

## Una breve exploración de la terminología amorosa en los *corpora catullianum, tibullianum* y *propertianum* con métodos y herramientas computacionales: etiquetado gramatical, lemas, bigramas y co-apariciones

*A brief exploration of the amorous terminology in the Catullan, Tibullan, and Propertian corpora using computational methods and tools: Part-of-speech Tagging, lemmas, bigrams, and co-occurrences*

Carlos Javier NUSCH  
PREBI-SEDICI, Universidad Nacional de La Plata  
[carlosnusch@prebi.unlp.edu.ar](mailto:carlosnusch@prebi.unlp.edu.ar)  
<https://orcid.org/0000-0003-1715-4228>

A María Delia Buisel,  
por haberme inspirado.  
*Et rursus in bellum resorbens*  
*Mars me fretis tulit aestuosus*  
[Versos modificados por el autor del artículo]

### RESUMEN

Este estudio expone los primeros análisis exploratorios en la poesía de Catulo, Propertio y Tibulo, basándose en etiquetado gramatical, lemas, bigramas y co-apariciones. Se emplearon técnicas de análisis de texto automático y herramientas computacionales mediante R Studio. Los resultados, presentados en tablas y gráficos, revelaron varios fenómenos. El análisis léxico destacó formas pronominales de primera y segunda persona y términos específicos del léxico poético elegíaco. El etiquetado gramatical mostró términos familiares, especialmente en poemas epitalámicos. Los bigramas resaltaron contrastes temporales y espaciales en el discurso amoroso, mientras que los grafos de bigramas destacaron adjetivos asociados al amor con rasgos positivos y negativos. Las co-ocurrencias subrayaron términos amorosos, pronombres y posesivos, y verbos en pasado que indican amor concluido. Se plantea la hipótesis de la co-dependencia de términos pronominales y elementos de la deixis temporal y espacial como un enfoque promisorio para la tipificación de la poesía elegíaca latina.

### PALABRAS CLAVE

Poetas elegíacos latinos, tópicos amorosos, poesía epitalámica, análisis estilométrico, Procesamiento del Lenguaje Natural (PLN).

### ABSTRACT

This study presents the first exploratory analyses in the poetry of Catullus, Propertius, and Tibullus, based on grammatical tagging, lemmas, bigrams, and co-occurrences. Automatic text analysis techniques and computational tools were employed using R Studio. The results, presented in tables and graphs, revealed various phenomena. Lexical analysis highlighted first and second-person pronominal forms and specific terms from the elegiac poetic lexicon. Grammatical tagging showed familiar terms, especially in epithalamic poems. Bigrams emphasized temporal and spatial contrasts in love discourse, while bigram graphs highlighted adjectives associated with love expressing positive and negative traits. Co-occurrences underscored terms related to love, personal pronouns, possessives, and past tense verbs denoting concluded love. The hypothesis of the co-dependence of pronominal terms and elements of temporal and spatial deixis is proposed as a promising approach to typify Latin elegiac poetry.

### KEYWORDS

Latin Elegiac Poets, Amatory Motifs, Epithalamic Poetry, Stylometric Analysis, Natural Language Processing (NLP).

#### Dirección

Clara Martínez  
Cantón

Gimena del Río  
Riande

Francisco Barrón

#### Editor asociado

Rubén Iñiguez  
Pérez

## 1. L'AMOR DE LONH: LA POESÍA AMOROSA Y LA LECTURA DISTANTE

En trabajos anteriores (Nusch, 2020, 2021, 2022) se han comentado las consideraciones que C. S. Lewis (1936) hizo acerca del amor cortés y la literatura occitana en su famoso libro *The Allegory of Love*. En ese trabajo el estudioso británico arguye que la idea del amor aún hoy vigente en Occidente nace en Occitania en el siglo XI con apenas una discreta influencia aportada por la tradición anterior. También se marcaron someramente en esas investigaciones algunas notables similitudes entre los tópicos amorosos, el tratamiento de la amada y ciertos términos provenientes de los campos político y militar que aparecen en los poemas eróticos, tanto en la literatura occitana como en la literatura latina del siglo I a.C. Se propuso entonces abordar el trabajo comparativo no solo utilizando las tradicionales técnicas de lectura cercana sino el potencial de las herramientas que aporta la noción de lectura distante.

En el ámbito de las Humanidades, la posibilidad de analizar grandes volúmenes textuales con herramientas computacionales se ha venido realizando y discutiendo desde mediados del siglo XX. Hace más de una década ya que Stephen Ramsay (2011) postuló la idea de que las computadoras podían ser consideradas como máquinas de lectura, capaces de analizar grandes cantidades de texto de manera rápida y eficiente con la finalidad de revelar patrones y conexiones en los textos que podrían pasar desapercibidos para los lectores humanos. Dos años después, Franco Moretti cuestionó el concepto de canon en su libro *Distant Reading* (2013) al preguntarse cómo confiar en que un conjunto de obras que constituyen el 1 % de lo que se ha publicado en su tiempo y que fueron seleccionadas por humanos sea realmente representativo del total de la literatura de su época. Ante el enorme poder computacional de las máquinas, ¿tiene sentido dedicarse a la lectura cercana si es posible utilizar el apabullante poder abarcador de las técnicas de la lectura distante? Aunque la indagación de Moretti parece más una provocación que un verdadero dilema, actualmente no parece posible concebir un investigador que no evalúe seriamente la utilización de herramientas digitales y cómo estas podrían contribuir a mejorar el alcance de sus indagaciones. Sin embargo, no debe cometerse el error de pensar que las técnicas de lectura distante vienen a reemplazar las tradicionales prácticas de lectura cercana. El rol de estas nuevas herramientas es el de complementar o asistir el trabajo del humanista y el desafío de cualquier trabajo que las contemple consiste en detectar cuál será la técnica adecuada y qué metodología o enfoque utilizar para ampliar correctamente el horizonte de su trabajo.

El presente artículo describe las primeras exploraciones estilométricas realizadas sobre tres poetas romanos del siglo I a.C. (Cayo Valerio Catulo, Albio Tibulo y Sexto Propercio), en el marco del proyecto Aetates Amoris<sup>1</sup>. En este acercamiento inicial, se intenta delimitar un conjunto de términos, de conceptos y características propias de los poetas elegíacos romanos para poder compa-

---

<sup>1</sup> Accesible desde: <http://aetatesamoris.com.ar>.

rarlos con un conjunto semejante de términos y conceptos de los poetas occitanos en trabajos de investigación futuros. En los apartados siguientes se procederá al análisis distribucional de los términos amorosos con la finalidad de detectar patrones y conceptos importantes en la construcción del imaginario amoroso. Para realizar dicha tarea se seleccionarán cuatro tipos de características: clases de palabras, lemas, bigramas y co-apariciones.

## 2. METODOLOGÍA

En este apartado se precisan algunas características y particularidades propias del imaginario amoroso griego y romano en la Antigüedad con la finalidad de explicar brevemente conceptos y términos que luego se buscarán en los textos por medio de los diferentes métodos de análisis automático.

Se han seleccionado cuatro tipos de características (clases de palabras, lemas, bigramas y co-apariciones-ocurrencias) para la captura del vocabulario amoroso elegíaco que se busca tipificar y comparar. Con la finalidad de realizar estas tareas se preparó el entorno de trabajo en R Studio y se utilizaron las librerías *rvest* en la cosecha de textos, y *tidyverse* y *tidytext* (Silge y Robinson, 2017; Wickham, 2022) en el análisis textual automático. Las clases de palabras proporcionan un insumo útil porque la representación del erotismo cuenta con sustantivos, verbos y adjetivos que le son propios. La reducción del vocabulario a lemas intenta simplificar y concentrar los términos amorosos eliminando accidentes gramaticales para identificar más fácilmente ciertos núcleos semánticos presentes en la poesía erótica. La búsqueda de bigramas responde al hecho de que muchas expresiones típicas del amor elegíaco latino se expresan por medio de pares de palabras: *puella divina* (muchacha divina), *servitium amoris* (literalmente *servicio de amor* pero entendido como esclavitud del amor), *Cintia mea* (mi Cintia), *Lesbia illa* (aquella Lesbia). En el armado de grafos se utilizaron las librerías *igraph*, *ggraph* y *grid* (Csárdi et al., 2023; Murrell, 2005; Pedersen, 2022). Las co-apariciones resultan importantes para la detección de tópicos o temas, ya que determinados términos nucleados alrededor de un mismo asunto tienden a co-aparecer en uno o diferentes textos. Para la comparación por pares (*pairwise*) se utilizó la librería *widyr* (Robinson et al., 2022). Esta última característica es el elemento principal para futuros trabajos centrados en el modelado de tópicos.

En la totalidad del artículo se realizan análisis automáticos y cuantitativos que luego se representan por medio de tablas o gráficos. Como se trata de una investigación en etapa inicial, se realizan algunas observaciones y descripciones de los resultados obtenidos en el intento de explicar la pertinencia de ciertos términos para la determinación de un vocabulario amoroso elegíaco. Las características analizadas serán utilizadas en futuras tareas de comparación de léxicos, modelado de tópicos, y *clustering* de documentos, entre otras. En el apartado final se ensayan algunas conclusiones y se proponen trabajos futuros y nuevas herramientas con el objeto de mejorar la precisión y alcances en el estudio de los vocabularios de los tres poetas.

### 3. UNA APROXIMACIÓN AL LÉXICO AMOROSO ANTIGUO

Antes de entrar de lleno en el análisis cuantitativo resulta necesario hacer una serie de precisiones acerca del léxico amoroso. La Antigüedad Clásica consideraba al amor y al estado de enamoramiento como algo de carácter divino, que estaba emparentado tanto con la enfermedad como con la locura. El sentimiento de amor, al ser infundido por un dios, otorgaba prestigio, pero a su vez debía ser tan temido como respetado. Esto era para los griegos el concepto de ἔρως, del deseo amoroso vinculado a lo sexual que se concebía “como una atracción casi automática, de base divina y cósmica, que experimenta un individuo hacia otro” (Rodríguez Adrados, 1996, p. 10). Este sentimiento de amor-pasión se manifestaba a través de una sintomatología precisa en la que el deseo oprimía al φρήν, al *diafragma*, al καρδιά, al corazón y al στῆθος, al pecho (Calame y Pérez Rodríguez, 2002, pp. 35-36)<sup>2</sup>. La familia de palabras del amor en el mundo latino también tenía sus peculiaridades: *amo* era más ambiguo que ἔράω ya que comportaba el significado de *hacer el amor*, *estar enamorado*, *tener una amante*, o bien *amar* tanto en el sentido de amor pasional como el de amistad o cariño.

En el mundo latino, *amo* se oponía a *diligo* que también significa querer o apreciar, pero basado en la elección y la reflexión. El término *amor* (el sentimiento o su personificación) abarcaba el campo erótico pero se utilizaba también en el sentido de amistad. Esta imprecisión se transfería a sus palabras derivadas: *amicus* designaba al amigo, mientras que *amica* siempre era usado para referirse a una amante (Ernout y Meillet, 2001, p. 29; Harley, 2012, p. 85). La constelación de conceptos del campo amoroso latino también compartía significados del campo político o militar: el término *amicitia* que correspondía a *amistad*, también comportaba otros sentidos especializados como el de *alianza* en el terreno político y propiciaba toda una serie de alusiones poéticas y eróticas. También del ámbito político provenía *fides* con el sentido de *confianza*, *crédito*, *lealtad*, *palabra dada*, y ambos términos aparecían en los poemas de amor junto con *foedus*, *tratado de alianza*, *pacto* (Ernout y Meillet, 2001, p. 232, p. 243; Harley, 2012, p. 85). Las cláusulas de este *foedus* tenían fuerza de *lex amatoria*, *ley de amor* y en él los dioses actuaban de testigos frente a juramentos al estilo de los *auctorati*, los ciudadanos libres que se enrolaban como gladiadores (Ramírez de Verger, 1986, p. 78).

A la par de los términos específicos, la poesía amorosa latina practicada por los poetas elegíacos desarrolló ciertos tópicos fijos que tuvieron la capacidad de agrupar conjuntos temáticos y situaciones con un alto grado de sofisticación y codificación, como el *Amor puellae visae*, el *Servitium amoris*, la *Militia amoris*, la *Flamma amoris*, el *Morbus amoris*, etc. A su vez, los poetas

<sup>2</sup> Además, los griegos antiguos concebían a φιλέω, un verbo amoroso, como opuesto al concepto de amor-pasión (Rodríguez Adrados, 1996, p. 10), que carece de componente sexual la mayor parte de las veces y que implica cierta camaradería, amistad, cariño, amor filial (Rodríguez Adrados, 1996, p. 31).

elegíacos desarrollaron ciertas escenas y núcleos narrativos alrededor de personajes y situaciones tópicas. El personaje por antonomasia era el de la amada elegíaca: Catulo tenía a Lesbia, Galo a Licoris, Tibulo a Delia y a Némesis, Propercio a Cintia, y Ovidio a Corina. Estas amadas eran homenajeadas y queridas con una lealtad a toda prueba y por ellas los enamorados declaraban ser capaces de toda clase de sacrificios y humillaciones.

Pero ese amor tan alabado y cantado por los poetas antiguos tenía otras connotaciones y repercusiones en los vericuetos de la vida pública. El amor en estado puro, ἔρως como sentimiento individual se consideraba peligroso no solo para el individuo sino también para la vida social (Grimal, 2000, p. 170). Su campo de acción debía ser controlado y se reservaba para las heteras y los efebos en una sociedad voluntariamente deserotizada (Rodríguez Adrados, 1996, p. 72). La sexualidad de los esposos permanecía oculta y no fue hasta la época helenística cuando empezó a aparecer como tema literario en la cultura griega (Rodríguez Adrados, 1996, p. 52). Esa separación o concepción compartimental del amor pasional y erótico por un lado y del matrimonio por el otro, puede observarse claramente en la existencia del género epitalámico separado de los géneros poéticos en los que el erotismo se expresa con mayor libertad. No sólo la especificidad del género epitalámico se relaciona con el rito, sino que el ideal amatorio expresado en los poemas matrimoniales es muy diferente y está claramente delimitado en sus alcances.

El matrimonio tenía para los romanos un carácter sagrado, puesto que su finalidad era la procreación humana y constituía, naturalmente, un asunto concertado entre dos familias que quedarían unidas por medio de su descendencia. La felicidad de los esposos no resultaba algo prioritario, ya que dicha unión estaba al servicio de la *gens* del mismo modo que estaba al servicio del γένος en el caso griego (Grimal, 2000, p. 17). Si el matrimonio era concebido para la generación de hijos legítimos, la mujer romana, y en mayor medida la mujer patricia, era educada para ser madre y esposa:

La importancia social de la *mater* estaba fundamentada no solo en su papel de procrear hijos para su marido, sino también en formar a los niños, futuros *cives*, en los deberes cívicos y los valores romanos: *pietas, fides, gravitas, virtus, frugalitas*. Las niñas eran instruidas en las labores propias del hogar: bordar, hilar, preparar la lana y actividades afines con su futura función de *mater familias, lanifica, casta, pia, frugi, domiseda, univira* eran actitudes deseables en toda mujer (Álvarez, 2012, p. 60).

#### 4. PRIMERAS EXPLORACIONES: ANÁLISIS ESTILOMÉTRICO BÁSICO DE TRES POETAS ROMANOS

La estilometría es el análisis estadístico de textos literarios a través de herramientas y métodos destinados al análisis cuantitativo. A través del uso de un software de análisis automatizado de textos, se aplican técnicas de minería y análisis computacional con el fin de detectar patrones en grandes conjuntos textuales (Fradejas Rueda, 2020). En este caso, los diferentes análisis estilométricos se realizaron con la herramienta R Studio. Este apartado describe el trabajo con los poe-

mas en relación con los dos universos amorosos comentados: campo de lo erótico y pasional y el campo del matrimonio y sus ritos.

#### 4.1. *Labor limae*: la cosecha y el armado del corpus

Para construir el corpus de análisis se recurrió a la biblioteca digital Perseus Project<sup>3</sup> del Departamento de Clásicas de la Universidad Tufts (Cerrato y Chavez, s. f.). En Perseus se encuentra la obra de los tres autores del siglo I a.C. y contiene, además, 2.412 obras en 3.192 ediciones y traducciones (1.639 en griego y 636 en latín) y 69,7 millones de palabras: 32,1 millones en griego, 16,3 millones en latín, y el resto consiste en traducciones al inglés y fuentes en otros idiomas. Los materiales allí publicados están debidamente cuidados por especialistas, compartidos bajo una licencia CC BY SA 3.0 (US), y los textos expuestos en formato XML disponible para su descarga. El proyecto proporciona además otros recursos como un modelo para el etiquetado gramatical o *Part-of-speech Tagging* (POS Tagging) y *stopwords* para el latín.

Se procedió entonces a realizar una cosecha de los poemas presentes en dicho sitio por medio de procedimientos de *web scraping*. El *web scraping*, literalmente *escarbar una web*, es una técnica que consiste en extraer y recopilar información de forma automática de la web transformando la información no estructurada del sitio en datos estructurados mediante programas de software (Martí, 2016; Munzert et al., 2014). El software utilizado para la cosecha fue el mismo R Studio con su librería *rvest*. Luego, por medio de las librerías *tidyverse* y *tidytext* se hicieron los primeros análisis del corpus:

Características del Corpus	Catulo (Merrill, 1893) <sup>4</sup>	Propercio (Müller, 1898) <sup>5</sup>	Tibulo (Postgate, 1915)
Número de Archivos	117	105	37
Número de Versos	2.302	4.257	1,987
Palabras-Token	12.883	25.310	12,701
Promedio por Poema	110,4701	243,6381	343,7838
Palabras-Tipo	5.816	9.025	5,330

Tabla 1. Comparación de características en los tres corpus analizados. Fuente: Elaboración propia.

<sup>3</sup> Accesible desde: <https://www.perseus.tufts.edu/hopper/>.

