín Pérez, M. J. (2016). Measuring the Degree of Specialisation of Sub-Technical Legal Terms through Corpus Comparison: A Domain-Independent Method. *Terminology*, 22(1), 80-102.
- Méndez Cendón, B. (2009). Combinatorial Patterns in Medical Case Reports: an English-Spanish Contrastive Analysis. *The Journal of Specialised Translation*, (11), 169-190.
- Montoro del Arco, E. T. & Roldán Vendrell, M. (2013a). Las Denominaciones del Aceite de Oliva en los Reglamentos Oficiales: Evolución, Normalización y Problemas Terminológicos. *Comunicación y Transmisión del Saber entre Lenguas y Culturas*, 265-282.
- Montoro del Arco, E. T. & Roldán Vendrell, M. (2013b). Terminología, Normalización y Comunicación: Las Categorías de Aceite de Oliva en Español, Inglés y Chino. *Terminology: International Journal of Theoretical and Applied Issues in Specialized Communication*, 19(1), 62-92.
- Montoro del Arco, E. T. (2012). La Fraseología Especializada del Área Científica y Socioeconómica del Olivar y el Aceite de Oliva: Entre la Tradición y la Modernidad. *Phraseology and Discourse*. A. Pamies, J. M. Pazos Breña, & L. Luque Nadal (Eds.), 338-345.
- Moya Corral, J. A. (1994). Sobre el Léxico del Olivo y la Almazara. *Lexicografía y Enseñanza de la Lengua Española*. J. A. Moya Corral & M. Sosinski (Eds.), 259-272.
- Muegge, U. (2012). 10 Things You Should Know About Automatic Terminology Extraction. *The American Translators Association Chronicle*, 41(9), 24-28.
- Orgaz Agüera, F., Moral Cuadra, S., López Guzmán, T., & Cañero Morales, P. (2017). Estudio de la Demanda Existente en Torno al Oleoturismo. El Caso de Andalucía. *Cuadernos de Turismo*, (39), 437-453.
- Paradis, C. (2005). Ontologies and Construals in Lexical Semantics. *Axiomathes*, (5), 541-573.
- Paradis, C. (2010). Touchdowns in Winespeak: Ontologies and Construals in Use and Meaning-Making. In M. Goded Rambaud, & A. Poves Luelmo (Eds.) *Proceedings from the 1st Conference on Linguistic Approaches to Food and Wine Descriptions*. Madrid: UNED.
- Piao, Scott, Francesca Bianchi, Carmen Dayrell, Angela D'Egidio and Paul Rayson. 2015. "Development of the Multilingual Semantic Annotation System." *Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015)*, Denver, Colorado, United States, 1268-1274.
- Ramón, N., & Labrador, B. (2015). The Rhetorical Structure of Technical Brochures: A Proposal for Technical Writing. *32nd International Conference of the Spanish Association of Applied Linguistics (AESLA): Language Industries and Social Change*. Sevilla: Elsevier.
- ReCor. (2016). *LEXYTRAD*. www.lexytrad.es/es/recursos/recor-2/.

- Rey, A. (1995). *Essays on Terminology*. (J. C. Sager, Trans.). Amsterdam: John Benjamins Publishing Company.
- Roberts, R. P. (1998). Phraseology and Translation. *La Traducción: Orientaciones Lingüísticas y Culturales*. P. Fernández Nistal & J. M. Bravo Gozalo (Eds.), 61-78.
- Roldán Vendrell, M. & Fernández Domínguez, J. (2012). Emergent Neologisms and Lexical Gaps in Specialised Languages. *Terminology*, (18), 9-26.
- Roldán Vendrell, M. (2007). Unidades de Conocimiento Especializado en Textos del Ámbito de la Oleicultura Antigua: Aspectos Sintagmáticos. *Estudios lingüísticos, Literarios e Históricos. Homenaje a Juan Martínez Marín*. G. Águila Escobar, P. Barros García & E. T. Montoro del Arco (Eds.), 327-336.
- Roldán Vendrell, M. (2010b). Olivaterm: Vocabulario Multilingüe del Aceite de Oliva. *El Español en Contexto. Actas de las XV Jornadas sobre Lengua Española y su Enseñanza*. E. T. Montoro del Arco & J. A. Moya Corral (Eds.), 312-319.
- Roldán Vendrell, M. (2010a). *Bases para la Terminología Multilingüe del Aceite de Oliva*. Granada: Comares.
- Roldán Vendrell, M. (2013). *Diccionario de Términos del Aceite de Oliva*. Madrid: Arco/Libros S.L.
- Santa María, S. (2013). The Terminology of Olive Oil Taste Testing: A Bilingual (Spanish-English and English-Spanish) Glossary. *Translation Journal*, 17(2).
- Seghiri, M. (2015). Determinación de la representatividad cuantitativa de un corpus ad hoc bilingüe (inglés-español) de manuales de instrucciones generales de lectores electrónicos. In M. T. Sánchez Nieto (Ed.), *Corpus-Based translation and interpreting studies: From description to application* (pp. 125-146). Berlin: Frank & Timme.
- Seghiri, M. (2016). *Compilación de un Corpus Trilingüe de Seguros Turísticos (Español-Inglés-Italiano): Aspectos de Evaluación, Catalogación, Diseño y Representatividad*. PhD Thesis. Málaga: Universidad de Málaga.
- Sinclair, J. M. (1994). Corpus Typology: A Framework for Classification. G. Melcher & B. Warren (Eds.), *Studies in Anglistics*. Stockholm: Almqvist and Wiksell International, 17-34.
- Suter, H. J. (1993). *The Wedding Report, A Prototypical Approach to the Study of Traditional Text Types*. Amsterdam: John Benjamins Publishing Company.
- Swales, J. M. (1990). *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.
- The Stanford Parser: A Statistical Parser. (2018). *The Stanford Natural Language Processing Group*. nlp.stanford.edu/software/lex-parser.shtml.
- Torres Quesada, G. J. (2011). El Léxico del Olivo y la Almazara en Sierra Mágina. *Revista de Estudios sobre Sierra Mágina* (29), 105-141.
- Un 70% de visitas a almazaras con Oleoturismo son extranjeras, especialmente EEUU y Centroeuropa. (2018). *La Vanguardia*. www.lavanguardia.com/local/sevilla/20180311/441446194639/masjaen-turismo-un-70-de-visitas-a-almazaras-con-oleoturismo-son-extranjeras-especialmente-eeuu-y-centroeuropa.html.

Índice

	Pág.
Resumen	9
1. Introducción	11
2. Estado de la cuestión	15
2.1. Lenguajes para Fines Específicos (LFE)	15
2.2. Género textual	17
2.3. Comunicación: aceite de oliva y lenguaje	18
3. Material y metodología	21
3.1. Material	21
3.2. Metodología	26
3.2.1. Análisis léxico	26
3.2.2. Análisis sintáctico-gramatical	27
3.2.3. Análisis semántico	28
3.2.4. Análisis retórico	28
4. Resultados	31
4.1. Nivel léxico	31
4.2. Nivel sintáctico-gramatical	33
4.3. Nivel semántico	35
4.4. Nivel retórico	39
4.5. Aplicación de los resultados a la creación del Generador 1.0	43
5. Generador de Notas de Cata de Aceite de Oliva	45
5.1. Uso del Generador de Notas de Cata de Aceite de Oliva 1.0	47
6. Conclusión	57
Referencias	59