<sup>4</sup> En el ámbito de la minería de textos o en el del análisis automatizado de textos una palabra es simplemente una secuencia de caracteres. Para una computadora cualquier letra o cualquier número y el guión bajo, situado entre dos espacios en blanco, o entre un espacio en blanco y un signo de puntuación es una palabra (Fradejas Rueda, 2020). Una palabra-token es cada una de las formas que aparecen en el texto, sin importar cuántas veces ocurra cada una (Sierra et al., 2018). En el caso de Catulo, serán cada una de las apariciones del término *Lesbia*, por ejemplo, sumadas a cada una de las apariciones del resto de las otras palabras. Las palabras-tipo o *types* son lo que en Procesamiento del Lenguaje Natural como el vocabulario del corpus, es decir, el conjunto de términos distintos que ocurren al menos una vez en el corpus. Son palabras tipo cada una de las formas o palabras diferentes que aparecen en un texto pero no todas sus apariciones; el número de palabras-tipo se obtiene contando una sola vez cada palabra diferente sin importar cuántas veces se repita en el texto.

<sup>5</sup> En los trabajos anteriores se utilizó la edición de Vincent Katz (Katz, 2016) disponible en Perseus. Sin embargo, dicha publicación se encuentra incompleta en el sitio por lo que en esta ocasión se recurrió a la edición de Lucian Müller. Los datos, gráficos y resultados de este artículo se encuentran curados y seguramente difieran en gran medida de las publicaciones anteriores.

Como puede observarse en los gráficos siguientes, si bien el poemario de Catulo es el que más poemas incluye resulta el más dispar en cuanto a la longitud de los poemas. Es decir, en el *Corpus Catullianum* nos encontramos frente a dos tipos de muestras bien delimitadas: un conjunto grande de poemas cortos con pocas palabras y un conjunto breve de poemas largos con muchas palabras. Tal disparidad seguramente produce un sesgo estadístico en el análisis automático.

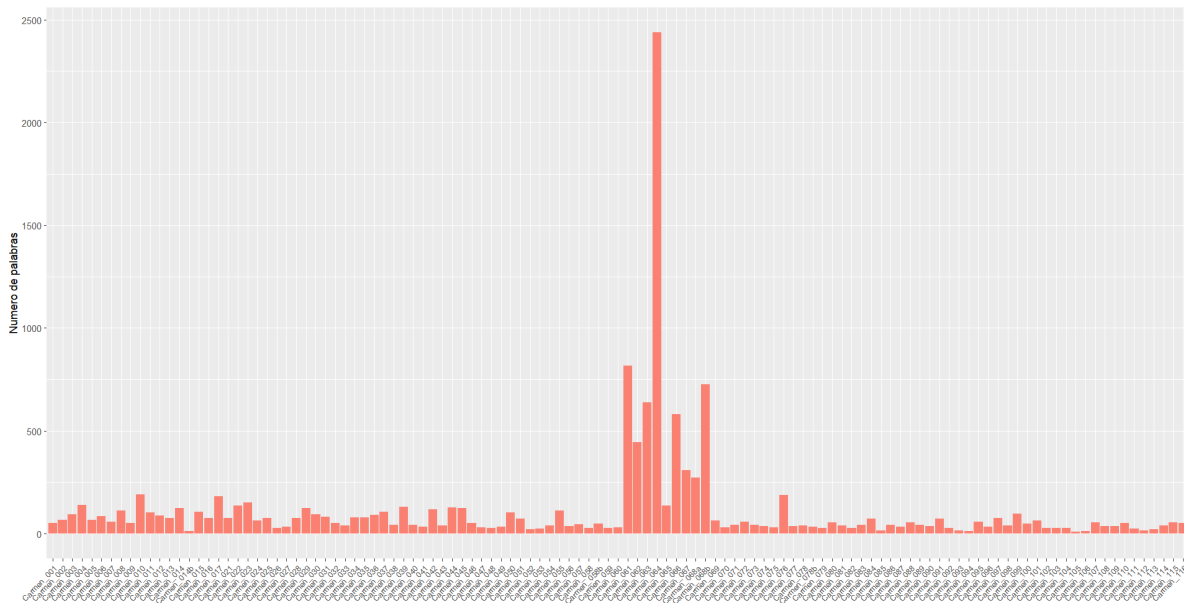


Figura 1. Poemas del *Corpus Catullianum*. La longitud de cada poema expresada en cantidad de palabras<sup>6</sup>. Fuente: Elaboración propia.

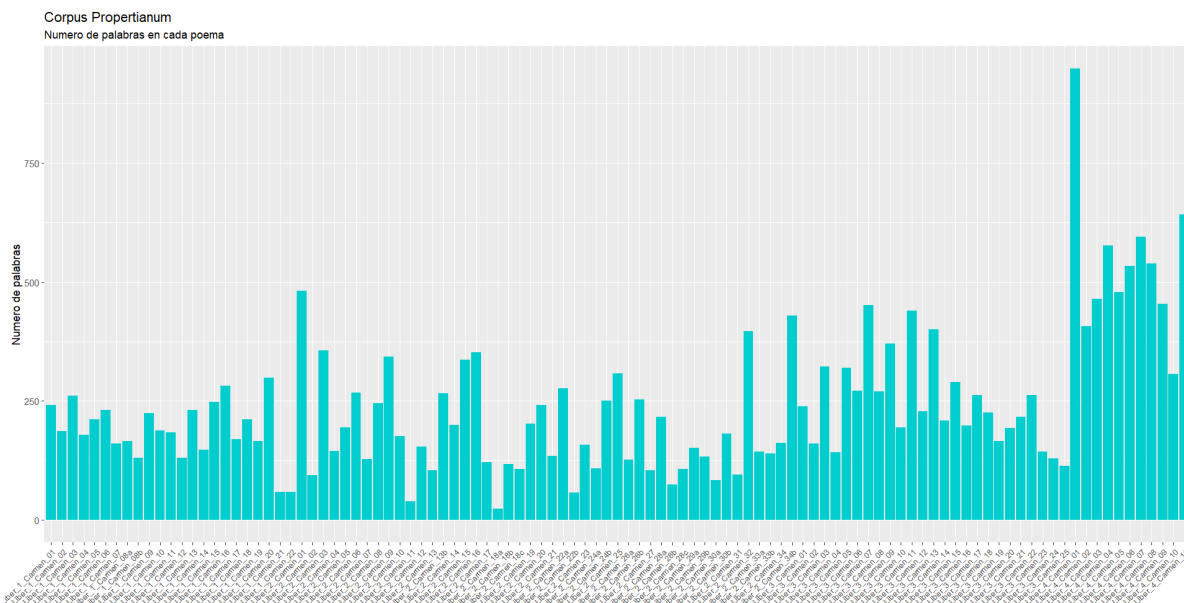


Figura 2. Poemas del *Corpus Propertianum*. La longitud de cada poema expresada en cantidad de palabras. Fuente: Elaboración propia.

<sup>6</sup> En su gran mayoría los gráficos de este artículo han sido generados siguiendo los scripts de José Manuel Fradejas Rueda (2020) con las modificaciones pertinentes para ser adaptados a los textos latinos.

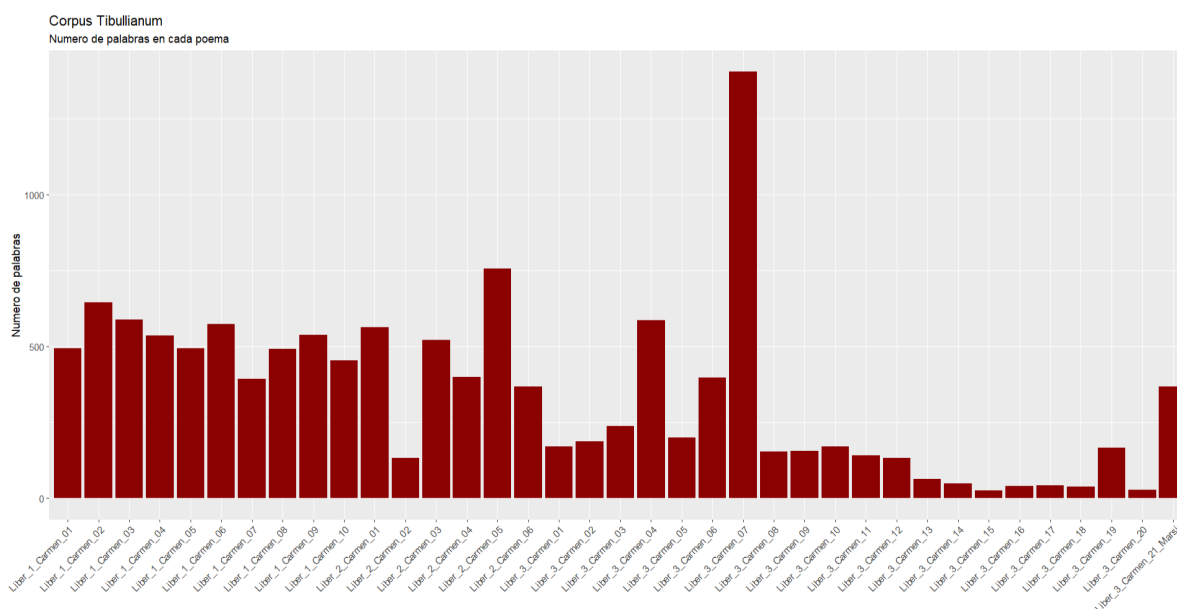


Figura 3. Poemas del *Corpus Tibullianum*. La longitud de cada poema expresada en cantidad de palabras. Fuente: Elaboración propia.

#### 4.2. Estribillos, versos repetidos y stopwords o palabras vacías

Con el objetivo de indagar algunas cuestiones básicas sobre el léxico del poemario de Catulo y sus palabras más usadas, se preparó el entorno de trabajo en R Studio y se utilizaron las librerías *tidyverse* y *tidytext* (Silge y Robinson, 2017). Un recurso importante cuando se busca trabajar con léxicos es la posibilidad de eliminar las palabras vacías o *stopwords*. Las palabras vacías son las palabras de una lista vacía (o un diccionario negativo) que se filtran (es decir, se detienen) antes o después del procesamiento de los datos del lenguaje natural (texto) porque son *insignificantes* (Rajaraman y Ullman, 2011). Se trata mayormente de preposiciones o términos con significado anafórico, catafórico, conectores, términos que contribuyen a la organización sintáctica. No siempre es procedente la eliminación de las palabras vacías puesto que no carecen totalmente de significado y muchas veces forman parte de expresiones importantes o son enormemente necesarias para la tarea a realizar, como en el caso de atribución de autoría. En la técnica presente, sin embargo, resultó preferible eliminarlas puesto que su ocurrencia suele ser mucho mayor que la de otros términos de mayor carga semántica en la temática que se busca detectar. Para el caso del latín se utilizaron las *stopwords* proporcionadas por el mismo Perseus Project. Se elaboró un gráfico con las palabras más utilizadas que aparecían más de diez veces en el poemario. Las más comunes eran algunas cuya importancia debía sopesarse porque pertenecían a estribillos (*currite*, por ejemplo) y los pronombres de primera y segunda persona en caso nominativo (*ego* y *tu*), así como el caso dativo (*mihi* y *tibi*).

Las palabras significativas que se obtuvieron y que están relacionadas al repertorio amoroso elegíaco son: *puella* (muchacha), *puellae* (de la muchacha), *Lesbia*, la amada de Catulo, *pectore* (en el pecho), *amores* (amores), *amore* (en el amor) en posición 56. Otro campo semántico posiblemente representado entre los primeros términos fue el del matrimonio con: *virgo* (virgen), *pater*



(padre), *Himeneae* (Himeneo) e *Hymen* (Himen). Sabemos que la amada elegíaca de Catulo era una mujer culta y casada, con lo cual es imposible que el término *virgo* se vincule a Lesbia. El padre es nombrado frecuentemente en los poemas de bocas y los términos *Hymen* e *Himeneae* que refieren al dios y al rito matrimonial aparecen con una frecuencia muy alta porque pertenecen a los estribillos de los epitalamios. Se trata de términos cuya repetición se debe a la característica disposición y musicalidad del género, poemas originalmente compuestos para ser cantados por coros en los ritos matrimoniales en los que se suelen repetir estrofas. No obstante, esa repetición también puede ser tomada como un rasgo de importancia en cuanto a la fijación de ciertos conceptos, sobre todo en la lírica compuesta para ser escuchada. Más precisamente, como el sujeto oyente es pasivo, y no puede alterar la velocidad de sucesión de los elementos del discurso como lo haría un lector, la repetición deviene un recurso compositivo valioso que proporciona un mayor rendimiento estético. Como lo explica en su tesis Fernández Pereiro (1968), las variaciones permiten presentar un objeto sucesivamente desde diferentes aspectos y la percepción del público oyente es sintética, llega al goce de manera acumulativa. Por estas razones, la opción de eliminar o no a las palabras que se repiten en un estribillo, deviene una decisión controvertida y afecta la concepción poética de aquello que se considera que el autor deseaba marcar como *importante* mediante la repetición.

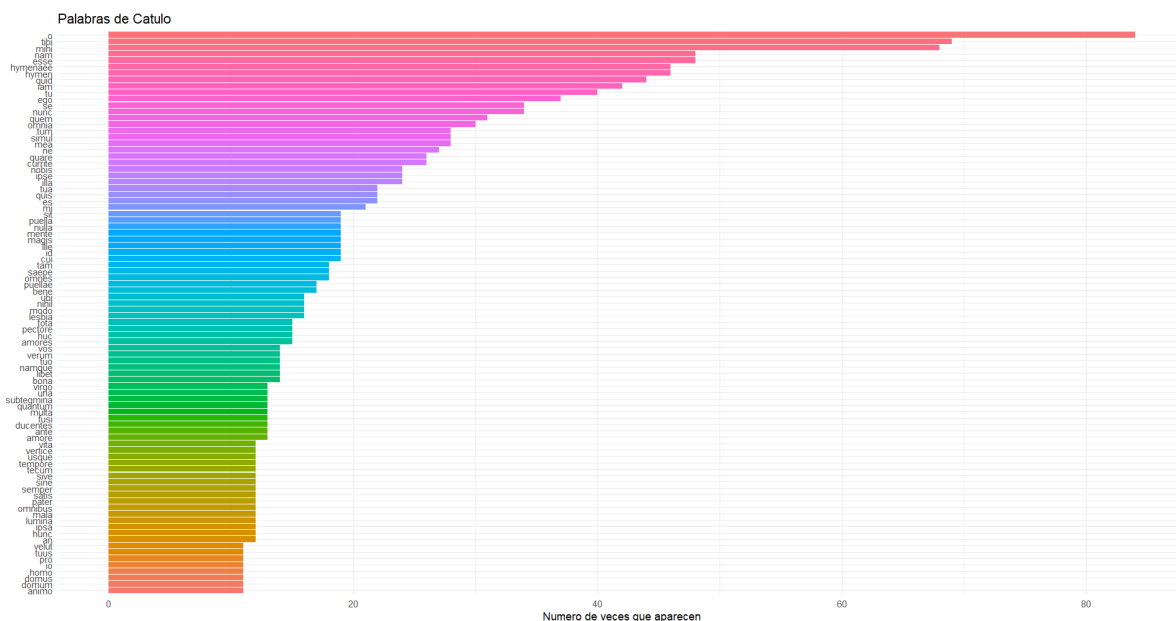


Figura 4. Palabras más frecuentes en Catulo habiendo quitado palabras vacías de Perseus. Fuente: Elaboración propia.

Atendiendo a las observaciones anteriores, en un trabajo anterior se procedió a crear palabras vacías propias con el fin de indagar un poco más en el corpus y eliminar términos que resultaban poco significativos para el vocabulario amoroso. Se elaboró, a modo experimental, un listado con términos frecuentemente utilizados en versos repetidos y estribillos en el *corpus catullianum*, como *hymeneae*, *hymen*, o *currite*. Sin embargo, el número de palabras significativas pertenecientes al vocabulario amoroso erótico que aparecían en el gráfico no cambió significativamente. Además, la eliminación de esos términos marcándolos como palabras vacías

suprimía todas sus ocurrencias cuando lo correcto, siguiendo aquel criterio, hubiera sido considerar los términos repetidos de estribillos una sola vez y conservar además el conteo de las ocurrencias de esos mismos términos en versos que no eran repeticiones o estribillos.

Para este artículo se elaboró un *script* con el que se detectaron los versos repetidos y se realizó un conteo de frecuencias al listado de palabras de versos repetidos. Este listado permitió restar esas palabras a las frecuencias totales, de modo que las palabras repetidas en estribillos se consideraran una sola vez sin modificar la frecuencia de las mismas palabras en otros versos no repetidos. El resultado de esta operación fue el siguiente; como puede observarse, no se ha modificado demasiado la distribución total de los términos.

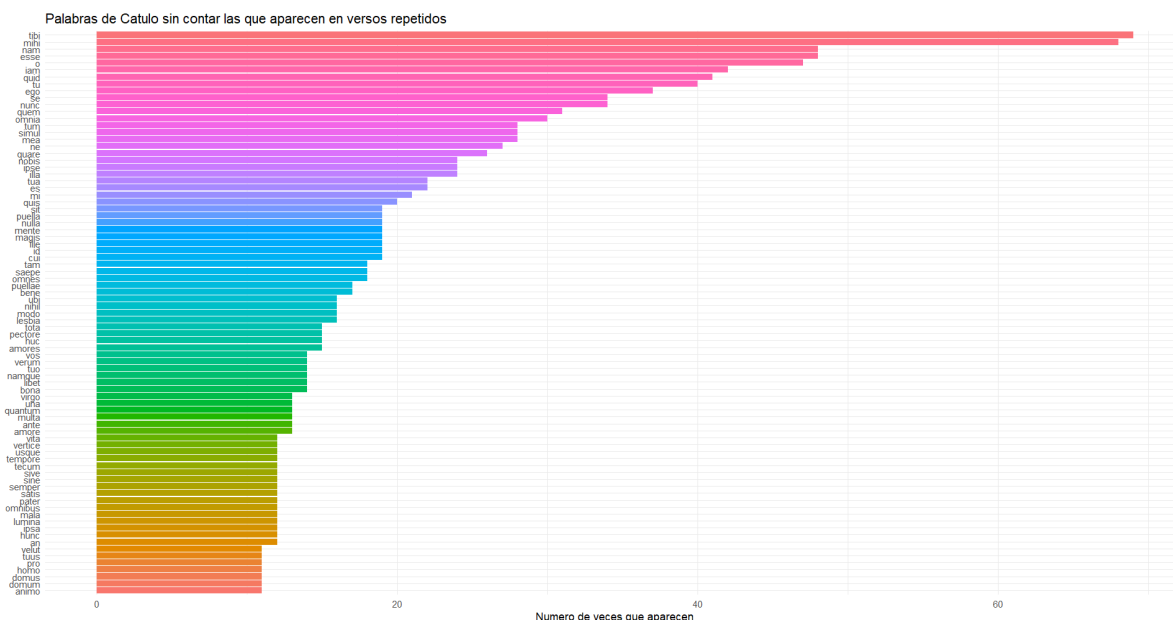


Figura 5. Palabras más frecuentes en Catulo sin términos repetidos en estribillos. Fuente: Elaboración propia.

No obstante, Catulo no solo repite versos en estribillos de poemas epitalámicos. El recurso de la repetición de versos es empleado en otros poemas. Por ejemplo, el verso *Hymen* o *hymenaeae*, *hymen ades* o *hymenaeae* (Oh Himen, o Himeneo, Himen acércate, o Himeneo) en el carmen 62 se utiliza unas ocho veces y *Currite ducentes subtegmina, currite, fusi* (“Girad conduciendo los hilos, girad, husos”) unas 11 en el carmen 64. En el marco de estos dos poemas esas repeticiones sí podrían considerarse como un estribillo. En otros poemas algunos versos aparecen apenas dos veces, como el caso de *Cinaede Romule, haec videbis et feres?* (“Puto Rómulo, ¿verás esto y lo soportarás?”). Evidentemente, en este último caso la intención autoral en el uso de la repetición de versos persigue otros fines rítmicos o simplemente enfatiza una idea. Siendo la repetición rítmica uno de los factores constructivos más importantes del discurso poético, ¿por qué debería ser correcto eliminar la repetición de los estribillos, un recurso evidentemente intencional por parte del autor y no deberíamos eliminar, según el mismo criterio, otro tipo de repeticiones como la anáfora, la epífora, la anadiplosis, la epanadiplosis o la epístrofe?

Luego de indagar en términos generales las palabras del corpus de Catulo, con el criterio de eliminar repeticiones de versos, se decidió desecharlo. En las técnicas siguientes se respeta la repetición de términos diseñada originalmente por el autor ya que, de todas formas, no afecta demasiado el análisis ni modifica en gran medida los resultados. Tampoco se continuó con la idea de utilizar palabras vacías *ad hoc* puesto que muchos términos del vocabulario amoroso (sustantivos, adjetivos y verbos) simplemente se hallaban utilizando técnicas tradicionales de análisis automático de textos como el etiquetado de partes del discurso y la lematización.

### 4.3. Comparación de léxicos de Catulo, Propertio y Tibulo

A continuación, se trabajó con el corpus de Catulo en comparación con el poemario de Propertio para ver si ambos autores presentaban vocabularios amorosos similares. Con ese objetivo se generaron los gráficos con las palabras más comunes en ambos poemarios aplicándoles las palabras vacías proporcionadas por Perseus Project. Los términos más frecuentes en ambos poetas son, como ocurrió anteriormente, los pronombres de primera y segunda persona singular y sus respectivas formas en nominativo y dativo. En el caso de Catulo aparecieron las palabras de los estribillos entre las primeras como ocurrió en el análisis anterior. Otros términos frecuentes fueron *puella*, *puellae* y *amor*, *amore*, *amores* y por supuesto los nombres de sus respectivas amadas: *Lesbia* y *Cynthia*.

En el poemario de Propertio *Cynthia* aparece en filas superiores frente a *Lesbia* en el poemario de Catulo. Las cincuenta palabras más usadas por Propertio y Catulo son:

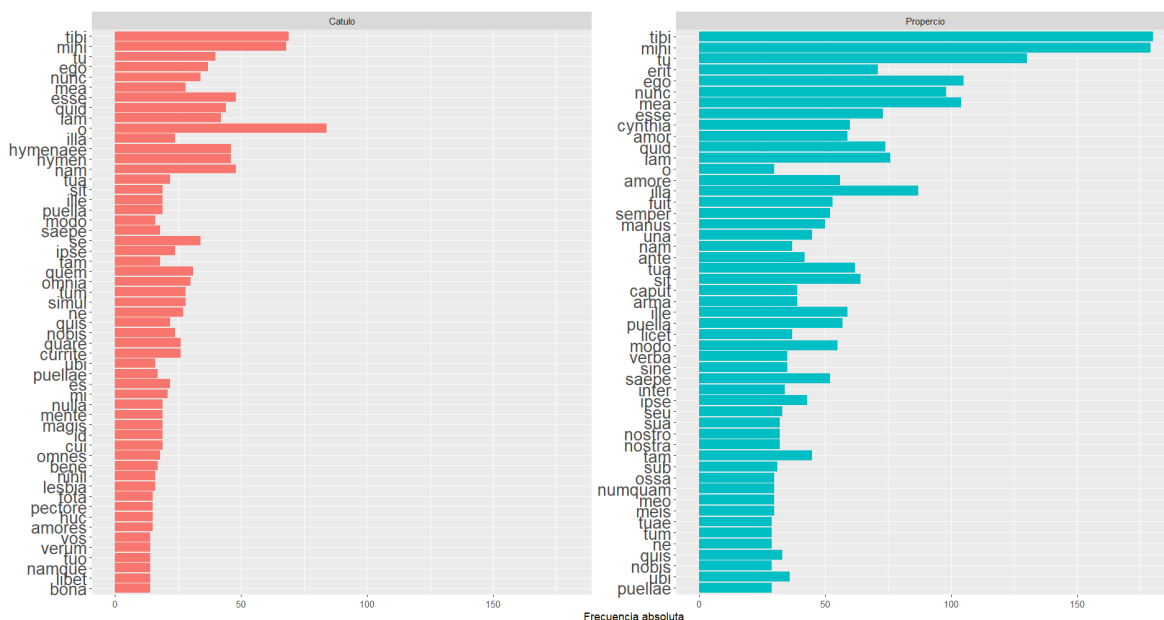


Figura 6. Términos más frecuentes en Catulo y Propertio habiendo retirado solo las palabras vacías de Perseus Project. Fuente: Elaboración propia.

Para comparar a Catulo y Tibulo se procedió de la misma manera, se exploró el corpus solo quitando las palabras vacías del Perseus en un primer momento para ver qué términos aparecían con mayor frecuencia. También en el caso de Tibulo encabezan ambas tablas los

pronombres personales *ego* y *tu* en sus correspondientes casos en dativo *mihi* (para mí) y *tibi* (para ti). El caso nominativo de ambos pronombres sigue entre las primeras posiciones.

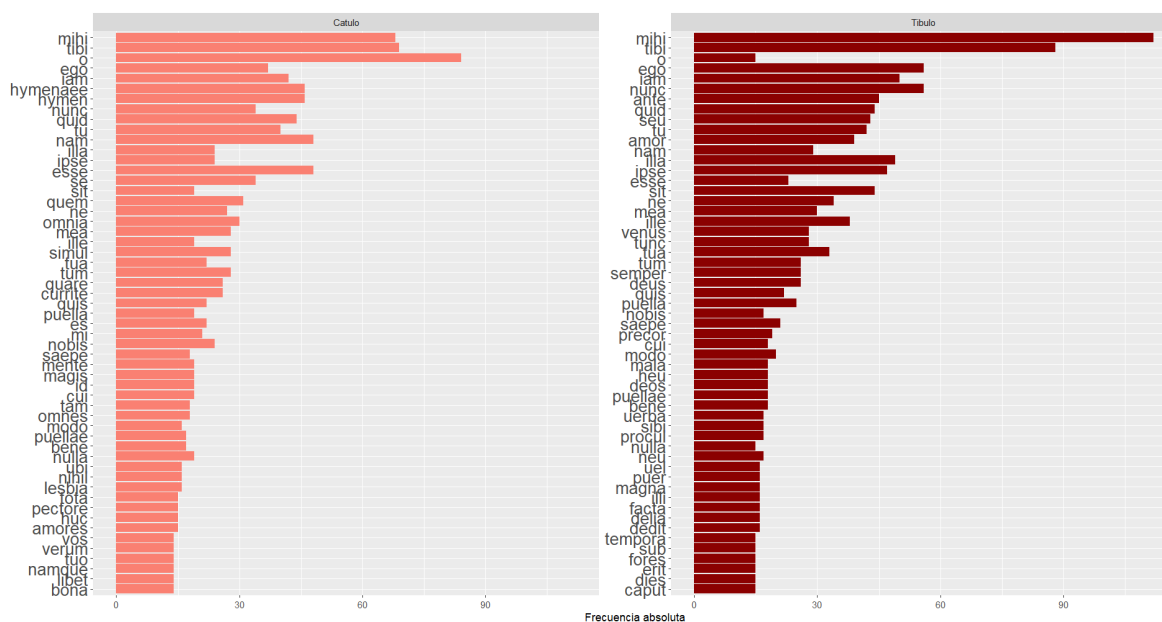


Figura 7. Términos más frecuentes en Catulo y Tibulo habiendo retirado solo las palabras vacías de Perseus Project. Fuente: Elaboración propia.

Pueden observarse también términos del léxico amoroso en Tibulo, como era de esperarse: *amor*, nada menos que *Venus* la diosa del amor y *puella*.

Otra de las tareas fue la búsqueda de palabras repetidas más de tres veces por poema en cada poeta. En Catulo aparecen los repertorios de los *carmina maiora* con términos vinculados al matrimonio: *virgo*, *nupta* (esposa), *concubine* (oh, concubino). El poema con mayor número de palabras repetidas es el carmen 64, también el más extenso de todo el poemario con una cantidad de palabras mucho mayor a la de los otros poemas. Se trata de un poema de carácter narrativo así que no es de extrañar que *Theseus* sea una figura prominente ya que se trata de quien protagoniza un episodio narrado por las Parcas. Las palabras más repetidas en el poemario de Propertio son de carácter amoroso erótico: *Cynthia* y *amore* las de mayor número de apariciones. En el Libro I, Tibulo presenta a *Delia* nombrada cuatro veces en el carmen 3, *Amor* tres veces en el sentido del dios y una en el sentido del sentimiento en el carmen 6, en el carmen 8 *Venus* es nombrada tres veces como diosa y una para referirse en sentido figurado al amor.

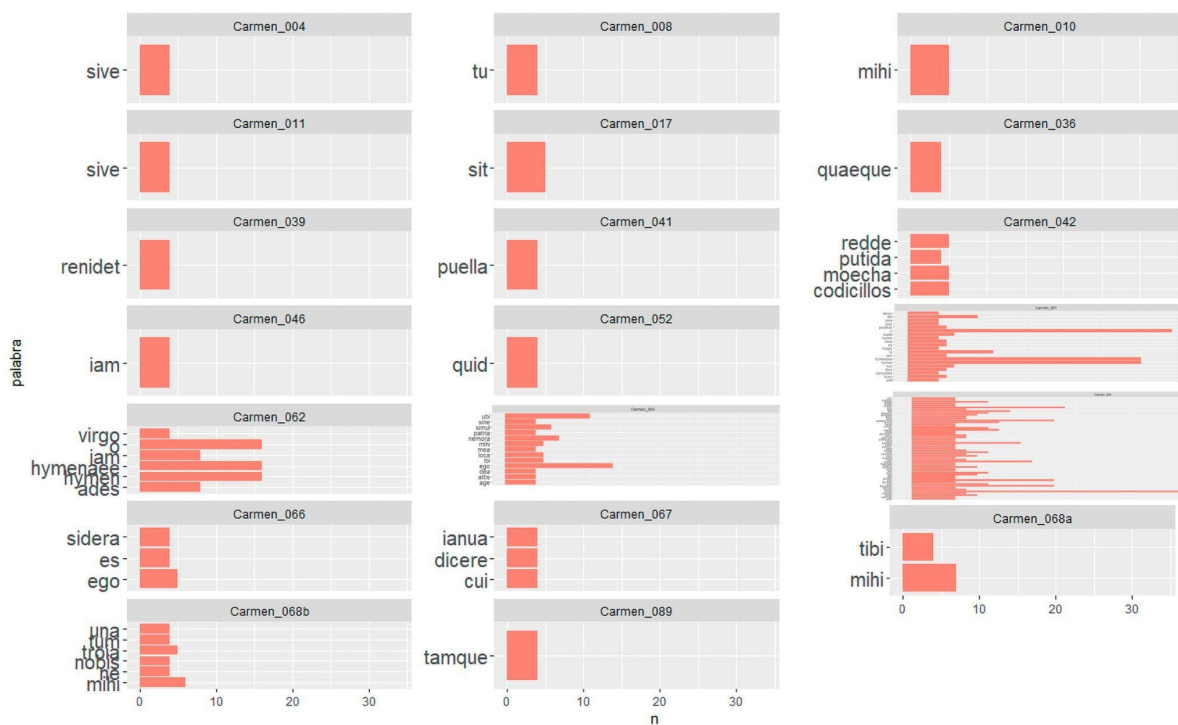


Figura 8. Palabras repetidas más de tres veces en el *Corpus Catullianum* ubicadas por el poema en el que aparecen. Fuente: Elaboración propia.



Figura 9. Palabras repetidas más de tres veces en el *Corpus Propertianum* ubicadas por el poema en el que aparecen. Fuente: Elaboración propia.

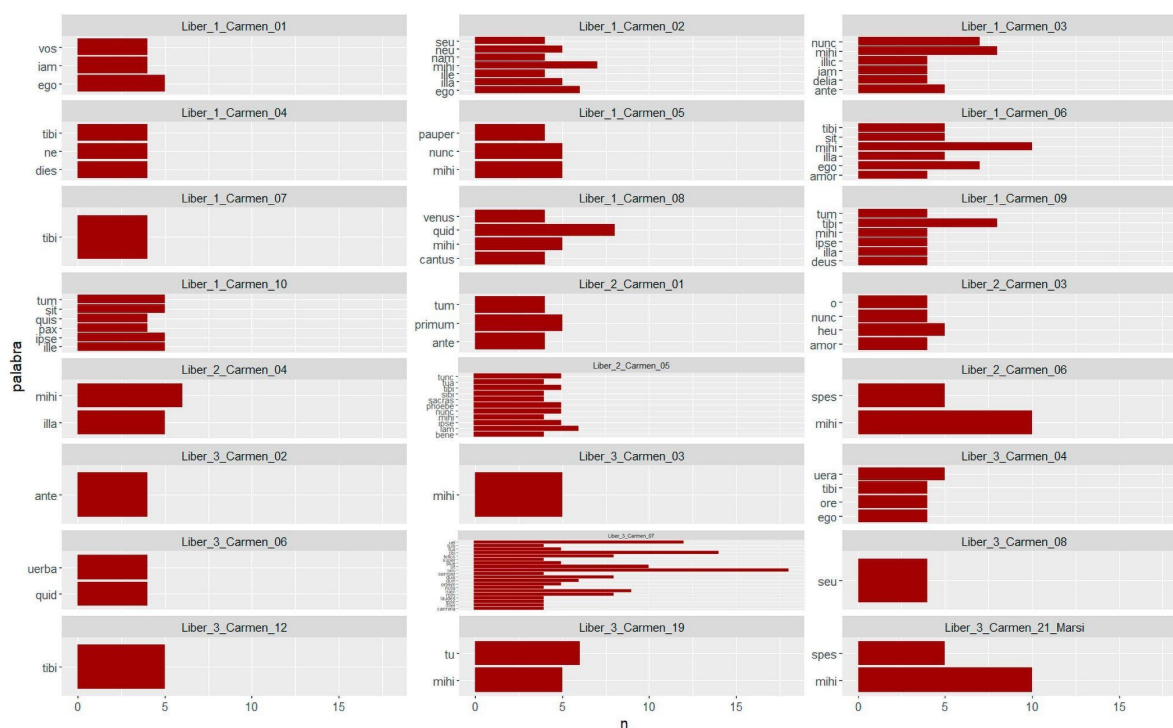


Figura 10. Palabras repetidas más de tres veces en el *Corpus Tibullianum* ubicadas por el poema en el que aparecen. Fuente: Elaboración propia.

#### 4.4. Clases de palabras en Catulo

Para este apartado se recurrió a las librerías `udpipe`<sup>7</sup> (Straka y Straková, 2019, 2020), `tidyverse` y `tidytext` y se realizaron los gráficos de los cincuenta sustantivos más utilizados por Catulo, el primero con palabras-token y el segundo con términos lematizados. Para la lematización se utilizó el etiquetado gramatical o *Part-of-speech Tagging (POS Tagging)* que provee la librería `udpipe` en su modelo `Perseus`. Las palabras-token aportan información interesante de los casos en los cuales están usados los sustantivos; por ejemplo, no es irrelevante cuando Catulo se refiere a sí mismo en caso vocativo, cuando es sujeto o cuando es el objeto de una acción. Cierta parte de la crítica ha analizado la poesía de Catulo desde una visión psicologista y genérica encontrando en su poesía fantasías de autocastración y travestismo o confesiones de homosexualidad veladas (Nusch, 2020, p. 72). Si nos interesara centrarnos en ese aspecto, analizar el rol sintáctico tanto de Catulo como de Lesbia en los poemas sería un acercamiento posible. En el caso de los lemas, lo provechoso es ver qué núcleos de significados son importantes para el poeta sin detenerse en la circunstancia aludida en los poemas por medio del caso o del tiempo verbal.

<sup>7</sup> Según la documentación de la librería `Udpipe` el modelo `latin-perseus-ud-2.5-191206` ha tenido un desempeño del 100.0% en la detección de palabras, 98.5% para las sentencias, 83.3% para el etiquetado de partes del discurso (UPOS, del inglés Universal Part-of-Speech Tagging Accuracy) y 78.0% en la detección de lemas (Straka y Straková, 2020/2021). Para una revisión más actualizada sobre el desempeño de diferentes modelos de POS Tagging y Lematizadores puede consultarse (Nehrdich y Hellwig, 2022).

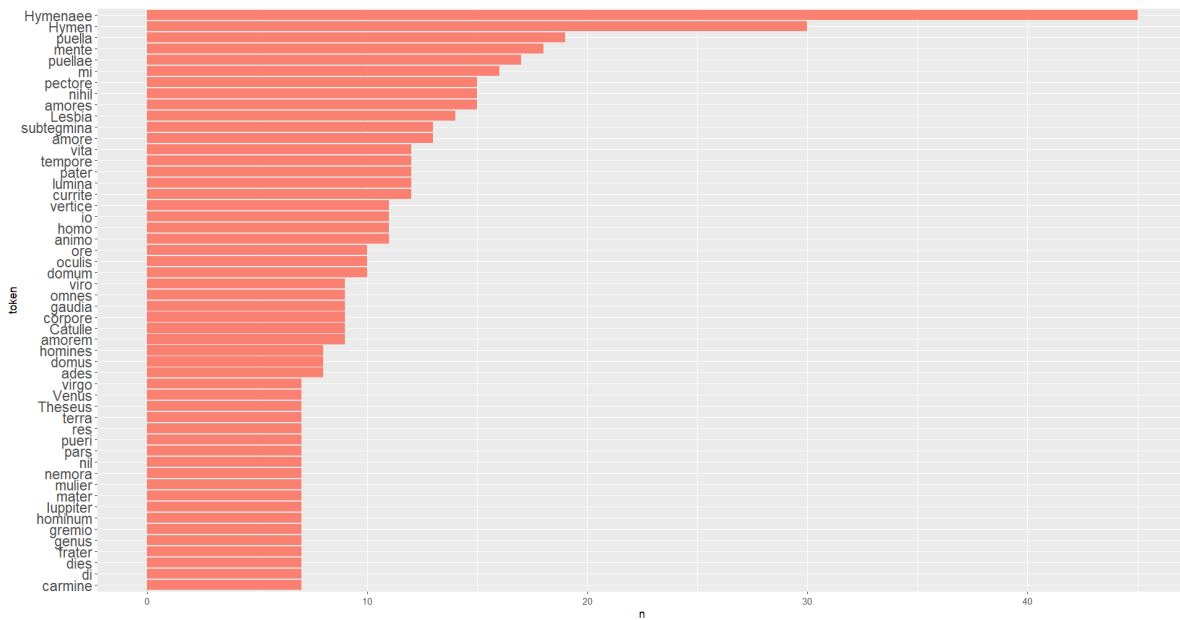


Figura 11. Sustantivos (palabras-token) más utilizados en Catulo. Fuente: Elaboración propia.

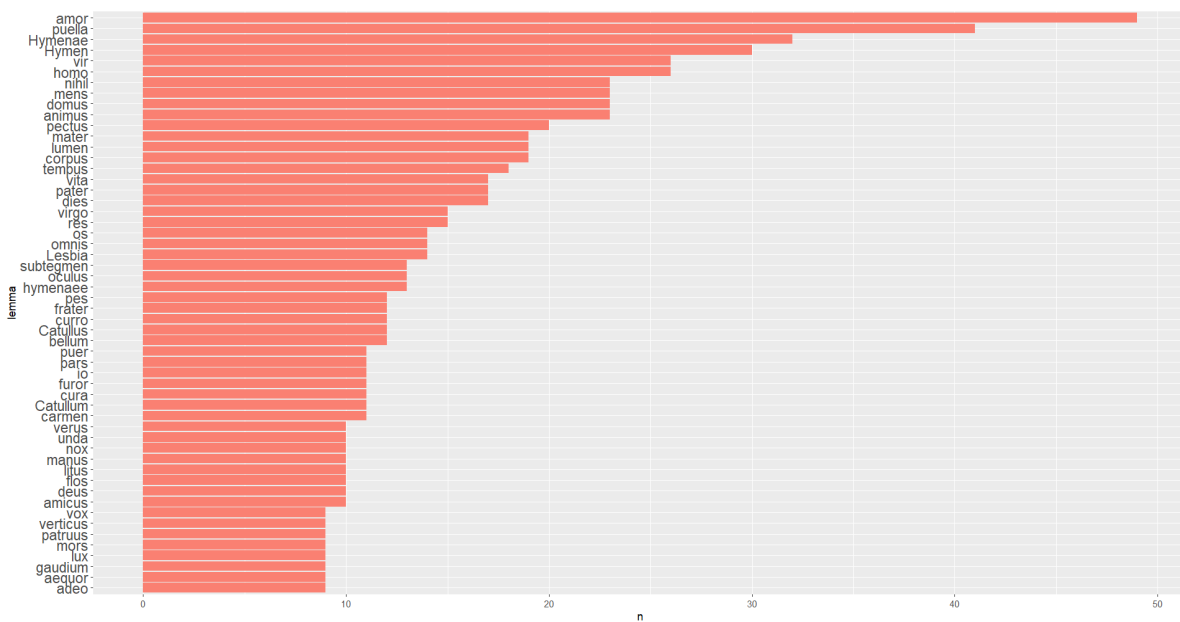


Figura 12. Sustantivos (lemas) más utilizados en Catulo ordenados por la raíz (lema) de la palabra. Fuente: Elaboración propia.

En el gráfico de los lemas *amor* y *puella* aparecen como más importantes, luego los dos referentes al himeneo que vienen de los epitalamios (*Hymen* e *Hymenaeae*). La presencia de *domus* (casa), *virgo*, *mater* (madre) y *pater* podría tener que ver con los poemas más extensos dedicados al matrimonio, aunque no necesariamente ya que *padre*, *madre*, y *virgen*, son también formas de dirigirse a determinados dioses. *Lesbia* aparece luego junto con *oculus* (ojos), *Catullus*, *furor* (locura, delirio). Son lemas pertenecientes a los poemas más breves y de temática amorosa. Un especial lugar ocupa la palabra *frater* (hermano), ya que el poeta lamentó varias veces su pérdida en su obra.





En el caso de los lemas verbales, los seis o siete primeros deben ser los más importantes en la lengua latina. *Amo* (amar) y *volo* (querer, desear) también ocupan un lugar importante en el poemario dedicado al amor.

Por último, se aplicó el mismo análisis para los adjetivos:

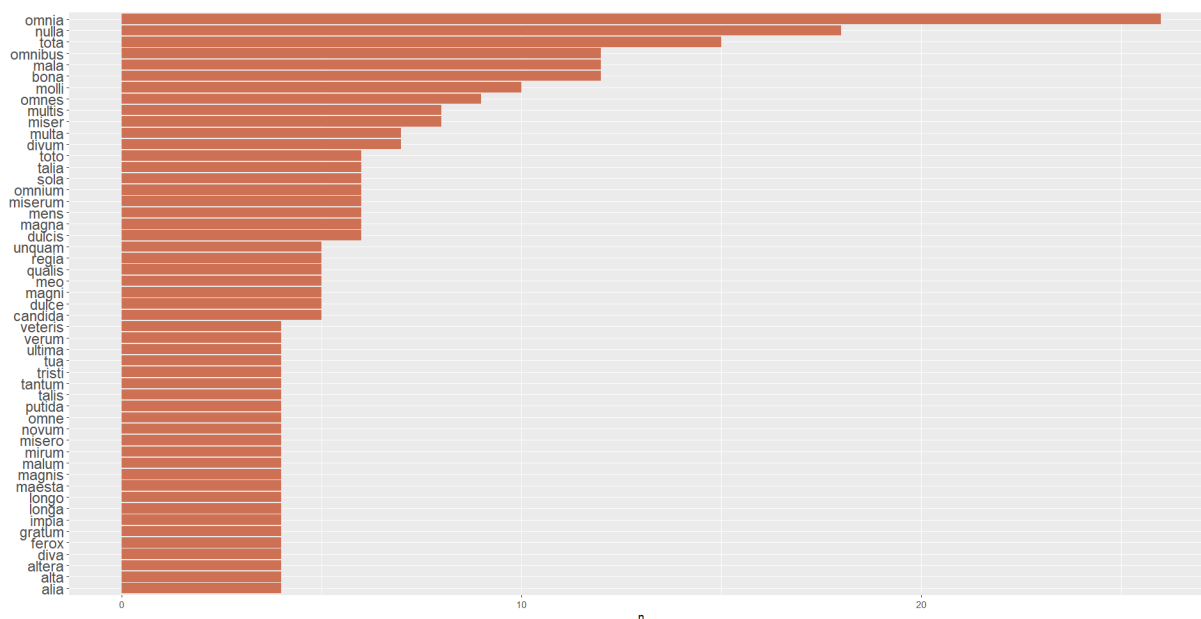


Figura 15. Adjetivos (palabras-token) más utilizados en Catulo. Fuente: Elaboración propia.

Los adjetivos *omnia*, *nulla*, *tota* son comunes. Luego aparecen *mala* y *bona*, estos últimos quizá digan algo sobre la tendencia del poeta a juzgar las acciones propias y ajenas desde un punto de vista ético o moral-religioso, un análisis más pormenorizado de estas apariciones sería necesario. *Miser* aparece entre los más frecuentes y algunas palabras erróneamente clasificadas como *mens*. Otra vez, alguna falla del modelo de etiquetado automático.

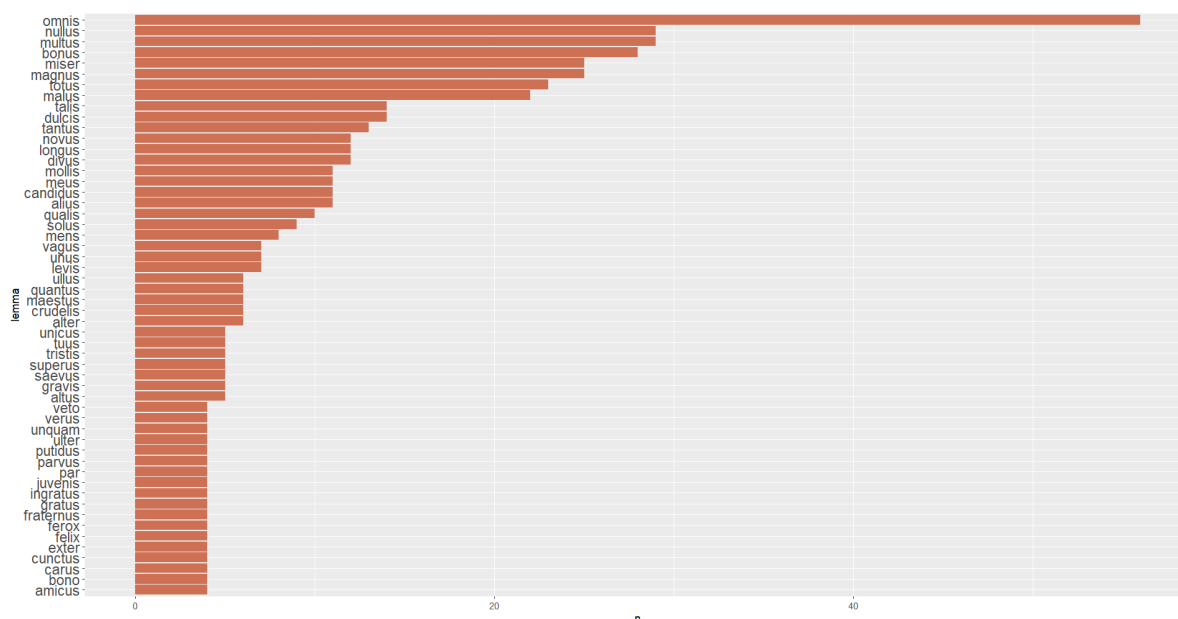


Figura 16. Adjetivos (lemas) más utilizados en Catulo ordenados por la raíz (lema) de la palabra. Fuente: Elaboración propia.

Con los adjetivos lematizados no se observan demasiados cambios, aunque *bonus*, *miser* y *malus* escalen posiciones. Sería interesante correr un análisis de sentimientos en el futuro y observar cuántos adjetivos positivos y cuántos negativos aparecen en el poemario.

#### 4.5. Los 50 lemas sustantivos más frecuentes de Catulo y Propertio

En este apartado se procedió a analizar las raíces sustantivas en los poemarios de Catulo y Propertio. En consonancia con la estética del amor utilizada y el vocabulario correspondiente, los sustantivos más importantes en ambos poetas son *amor* y *puella*. En Propertio aparece *Cynthia* y en Catulo los sustantivos del campo del matrimonio que se repiten en los estribillos de los epitalamios. Por encima de *Lesbia* aparecen los sustantivos *mater*, *pater* y *virgo*, que podrían vincularse al campo del matrimonio. En Propertio, en cambio, otros del campo erótico tienen mayor importancia: *Venus*, *amicus* (amigo), *fides* (fe, confianza), *domina* (señora).

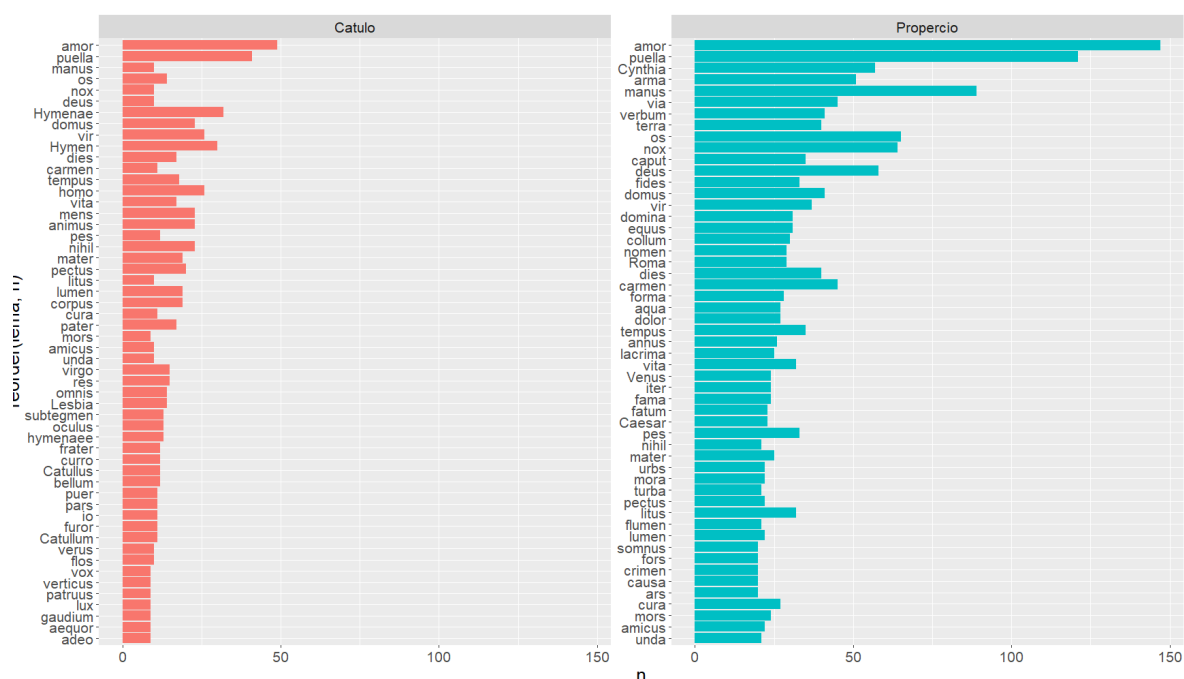


Figura 17. Los 50 sustantivos (lemas) más frecuentes en Catulo y Propertio ordenados por la raíz (lema) de la palabra. Fuente: Elaboración propia.

#### 4.6. Los 50 lemas verbales más frecuentes de Catulo y Propertio

En el caso de los verbos, en ambos poetas hay una coincidencia casi exacta de las formas más comunes del latín *sum*, *possum* (poder), *dico* (decir), *do* (dar), *fero* (llevar), *facio* (hacer) y luego aparece en séptimo lugar *amo*. *Video* (ver) podría considerarse muy importante para el imaginario amoroso, así como *gaudeo* (gozar) y *cupio* (desear); sin embargo, en Catulo ocupan los lugares más bajos de la tabla. De todas formas, que un término sea importante para una poética no quiere decir que será reiterado hasta el hartazgo en un poemario.

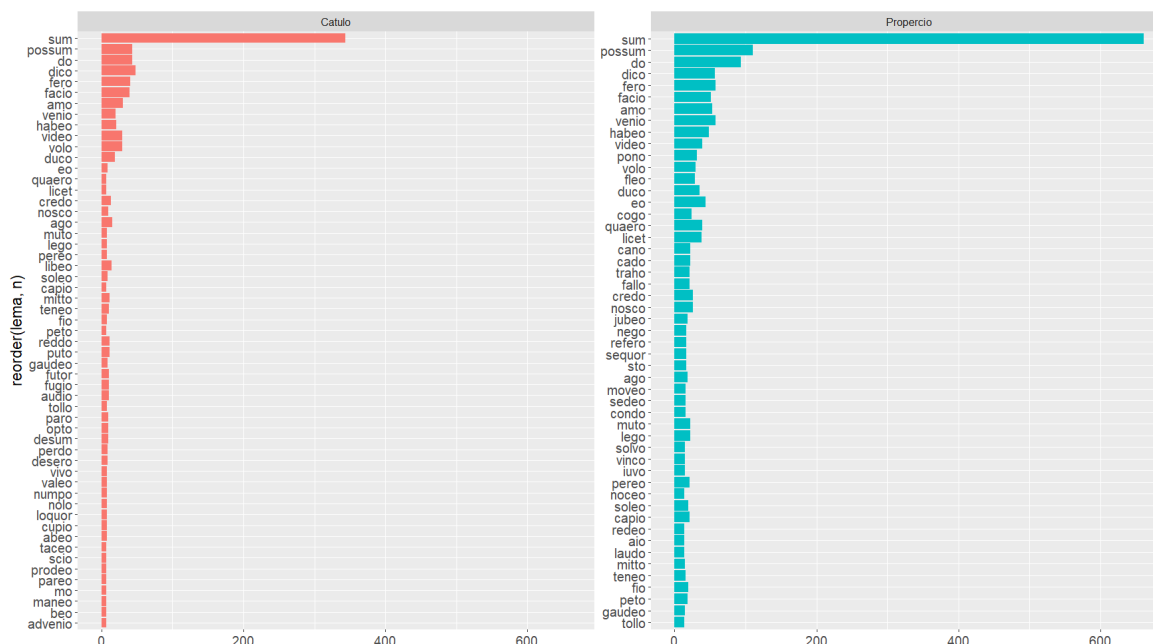


Figura 18. Los 50 verbos (lemas) más frecuentes en Catulo y Propertio ordenados por la raíz (lema) de la palabra. Fuente: Elaboración propia.

#### 4.7. Los 50 lemas sustantivos más frecuentes en Catulo y Tibulo

En la comparación de sustantivos lematizados con Tibulo los dos primeros términos amorosos continúan siendo *amor* y *puella*. En el caso de Tibulo, *Venus* y *deus* (dios) ocupan el tercer y cuarto lugar respectivamente, cuando en Catulo aparecen *Hymenae* y *deus*. Tibulo también muestra un conjunto de términos compatible con la moral elegíaca: *os* (boca), *pes* (pies), *cura* (preocupación), *nox* (noche), *corpus* (cuerpo). Otros términos menos comunes en el vocabulario erótico, como *pater*, aparecen mucho más abajo en el ranking. La primera enamorada de Tibulo, *Delia*, es nombrada menos veces que la Lesbia de Catulo.

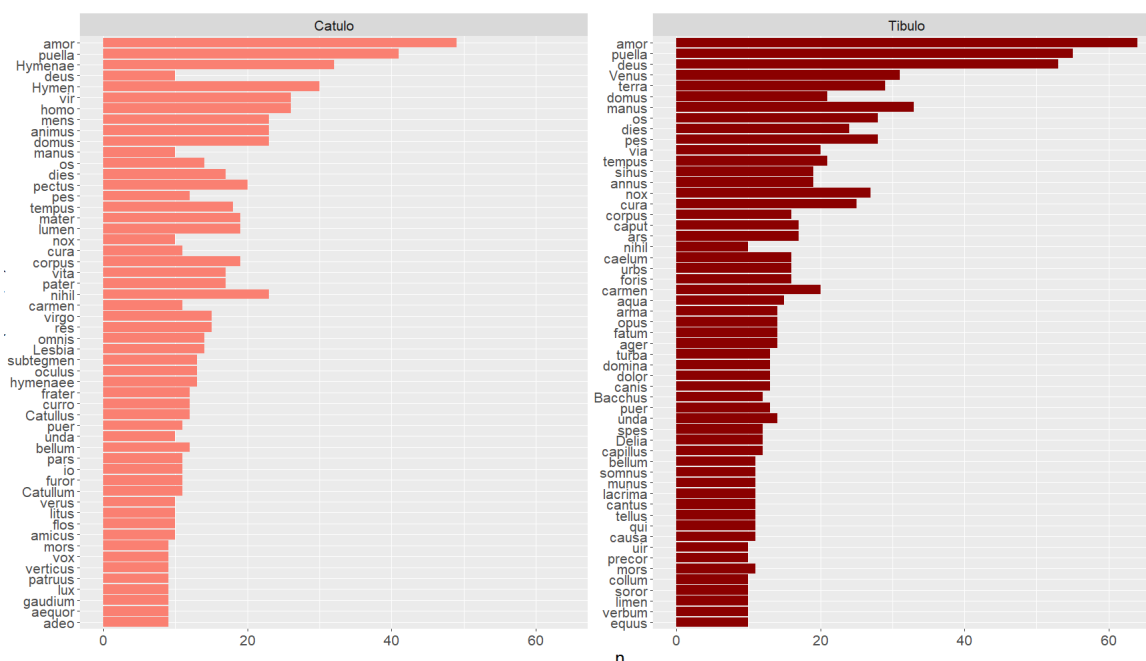


Figura 19. Los 50 sustantivos (lemas) más frecuentes en Catulo y Tibulo ordenados por la raíz (lema) de la palabra. Fuente: Elaboración propia.

#### 4.8. Los 50 lemas verbales más frecuentes en Catulo y Tibulo

En el caso de los verbos, Tibulo utiliza *venio* en séptimo lugar por sobre *amo*. A no ser por este último verbo, no aparecen verbos explícitamente eróticos, aunque algunos podrían indicar diferentes estados de ánimo del poeta como *timeo* (temer), *placeo* (causar placer), *sentio* (sentir, experimentar), no obstante, no tienen una aparición tan elevada.

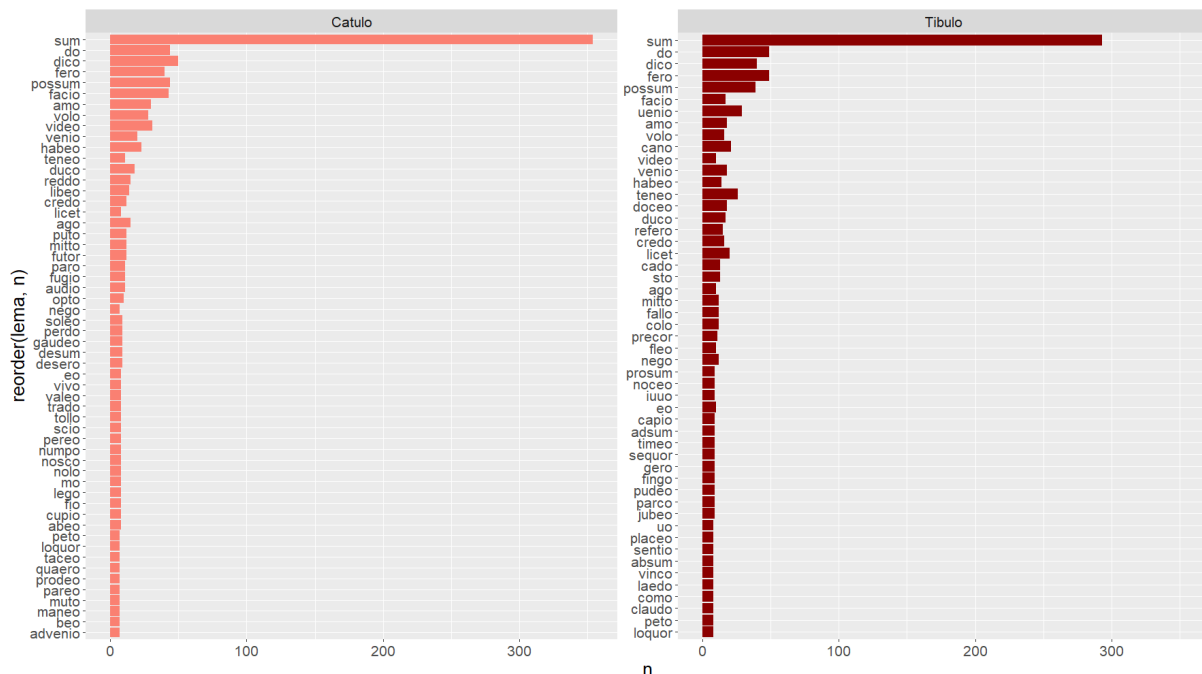


Figura 20. Los 50 verbos (lemas) más frecuentes en Catulo y Tibulo ordenados por la raíz (lema) de la palabra. Fuente: Elaboración propia.

### 5. BIGRAMAS

Un bigrama de palabra se refiere a una secuencia continua de dos palabras consecutivas en un texto. El análisis de los bigramas permitió indagar qué pares de términos aparecían juntos en la poesía de Catulo, Propercio y Tibulo y si esto tenía que ver o no con el imaginario amoroso elegíaco.

#### 5.1. Bigramas más frecuentes en cada poeta

La mayor parte de los bigramas de Catulo pertenecen a estribillos. En el caso de Propercio el número de bigramas es evidentemente mayor, lo cual es esperable puesto que su corpus es el más extenso en número de palabras.

Gran parte de los bigramas presentes en Catulo pertenecen a los *carmina maiora*, cuya temática es mayormente nupcial, y están especialmente ubicados en los estribillos. Del universo erótico aparece *meae puellae*, *meos amores*, y *Lesbia mi*. En Propercio la proporción de bigramas eróticos es mayor, con los inequívocos *puella fuit*, *nox mihi*, *mea Cynthia*, *mea musa* e *ipsa Venus*. En Tibulo también aparece *ipsa Venus*, junto con dos bigramas de temática erótica en la parte inferior del gráfico: *dura puella* (muchacha dura u obstinada) y *acer amor* (amor penetrante o agrio).

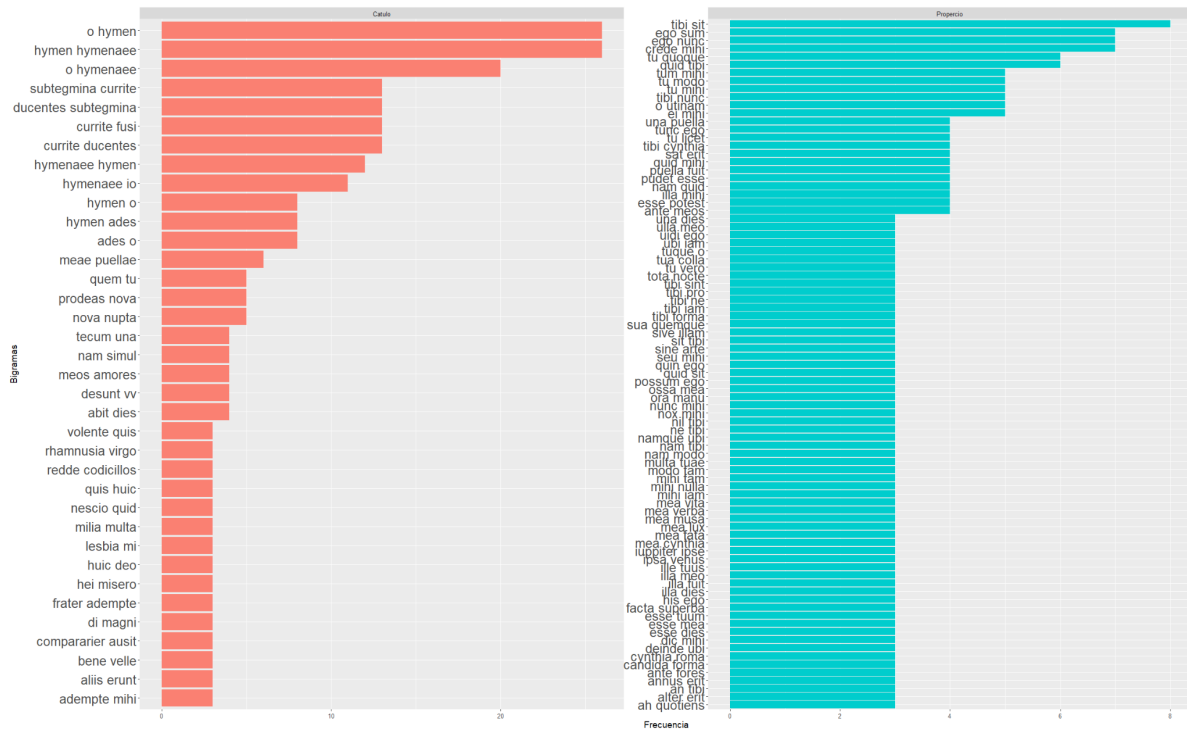


Figura 21. Bigramas más frecuentes en Catulo y Propertio. Fuente: Elaboración propia.

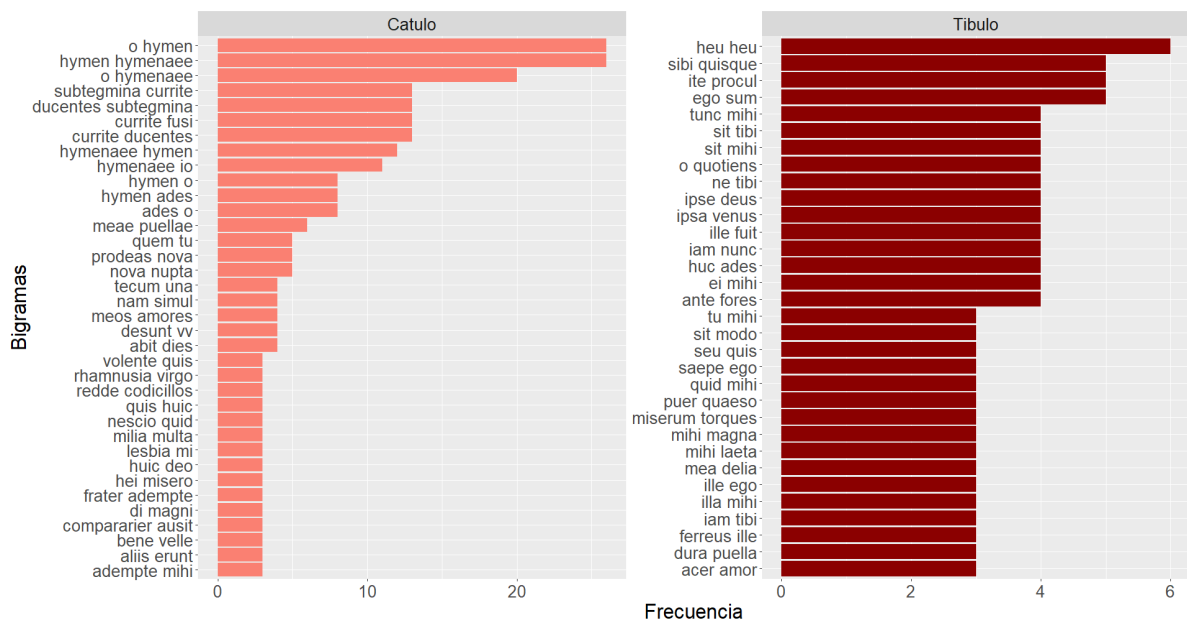


Figura 22. Bigramas más frecuentes en Catulo y Tibulo. Fuente: Elaboración propia.

En Catulo dos bigramas pertenecientes a los poemas dedicados al hermano muerto resultan, aparentemente, un caso curioso: *frater adempte* y *adempte mihi* (utilizado en los poemas 68 a y b así como en el 101). En realidad, se trata de un trigramas, *frater adempte mihi* que se repite en la misma posición en el verso en los *carmina* 68 y 101. Es decir, siempre en el segundo hemistiquio de los pentámetros (ambos poemas están compuestos en dísticos elegíacos).

*sed totum hoc studium luctu fraterna mihi mors  
abstulit. o misero frater adempte mihi,* (Carm. 68, 19-20)

Pero la muerte fraterna con su luto de toda esa dedicación me apartó. ¡Oh pobre mi hermano arrebatado!

*Troia (nefas) commune sepulcrum Asiae Europaeque,  
Troia virum et virtutum omnium acerba cinis:  
quaene etiam nostro letum miserabile fratri  
attulit. Hei misero frater adempte mihi, (Carm. 68, 89-92)*

Troya nefasta, sepulcro común de Asia y de Europa  
Troya amarga ceniza de todos los varones y las virtudes:  
ella inclusive trajo la muerte a nuestro pobre hermano.  
¡Oh mi pobre hermano arrebatado!

*heu miser indigne frater adempte mihi. (Carm. 101, 6)*

¡Oh mi pobre hermano arrebatado indignamente!

En los tres casos el trigrama aparece en la misma posición en el verso. Seguramente Catulo, un poeta enormemente consciente y en dominio de su técnica, haya tenido razones más que rítmicas para la utilización de tal esquema. Puede haberse tratado de un intertexto explícito entre sus poemas o quizá una simple preferencia debido a la sonoridad y elocuencia de la expresión. Este sería un tema interesante para estudiar más en profundidad en relación con la función métrica (inicio, final de verso y hemistiquios) y semántica de tales bigramas.

## 5.2. Grafos de bigramas en Catulo, Propertio y Tibulo

Para trabajar con los grafos de bigrama en Catulo se utilizaron las librerías *tidyverse*, *tidytext*, *igraph*, *ggraph* y *grid*. Por tratarse de poesía, pareció conveniente dejar expuestos los bigramas más significativos, es decir, aquellos cuya frecuencia de aparición es superior a dos.

El grafo de los bigramas que contiene los términos que aparecen en los estribillos los muestra en redes separadas y con una arista más oscura de acuerdo su número de ocurrencias (*Hymen, Hymenaeae, currite, fusi, etc.*).

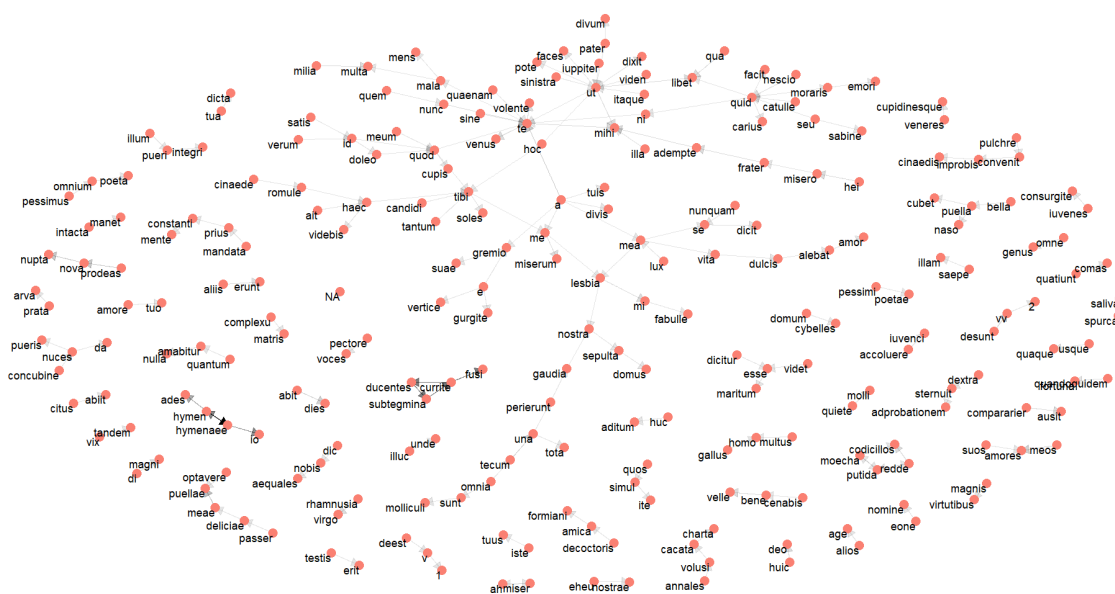


Figura 23. Grafo de bigramas en Catulo. Fuente: Elaboración propia.

El grafo también refleja el conjunto de palabras que refieren al dolor de haber perdido al hermano, *misero*, *frater*, *adempte*, *mihi* e *illa*. Con respecto al pronombre se debe ser cuidadoso, pues en un análisis apresurado podría pensarse que se refiere siempre a Lesbia, ya que en el carmen 68 Catulo pasa de hablar del dolor por la muerte de su hermano al amor de Lesbia y sus vicisitudes. Sin embargo, no es así: una lectura del poema muestra que aparece dos veces referido a Roma; luego, sí se refiere a Lesbia en dos oportunidades, cuando relata que no vino a su encuentro conducida por la diestra paterna en el verso 143 y narra además que ella podría marcar el día más feliz (el del encuentro amoroso) con la piedra más blanca (v. 148). Hay una tercera aparición del pronombre en el verso 153 referido a días. La forma del nominativo del femenino singular y el plural neutro es la misma en latín. Varias veces en el carmen 68 Catulo refiere a su hermano con el término *frater* como se dijo más arriba.

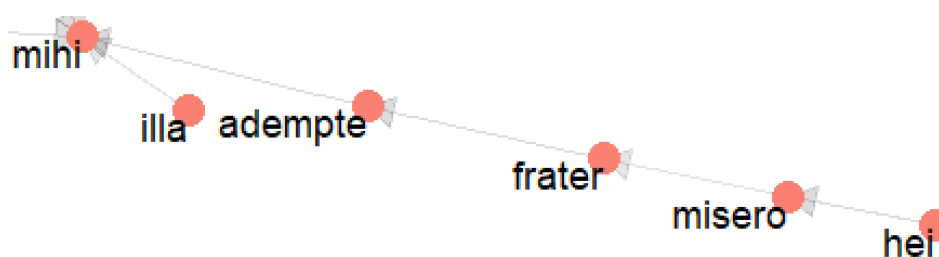


Figura 24. Vista del sector del grafo que refiere al conjunto de palabras vinculadas a la muerte del hermano de Catulo. Fuente: Elaboración propia.

Resulta interesante la vinculación de *Lesbia* con términos como *vita*, *lux*, *mea*, *amor*, *dulcis*, *alebat*, pero también *se*, *dicit* y *nunquam*. Lesbia se vincula a la vida, la luz, la dulzura, pero también podríamos pensar *dice*, *que ella*, *nunca*.

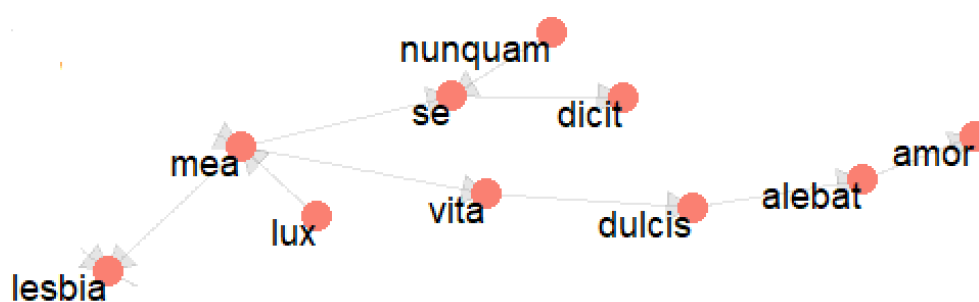


Figura 25. Vista del sector del grafo que refiere a Lesbia. Fuente: Elaboración propia.

El mismo grafo, diseñado con el poemario de Propertio proporciona otros pequeños detalles sobre las co-apariciones de términos amorosos: *Cynthia* se vincula con *mea*, *mihi* y *tibi*. Al igual que en el caso de Catulo a través de *mea* se vincula *musa*, *vita* y *lux*. Otro caso especial para destacar es el término *puella* que aparece junto a *docta* (sabia), pero también *lenta* (flexible pero también tenaz o, por el contrario, perezoso) y *dura* (dura, áspera).





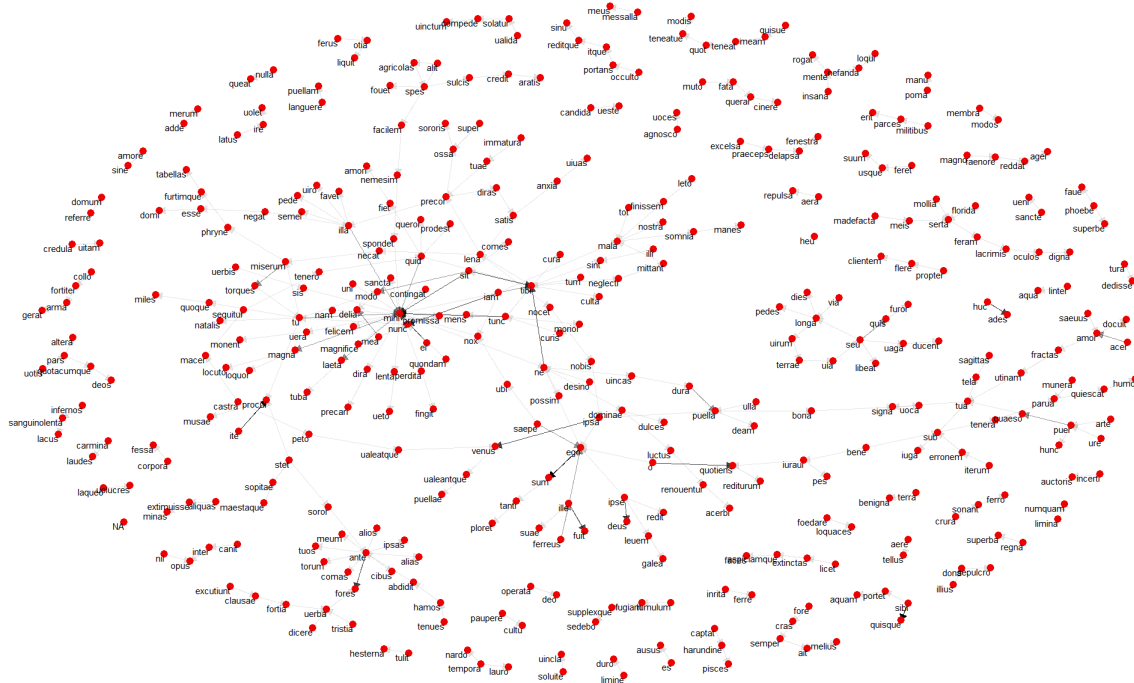


Figura 27. Grafo de bigramas en Tibulo. Fuente: Elaboración propia.

### 6. CO-CURRENCIAS

También se analizó la co-aparición de términos, es decir, la tendencia que tienen dos palabras a aparecer juntas en el texto. En este caso, para la comparación por pares (*pairwise*) se utilizó la librería *widyr*. La librería permite analizar el coeficiente de correlación *phi*, un indicador estadístico que permite saber cuán a menudo aparecen juntas dos palabras con referencia a cuántas veces están separadas. Se filtraron los términos *amor*, *Lesbia*, *pater* y *virgo* para ver con qué otras palabras co-ocurrían. Una mirada breve a las tablas pudo dar una cierta idea de algunos pares conceptuales. En la tercera columna se exhibe el número de veces que estos términos co-aparecen.

1	<i>amor tuus</i>	3
2	<i>amor dulcis</i>	3
3	<i>amor vita</i>	2
4	<i>amor alebat</i>	2
5	<i>amor tibi</i>	1
6	<i>amor hoc</i>	1
7	<i>amor ut</i>	1
8	<i>amor fuit</i>	1
9	<i>amor magnus</i>	1
10	<i>amor cuius</i>	1

Tabla 2. Co-apariciones del sustantivo *amor* en Catulo. Fuente: Elaboración propia.

El *amor tuus* (amor tuyo) es *dulcis* (amor), es *magnus* (grande) y se vincula a *vita* (vida). Este amor *alebat* (alimentaba), curiosamente en pasado, así como *fuit* (fue). Es un amor que ocupa el espacio del pasado.

1	<i>lesbia mi</i>	4
2	<i>lesbia mea</i>	3
3	<i>lesbia me</i>	3
4	<i>lesbia illa</i>	2
5	<i>lesbia dicit</i>	2
6	<i>lesbia nostra</i>	2
7	<i>lesbia quod</i>	1
8	<i>lesbia quem</i>	1
9	<i>lesbia tenere</i>	1
10	<i>lesbia quid</i>	2

Tabla 3. Co-apariciones del nombre *Lesbia*. Fuente: Elaboración propia.

*Lesbia* está vinculada mayormente a *ego*, el pronombre personal de primera persona y a *mea*, el posesivo. En tercer lugar, coaparecen *Lesbia illa*, *aquella Lesbia*, que vincula a *Lesbia* con el deíctico que indica mayor lejanía física y emocional.

1	<i>virgo rhamnusia</i>	3
2	<i>virgo quod</i>	2
3	<i>virgo manet</i>	2
4	<i>virgo intacta</i>	2
5	<i>virgo quaecumque</i>	1
6	<i>virgo patrona</i>	1
7	<i>virgo quicum</i>	1
8	<i>virgo primum</i>	1
9	<i>virgo ut</i>	1
10	<i>virgo mihi</i>	1

Tabla 4. Co-apariciones del sustantivo *virgo* en *Catulo*. Fuente: Elaboración propia.

En el caso de *virgo*, aparece en primer lugar con el apelativo con el que se llama a *Némesis*, cuyo papel es importante en el carmen 68, dedicado a las bodas de *Tetis* y *Peleo*. Lo interesante es que el concepto de *virgen* se vincule a *manet*, parece ser algo que debe *permanecer* o *mantenerse* y cuyo adjetivo más frecuente es *intacta*, es decir, sin tocar, algo muy sugerente en la mentalidad romana y la moral cívica de los epitalamios de *Catulo*. La esposa debe ser virgen porque debe garantizar que los hijos serán herederos legítimos del padre. La virginidad es la garantía sucesoria de los bienes de la familia, de la *gens* (Nusch, 2020, p. 89).

1	<i>pater ut</i>	2
2	<i>pater divum</i>	2
3	<i>pater cui</i>	1
4	<i>pater illius</i>	1
5	<i>pater saepe</i>	1
6	<i>pater inde</i>	1
7	<i>pater verum</i>	1
8	<i>pater cubile</i>	1
9	<i>pater quibus</i>	1
10	<i>pater necesse</i>	1

Tabla 5. Co-apariciones del sustantivo *pater* en *Catulo*. Fuente: Elaboración propia.

Con respecto al caso de *pater*, es necesario reconocer que la palabra es más ambigua porque no solo aparece vinculada al matrimonio y los epitalamios sino también a Júpiter, el dios padre, y al hermano en los poemas dedicados a recordar su muerte. Términos como *virgo*, *pater* y *mater* no son exclusivos de la temática epitalámica aunque se refiera a ellos con frecuencia en estos poemas. Además, el significado de estos tres adjetivos debe elucidarse en el contexto del poema en cuestión para saber si no se refieren en realidad a una deidad.

Luego de observar las co-apariciones de términos en tablas se calculó el coeficiente de correlación y se generó un gráfico para los cuatro términos. Como el resultado de calcular el coeficiente *phi* es siempre un número comprendido entre +1 y -1, para interpretar los resultados se utilizó la siguiente tabla:

- 1 Relación lineal descendente (negativa) perfecta
- 0.7 Relación lineal descendente (negativa) fuerte
- 0.5 Relación lineal descendente (negativa) moderada
- 0.3 Relación lineal descendente (negativa) débil
- 0 No existe relación lineal
- + 0.3 Relación lineal ascendente (positiva) débil
- + 0.5 Relación lineal ascendente (positiva) moderada
- + 0.7 Relación lineal ascendente (positiva) fuerte
- + 1 Relación lineal ascendente (positiva) perfecta

Tabla 6. Interpretación de los valores de la correlación. Fuente: Rumsey (2013).

Además, para ahorrar tiempo y recursos de procesamiento se calculó previamente la correlación de los términos que aparecían al menos cinco veces en el poemario. Veremos que esta metodología no fue del todo útil para el caso de Propertio.

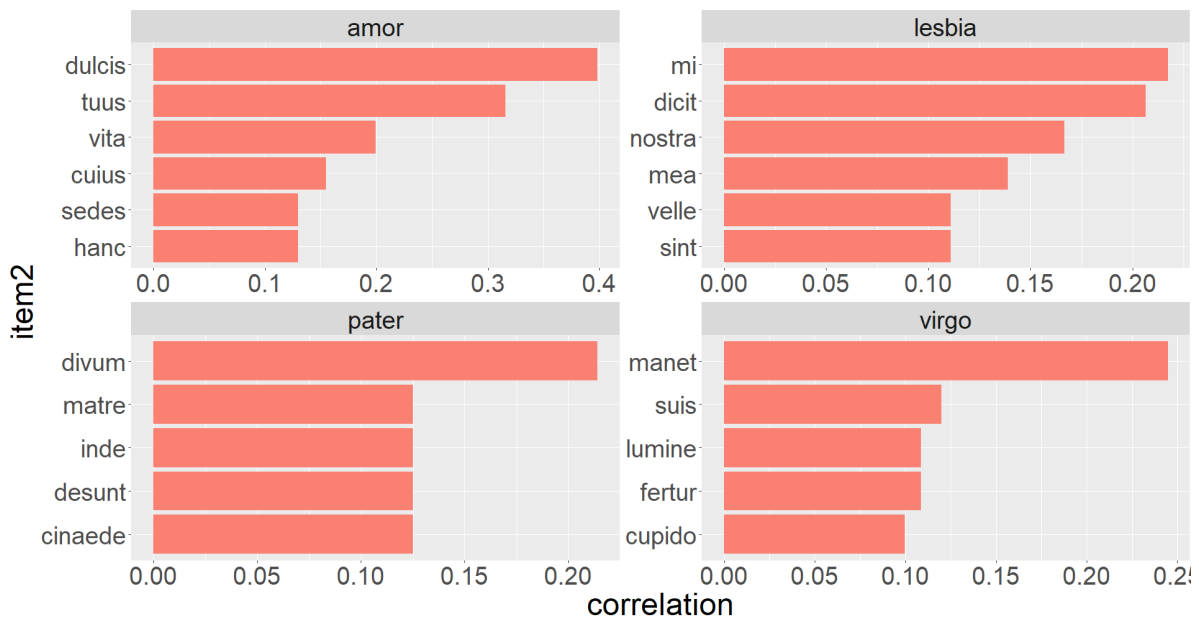


Figura 28. Cuadros de correlación de los términos *amor*, *Lesbia*, *virgo* y *pater* con sus respectivos coeficientes  $\phi$ . Los términos también se filtraron según aparecieran al menos cinco veces en el poemario. Fuente: Elaboración propia.

En el caso de Catulo se trata siempre de correlaciones de relación lineal ascendente débiles, salvo en el caso de *amor* que es moderada, pero la cantidad de términos y las particularidades del lenguaje figurativo comentadas en un trabajo anterior (Nusch, 2021) podrían explicar este tipo de correlación.

En este cuadro se podría interpretar, por ejemplo, que la *virgo* es algo que es *fertur*, *llevado*, *conducido*. La novia del ritual romano siempre es llevada, *deducta*, *conducida*, en la ceremonia matrimonial. El amor es dulce y se vincula con la vida, *Lesbia* es el sujeto que hace cosas al poeta, aparece junto al pronombre personal en acusativo *mi* y además *Lesbia* habla frecuentemente *dicit* a Catulo y de Catulo en los carmina 83 y 92.

Para el caso de Propertio, *Cynthia* aparece relacionada con más o menos el mismo adjetivo que *Lesbia*: *mea*, y con toda la constelación de pronombres posesivos y personales del yo poético más de una vez. También debe aclararse que el posesivo *mea* tiene la función de identificar a esa *Cynthia* en particular de otras del mismo nombre. No necesariamente el poeta esté pensando en una posesión concreta. *Cynthia* además es presentada como *sola* (única) y *prima* (la primera).

1	<i>cynthia tibi</i>	7
2	<i>cynthia mihi</i>	5
3	<i>cynthia mea</i>	4
4	<i>cynthia erit</i>	4
5	<i>cynthia sola</i>	4
6	<i>cynthia tu</i>	3
7	<i>cynthia fuit</i>	3
8	<i>cynthia roma</i>	3
9	<i>cynthia nobis</i>	3
10	<i>cynthia prima</i>	2

Tabla 7. Co-apariciones del nombre *Cynthia* en Propertio. Fuente: Elaboración propia.

Curiosamente, las apariciones de *virgo* son mucho menores en el poemario de Propertio, mucho más extenso que el de Catulo. A pesar de las pocas apariciones, el término sí co-aparece con otros dos del campo del matrimonio *dotem* (dote) y *desponderat* (desposaba, prometía en matrimonio), pero aplicado a Escila en el poema 19 del Libro III.

1	<i>virgo hanc</i>	1
2	<i>virgo igitur</i>	1
3	<i>virgo dotem</i>	1
4	<i>virgo desponderat</i>	1
5	<i>virgo hosti</i>	1

Tabla 8. Co-apariciones del término *virgo* en Propertio. Fuente: Elaboración propia.

El término *pater*, en Propertio, no da pistas de vincularse a términos exclusivamente matrimoniales en el poemario.

1	<i>pater ante</i>	1
2	<i>pater mihi</i>	1
3	<i>pater lumina</i>	1
4	<i>pater iam</i>	1
5	<i>pater illa</i>	1
6	<i>pater saepe</i>	1
7	<i>pater meae</i>	1
8	<i>pater vela</i>	1
9	<i>pater iura</i>	1
10	<i>pater mollia</i>	1

Tabla 9. Co-apariciones del término *pater* en Propertio. Fuente: Elaboración propia.

El término *Amor*, por supuesto, tiene muchas co-apariciones:

1	<i>amor ipse</i>	5
2	<i>amor tempore</i>	3
3	<i>amor caput</i>	2
4	<i>amor nullo</i>	2
5	<i>amor tardus</i>	2
6	<i>amor deus</i>	2
7	<i>amor amat</i>	2
8	<i>amor artificem</i>	2
9	<i>amor tibi</i>	2
10	<i>amor carmina</i>	2

Tabla 10. Co-apariciones del sustantivo *amor* en Propertio. Fuente: Elaboración propia.

Se hace necesario notar que *amor* aparece referido enfáticamente como *ipse* (el mismo) y que es algo que acontece *tempore* (en el tiempo o en un tiempo particular que puede ser lejano). El amor es algo que además se relaciona a *caput* (cabeza), no olvidemos que es el dios Amor quién posa la planta del pie sobre la cabeza del poeta en el primer poema del Libro Primero. El amor además es presentado como *tardus* (lento o tardío). Es un *deus* (un dios) que no ama los *artificem* (artífices), el poeta lo dice en los dos primeros poemas del Libro I.

En cuanto a la correlación, también es débil en el caso de Propertio. Como en el cuadro anterior de Catulo, aquí se muestran los términos que aparecen al menos cinco veces en el poemario.

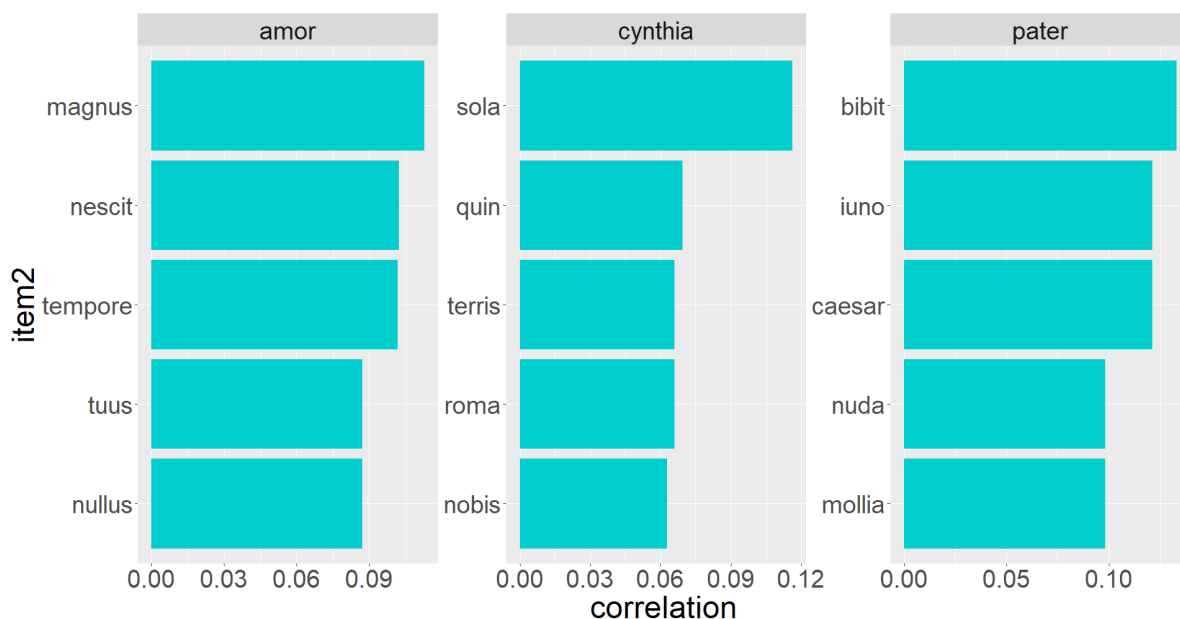


Figura 29. Cuadros de correlación de los términos *amor*, *Cynthia* y *pater* en Propertio con sus respectivos coeficientes  $\phi$ . El término *virgo* no se encuentra en este cuadro porque no aparece al menos cinco veces en el poemario. Fuente: Elaboración propia.

Al generar el cuadro anterior el término *virgo* no aparecía porque no tenía un nivel de co-apariciones superior a 5, con lo cual se debió volver a calcular la correlación, esta vez de todos los términos del corpus para poder acceder a las co-apariciones de *virgo*. El resultado fue que, a

pesar de ser un término que aparecía pocas veces a diferencia de otros, su correlación con los términos que co-aparecía era muy alta.

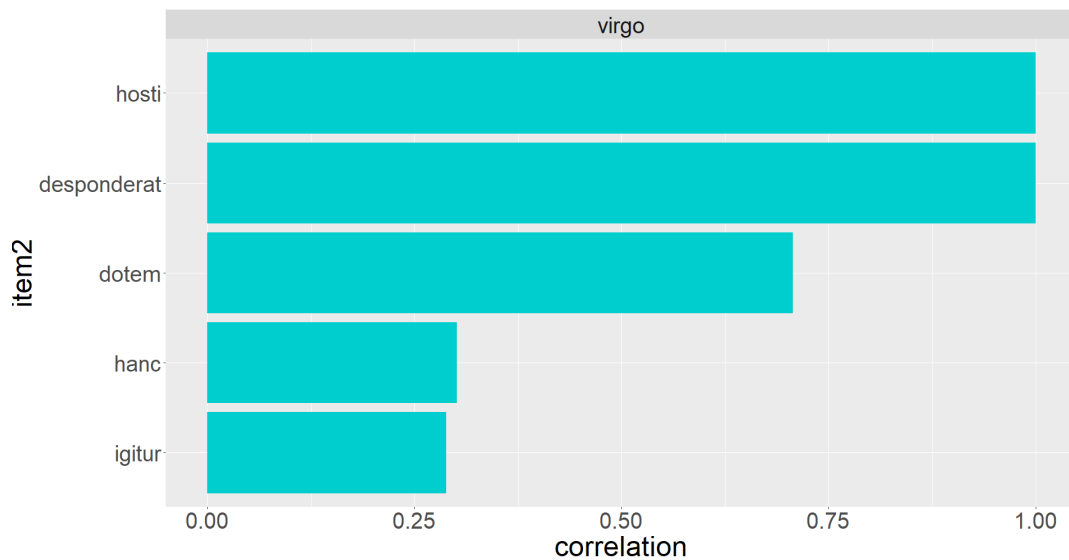


Figura 30. Cuadros de correlación de los términos *virgo* en Propertio. Para poder mostrarlo se debió calcular la correlación del poemario completo, el proceso fue más largo y costoso en términos de procesamiento, pero la correlación mostrada fue muy alta. Fuente: Elaboración propia.

Para Tibulo los pares de palabra referentes a la amada son similares a los de Propertio en cuanto a los pronombres personales y posesivos o el circunstancial *tecum* (contigo).

1	<i>delia mea</i>	5
2	<i>delia mihi</i>	2
3	<i>delia modo</i>	2
4	<i>delia tecum</i>	2
5	<i>delia labor</i>	1
6	<i>delia ipse</i>	1
7	<i>delia nam</i>	1
8	<i>delia tibi</i>	1
9	<i>delia custos</i>	1
10	<i>delia quoque</i>	1

Tabla 11. Co-apariciones del nombre *Delia* en Tibulo. Fuente: Elaboración propia.

Con Némesis<sup>10</sup>, las co-apariciones son menores, encontramos nombres como *luxuria* (exhuberancia, suntuosidad, lujuria) y adjetivos más positivos como *placido* (agradable, apacible) o términos como *vultu* (rostro).

<sup>10</sup> Tibulo tuvo dos amadas y se refería a la segunda de ellas como Némesis (Venganza). No se trata aquí de la misma diosa Némesis de la que habla Catulo, aunque claramente Tibulo elige el seudónimo de su amada en honor a la divinidad.

1	<i>nemesis mea</i>	2
2	<i>nemesis modo</i>	1
3	<i>nemesis utque</i>	1
4	<i>nemesis luxuria</i>	1
5	<i>nemesis fluat</i>	1
6	<i>nemesis urbem</i>	1
7	<i>nemesis placido</i>	1
8	<i>nemesis videat</i>	1
9	<i>nemesis vultu</i>	1

Tabla 12. Co-apariciones del nombre Némesis en Tibulo. Fuente: Elaboración propia.

*Virgo* co-aparece solo una vez con otras cuatro palabras: con *referre* (llevar de nuevo), con *domum* (a la casa), con *lumina* (luces u ojos en sentido figurado) y *sicca* (seca, árida pero también sana, robusta).

1	<i>virgo referre</i>	1
2	<i>virgo domum</i>	1
3	<i>virgo lumina</i>	1
4	<i>virgo sicca</i>	1

Tabla 13. Co-apariciones del sustantivo *virgo* en Tibulo. Fuente: Elaboración propia.

El término *pater* aparece en Tibulo junto a términos como *parce* (perdona) o *dicere* (decir), *dedit* (dio), *tristia* (cosas tristes), etc. En ningún caso se podría afirmar a simple vista que se trata de un contexto epitalámico.

1	<i>pater ipse</i>	3
2	<i>pater musta</i>	3
3	<i>pater possim</i>	2
4	<i>pater tua</i>	2
5	<i>pater parce</i>	2
6	<i>pater dicere</i>	2
7	<i>pater dedit</i>	2
8	<i>pater tristia</i>	2
9	<i>pater posse</i>	2
10	<i>pater abdere</i>	2

Tabla 14. Co-apariciones del sustantivo *pater* en Tibulo. Fuente: Elaboración propia.

En cuanto a *amor*, aparece junto a *mutuus*, *mutuo* o *recíproco*, *misero*, *pobre*, *utinam*, *ojalá* y *saevusm cruel*. En la personificación como deidad *Amor* es lógico que porte *sagittas*, *flechas* y cuando conduce al desamor es también esperable que co-aparezca con *misero*, *pobre*.



1	<i>amor mutuus</i>	3
2	<i>amor acer</i>	3
3	<i>amor mihi</i>	2
4	<i>amor tua</i>	2
5	<i>amor quid</i>	2
6	<i>amor utinam</i>	2
7	<i>amor docuit</i>	2
8	<i>amor sagittas</i>	2
9	<i>amor misero</i>	2
10	<i>amor tela</i>	2

Tabla 15. Co-apariciones del sustantivo *amor* en Tibulo. Fuente: Elaboración propia.

La correlación de estos pares de palabras también es débil.

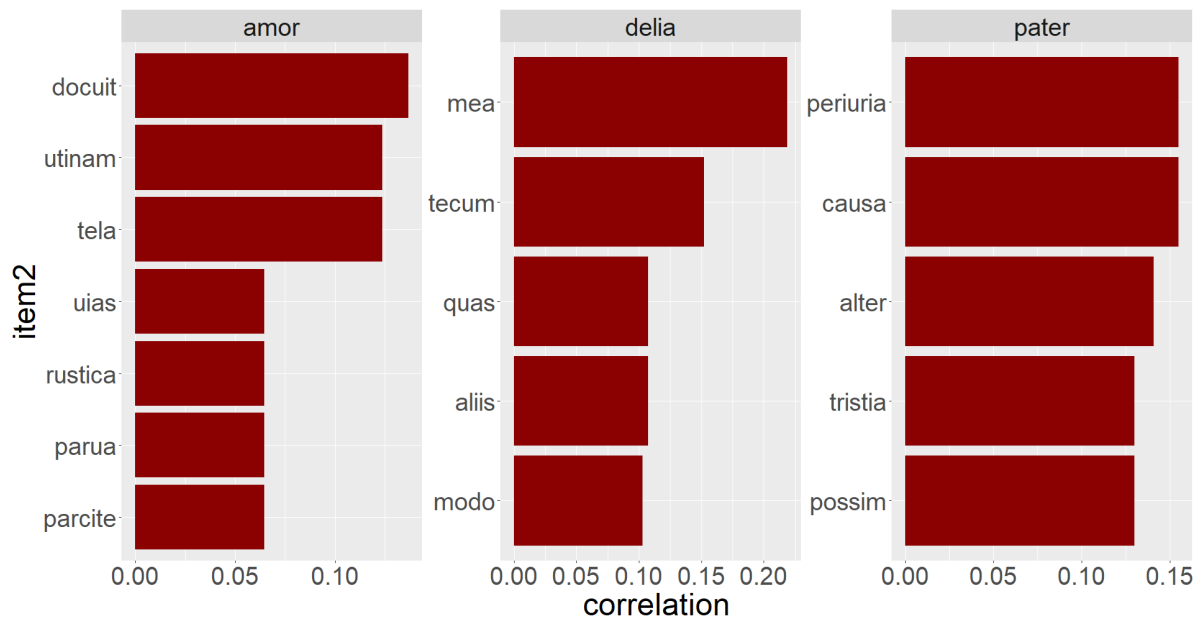


Figura 31. Cuadros de correlación de los términos *amor*, *Delia* y *pater* en el poemario de Propertio con sus respectivos coeficientes phi. Fuente: Elaboración propia.

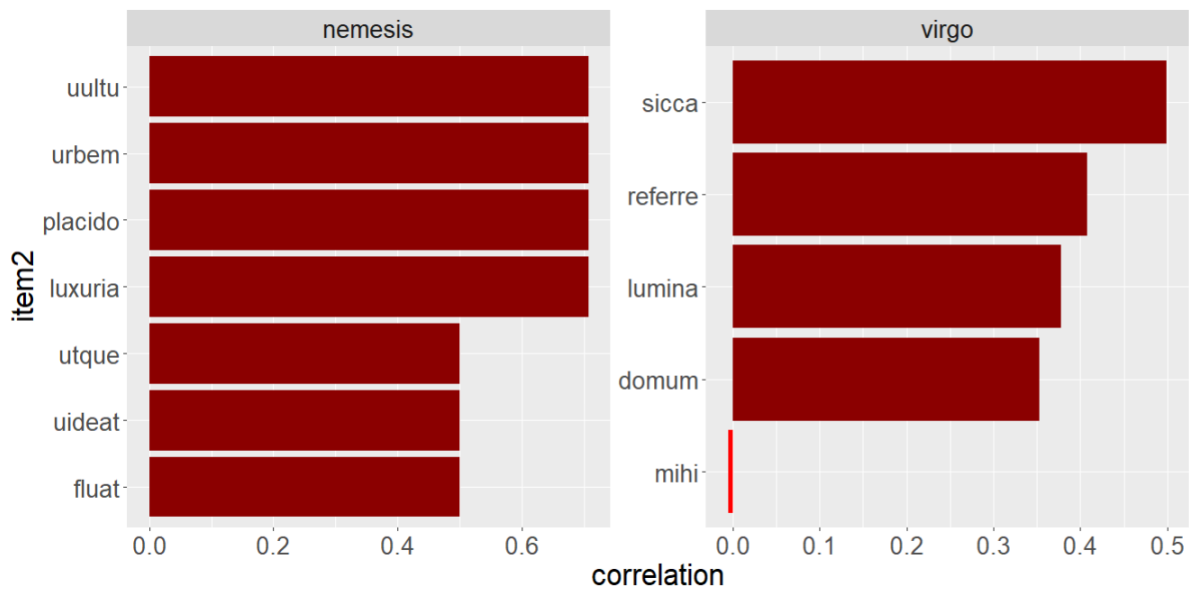


Figura 32. Cuadros de correlación de los términos *amor*, *Delia* y *pater* en el poemario de Propertio con sus respectivos coeficientes  $\phi$ . Fuente: Elaboración propia.

El gráfico de correlación de palabras de Tibulo muestra un campo semántico nuevo vinculado a su ideal de amor ubicado en los campos por medio del adjetivo *rustica* (rural). El amor se eleva a un plano en que ocupa un lugar central entre las cosas más maravillosas de la naturaleza (Grimal, 2000, pp. 191-192). Hay en el poeta un intento de sacralización del amor que va mucho más allá de lo efímero o sensual. *Mihi* aparece destacado en otro color puesto que su correlación es negativa respecto de *virgo*, es decir, tiene muy pocas probabilidades de co-ocurrir.

Por último, se realizó un grafo para los grupos (agrupaciones) de palabras presentes en los poemas de Catulo.

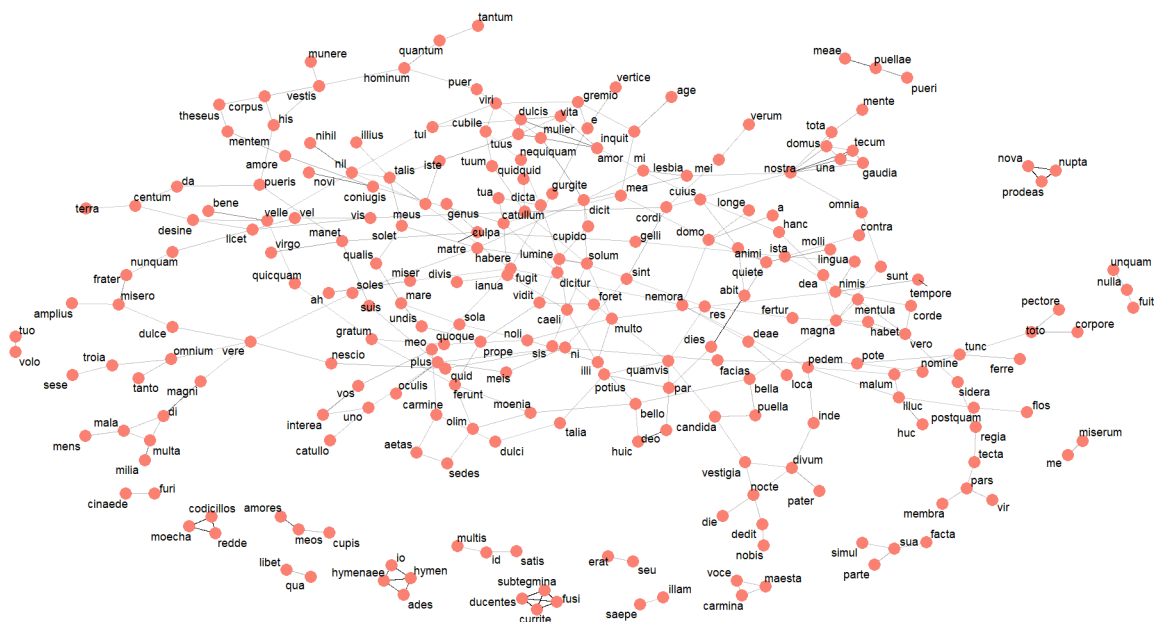


Figura 33. Grafo de agrupaciones de palabras en Catulo con un coeficiente de correlación mayor a 0.15. Fuente: Elaboración propia.



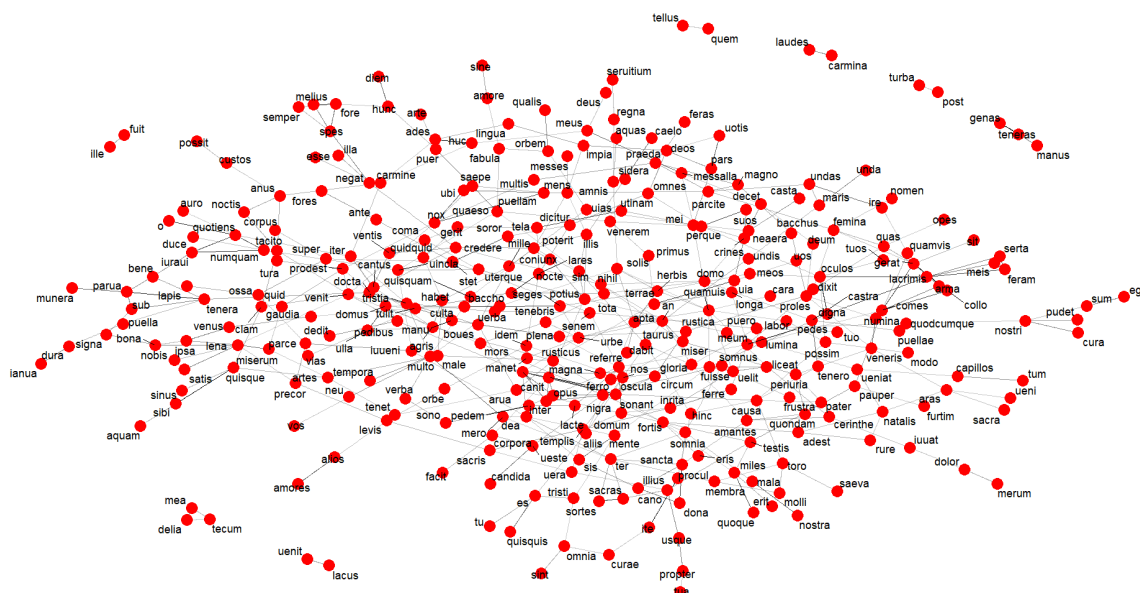


Figura 35. Grafo de agrupaciones de palabras en Tibulo con un coeficiente de correlación mayor a 0.15.  
Fuente: Elaboración propia.

## 7. CONCLUSIONES

El presente artículo exploró las posibilidades de análisis de los tres corpus con herramientas computacionales. El resultado de esta primera aproximación no revela un avance significativo ni sumamente original puesto que se trata de una investigación en estado inicial. Sin embargo, algunos resultados no dejan de despertar optimismo sobre el potencial de las herramientas empleadas (y de otras aún por emplear) en futuras tareas de análisis textual. Tanto el universo erótico elegíaco como el imaginario matrimonial aparecen en el análisis del *Corpus Catullianum*. La especificidad de los términos familiares para los epitalamios, sin embargo, debe tomarse con cautela.

En cuanto a la exploración del léxico de Catulo y su comparación con el de los otros dos poetas puede concluirse que la principal característica es el predominio de las formas pronominales de primera y segunda persona (*ego, tu, tibi, mihi*) así como algunos términos propios del léxico poético elegíaco (*puella, Lesbia, y amores*). Otros términos que podrían identificarse con el campo semántico del matrimonio fueron *virgo, pater, Himenaeae* e *Hymen*. En el caso de Propertio los términos *puella, amor* y *Cynthia* fueron los preponderantes. En Tibulo aparece *Venus* como una novedad, una diosa no tan referida por los otros dos poetas. Para el caso de los términos repetidos más de tres veces en algún poema, *virgo, nupta, concupine, amor, Cynthia* y *Delia* fueron algunos de los más utilizados por los poetas.

A su vez, la comparación con los corpus de Propertio y Tibulo arroja algunas pistas sobre el vocabulario más específicamente amoroso-elegíaco en lo que refiere a los términos más utilizados.

Muchos términos eróticos prevalecieron en los gráficos y en los diferentes elementos analizados. Una mejor ingeniería de características con certeza podrá revelar más detalles concluyentes sobre la naturaleza de los poemas amorosos y los poemas epitalámicos, así como sus particularidades y rasgos distintivos susceptibles de ser cuantificados y procesados automáticamente.

En cuanto al análisis general de los textos, la limpieza y el pre-procesamiento de los datos se procuró mejorar aspectos relativos a la metodología que se había utilizado en publicaciones anteriores. Se examinó esta vez una edición completa de Propertius y se descartó el criterio de eliminación de palabras vacías propias por considerarlo un enfoque espúreo que excluye términos que no deben ser eliminados. Se ensayó un enfoque que tuviera en cuenta y redujera una supuesta interferencia de los términos repetidos en estribillos, pero luego también se desechó tal metodología ya que la repetición de versos y la presencia de estribillos es sin duda uno de los tantos aspectos estilísticos delineados deliberadamente por el autor. Ese breve apartado, en particular, debería considerarse como una reflexión sobre datos negativos que, además, posibilita la explicación de determinadas decisiones de análisis en el resto del trabajo, y cuya publicación se mantiene como ayuda para la reflexión de otros investigadores que se encuentren con las mismas problemáticas.

El uso de etiquetado gramatical o *Part-of-speech Tagging* (POS Tagging) volvió a mostrar la aparición de términos como *virgo*, *pater* y *mater*. Los términos familiares son comúnmente referidos en los poemas epitalámicos ya que por su propia naturaleza se trata de roles protagónicos en los ritos y ceremonias matrimoniales. En estos casos, no obstante, deben ser examinados con prudencia ya que muchas veces se trata de palabras usadas en contextos no nupciales o bien como formas de referirse a determinados dioses. Por ejemplo, suele aludirse a Júpiter como *pater divum*, es decir, *padre de los dioses*, y Catulo utiliza *virgo* para referirse a la diosa Némesis. Entre los sustantivos recuperados con la técnica de *Part-of-speech Tagging* (POS Tagging) aparecieron *oculus*, *Catullus*, *furor Venus*, *amicus*, *fides*, *domina*, *os*, *pes*, *cura*, *nox*, y *corpus* y entre los verbos se destacaron moderadamente *libet*, *duco*, *milito*, *video*, *gaudeo*, *cupio*, *timeo*, *placeo* y *sentio*. Los adjetivos por sí solos no mostraron una clara pertenencia temática al vocabulario amoroso.

El estudio de los bigramas también exhibió ciertos pares de términos propios del discurso amoroso elegíaco latino: *meae puellae*, *meos amores*, *Lesbia mi*, *puella fuit*, *nox mihi*, *mea Cynthia*, *mea musa* e *ipsa Venus*. Tanto en el plano de la deixis temporal (tiempos verbales en pasado y de aspecto perfecto) como en el plano de la deixis espacial (demostrativos con el mayor grado de lejanía) parece evidenciarse un cierto contraste entre dos posibles situaciones amorosas: un tiempo dichoso en el que el amor se presenta como algo que ocurre y no ha terminado y otro tiempo presente desde el que se recuerda un pasado lejano y dichoso, pero en el que el amor ya ha concluido y cuya evocación resulta amarga.

En cuanto a la repetición de los bigramas y trigramas y su idéntica colocación en el verso con fines rítmicos (como en el caso de los tres ejemplos en los que Catulo se refiere a su hermano) habría que analizar con mayor detenimiento, por ejemplo, si el mismo fenómeno ocurre en otros géneros poéticos más dados a las fórmulas repetitivas con fines mnemotécnicos como la épica y cuán común es su uso en los poemas del poeta veronés.

A diferencia de la exploración de términos sueltos, en los grafos de bigramas cobraron preponderancia, entre otros, los adjetivos aplicados a conceptos del discurso amoroso. Aparecen

palabras asociadas a las amadas elegíacas y el mismo contraste entre rasgos positivos (*vita, lux, mea, mihi, tibi, amor, dulcis, musa, puella docta, alebat*) y negativos (*nunquam, puella dura, acer, saevus, misero*).

En el caso de las co-ocurrencias también se puede observar el predominio de los términos anteriores referidos al amor y de los pronombres personales y posesivos y verbos en tiempo pasado que denotan el curso amoroso concluido. Un detalle destacable son las co-apariciones de *virgo* en el poemario de Propertio, con *dotem* y *desponderat*. No se trata de un poema epitalámico, pero el imaginario del matrimonio es utilizado para ilustrar el destino de Escila en el poema 19 del Libro III.

Todas las características anteriores podrían contribuir en futuras tareas de lectura distante a identificar y caracterizar poemas amorosos y epitalámicos. Por supuesto que no serán las únicas utilizadas y se deberá recurrir a otras que incluyan métodos sensibles al contexto semántico y sintáctico. Si bien para un humano no requiere mucho esfuerzo el identificar un poema amoroso y diferenciarlo de otro que no lo es, para tareas de análisis automático de textos un enfoque de bolsa de palabras como el utilizado en este artículo puede resultar insuficiente. Una comparación de léxicos más precisa por medio de métodos de clustering o el uso de técnicas como las que se aplican en el modelado de tópicos podrían ser vías interesantes que seguir.

En futuros trabajos se procurará utilizar tecnologías sensibles al contexto lingüístico como las nuevas herramientas disponibles que ya proporciona la inteligencia artificial y el uso de *embeddings*. Estas metodologías permitirán tener una mayor precisión en tareas que requieran la desambiguación de casos de homonimia o la detección de casos de sinonimia. El manejo de textos como secuencias en las que existe una clara interdependencia (anafórica, catafórica, etc.) entre sus diferentes elementos formantes constituye uno de los grandes avances de la última década en el área del Procesamiento del Lenguaje Natural (Devlin et al., 2019; Mikolov et al., 2013; Vaswani et al., 2017).

Como un pequeño corolario surgido de la interpretación hermenéutica de los resultados podría arriesgarse la hipótesis de que el análisis de una posible co-dependencia de términos pronominales de primera y segunda persona combinados con algunos de términos típicos de la deixis temporal y espacial y sumados a un vocabulario de palabras propias del discurso amoroso ofrecen un promisorio camino por el cual continuar con la tipificación de la poesía elegíaca latina.

## REFERENCIAS BIBLIOGRÁFICAS

Calame, C., y Pérez Rodríguez, E. (2002). *Eros en la antigua Grecia*. Akal.

Cerrato, L. M., y Chavez, R. F. (s. f.). *Perseus Classics Collection: An Overview*. Recuperado 24 de marzo de 2023, de <https://www.perseus.tufts.edu/hopper/text?doc=Perseus:text:1999.04.0053>

Csárdi, G., Nepusz, T., Müller, K., Horvát, S., Traag, V., Zanini, F., y Noom, D. (2023). *Igraph for R: R Interface of the Igraph Library for Graph Theory and Network Analysis* (v1.5.1) [Software]. Zenodo. <https://doi.org/10.5281/zenodo.8240644>

- Devlin, J., Chang, M.-W., Lee, K., y Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>
- Ernout, A., y Meillet, A. (2001). *Dictionnaire etymologique de la langue latine* (4e ed.).
- Fradejas Rueda, J. M. (2020). *Cuentapalabras*. <http://www.aic.uva.es/cuentapalabras/>
- Grimal, P. (2000). *El amor en la Roma antigua* (J. P. Tauste, Trad.). Ediciones Paidós.
- Harley, R. M. (2012). Amicitia en los poemas de Catulo. *Pensamiento Actual*, 12(18-19). <https://dialnet.unirioja.es/servlet/articulo?codigo=5897911>
- Katz, V. (2016). *The Complete Elegies of Sextus Propertius*. Princeton University Press.
- Lewis, C. S. (1936). *La alegoría del amor: Un estudio sobre tradición medieval* (2015.ª ed.). Encuentro.
- Martí, M. (2016, abril 8). Qué es el Web scraping? Introducción y herramientas. *Sitelabs*. <https://sitelabs.es/web-scraping-introduccion-y-herramientas/>
- Merrill, E. T. (1893). *Catullus; edited by Elmer Truesdell Merrill*. Boston Ginn. <http://archive.org/details/catulluseditedby00catuoft>
- Mikolov, T., Sutskever, I., Chen, K., Corrado, G., y Dean, J. (2013). *Distributed Representations of Words and Phrases and their Compositionality* (arXiv:1310.4546). arXiv. <https://doi.org/10.48550/arXiv.1310.4546>
- Moretti, F. (2013). *Distant Reading*. Verso.
- Müller, L. (1898). *Sex. Propertii Elegiae*. Teubner.
- Munzert, S., Rubba, C., Meißner, P., y Nyhuis, D. (2014). *Automated Data Collection with R: A Practical Guide to Web Scraping and Text Mining*. John Wiley & Sons.
- Murrell, P. (2005). *R Graphics*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781420035025>
- Nehrdich, S., y Hellwig, O. (2022). Accurate Dependency Parsing and Tagging of Latin. En R. Sprugnoli y M. Passarotti (Eds.), *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages* (pp. 20-25). European Language Resources Association. <https://aclanthology.org/2022.lt4hala-1.3>
- Nusch, C. J. (2020). *Los géneros del amor: La presencia de Safo en la poesía de Catulo* [Tesis, Universidad Nacional de La Plata]. <http://sedici.unlp.edu.ar/handle/10915/108707>
- Nusch, C. J. (2021). *Las Edades del Amor: Una propuesta para el proyecto Aetates Amoris destinado a la poesía amorosa* [Tesis, Universidad Nacional de Educación a Distancia (UNED)]. <https://doi.org/10.35537/10915/125629>
- Nusch, C. J. (2022). Lectura Cercana y Distante: Una propuesta de aproximación al estudio de la poesía amorosa desde la codificación en XML-TEI y la minería de textos. *Publicaciones de la Asociación Argentina de Humanidades Digitales*, 3. <https://doi.org/10.24215/27187470e037>
- Pedersen, T. L. (2022). *ggraph: An Implementation of Grammar of Graphics for Graphs and Networks* (2.1.0) [Software]. <https://cran.r-project.org/web/packages/ggraph/index.html>

- Pereiro, N. G. B. de F. (1968). *Originalidad y sinceridad en la poesía de amor trovadoresca*. <http://www.memoria.fahce.unlp.edu.ar/library?a=d&c=libros&d=Jpm122>
- Postgate, J. P. (1915). *Tibulli aliorumque carminum libri tres*. Scriptorum classicorum bibliotheca Oxoniensis.
- Rajaraman, A., y Ullman, J. D. (Eds.). (2011). Data Mining. En *Mining of Massive Datasets* (pp. 1-17). Cambridge University Press. <https://doi.org/10.1017/CBO9781139058452.002>
- Ramsay, S. (2011). *Reading Machines: Toward an Algorithmic Criticism*. University of Illinois Press.
- Robinson, D., Misra, K., Silge [aut, J., y cre. (2022). *widyr: Widen, Process, then Re-Tidy Data* (0.1.5) [Software]. <https://cran.r-project.org/web/packages/widyr/index.html>
- Rodríguez Adrados, F. (1996). *Sociedad, amor y poesía en la Grecia antigua* (Edición: edición). Alianza.
- Sierra, G., Bel, G., y Hernández, J. A. L. (2018). *Conteo de palabras*. Lingüística de corpus - Procesamiento de corpus textuales y orales. [http://www.corpus.unam.mx/cursocorpus/4\\_1\\_Conteo\\_de\\_palabras.html](http://www.corpus.unam.mx/cursocorpus/4_1_Conteo_de_palabras.html)
- Silge, J., y Robinson, D. (2017). *Text Mining with R: A Tidy Approach*.
- Straka, M., y Straková, J. (2019). Universal Dependencies 2.5 Models for UDPipe (2019-12-06). <http://ufal.mff.cuni.cz/udpipe>. <https://lindat.mff.cuni.cz/repository/xmlui/handle/11234/1-3131>
- Straka, M., y Straková, J. (2020). *UDPipe at Evalatin 2020: Contextualized Embeddings and Treebank Embeddings* (arXiv:2006.03687). arXiv. <https://doi.org/10.48550/arXiv.2006.03687>
- Straka, M., y Straková, J. (2021). *udpipe.models.ud.2.5—R package containing Universal Dependencies 2.5 Models for UDPipe* [Software]. <https://github.com/jwiffels/udpipe.models.ud.2.5> (Obra original publicada en 2020)
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., y Polosukhin, I. (2017). Attention is All you Need. *Advances in Neural Information Processing Systems*, 30. <https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html>
- Wickham, H. (2022). *rvest: Easily Harvest (Scrape) Web Pages* (1.0.3) [Software]. <https://www.vps.fmvz.usp.br/CRAN/web/packages/rvest/>