

“Valenciano es un idioma? Yo pensaba que era un catalán mal hablado...”

Análisis de discursos sobre el valenciano y el catalán en las redes sociales

Maija Rantala

Trabajo de Fin de Máster

Programa de Especialista en Lenguas, Español

Instituto de Lenguas y Traducción

Facultad de Humanidades

Universidad de Turku

Abril de 2026

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service

Trabajo de Fin de Máster

Programa de Especialista en Lenguas, Departamento de Español

Maija Rantala

“Valenciano es un idioma? Yo pensaba que era un catalán mal hablado...” Análisis de discursos sobre el catalán y el valenciano en las redes sociales

Páginas: 65 páginas, 7 páginas de apéndices

El propósito de este trabajo es analizar los discursos sobre el valenciano y el catalán en las redes sociales e inspeccionar qué tipo de ideologías lingüísticas reflejan. El debate sobre si el valenciano y el catalán son la misma lengua o lenguas separadas sigue dividiendo opiniones. Desde la aceptación del nombre “valenciano” en el Estatuto de Autonomía del año 1982, además de lo lingüístico, esta polémica posee un carácter político significativo. Por ende, hace falta adquirir una imagen del estado actual del conflicto sobre la posición del valenciano desde la perspectiva de los hablantes.

El corpus de este trabajo consiste en 2 164 comentarios recogidos de nueve hilos de discusión de ForoCoches que tratan sobre la relación entre el valenciano y el catalán. Se trata de una investigación de métodos mixtos, ya que se utilizan tanto métodos cuantitativos, ejecutando un modelado de tópicos y un análisis de colocaciones, como cualitativos, llevando a cabo un análisis crítico del discurso.

A través del modelado de tópicos se pueden determinar tres temas predominantes en el corpus: la comparación entre el catalán y el valenciano, el valenciano en el sistema educativo y la identidad nacional de los valencianos. El análisis de colocaciones permite la confirmación de los temas determinados en la fase anterior, así como el hallazgo de discursos que reflejan dichos temas e ideologías lingüísticas relacionadas con ellos para el subcorpus utilizado en el análisis cualitativo.

El *close reading* del subcorpus, compuesto de 22 comentarios, permite la inspección tanto de las ideologías como de la manera en que se aborda a la polémica. Se puede concluir que el factor más determinante es la identidad nacional, ya que influye significativamente en las actitudes e ideologías lingüísticas de los hablantes. Las actitudes e ideologías, por su parte, dan forma a la cultura lingüística de una sociedad, en la que se fundamentan la planificación y política lingüísticas.

Palabras clave: lingüística de corpus, sociolingüística, valenciano, modelado de tópicos, análisis crítico del discurso, ideologías lingüísticas

Índice

1	Introducción	5
2	Lengua y sociedad	8
2.1	Planificación y política lingüísticas	8
2.2	Ideologías y actitudes lingüísticas	10
2.2.1	Aproximación general a las ideologías y actitudes lingüísticas	11
2.2.2	Lengua e identidad	13
2.2.3	Ideologías lingüísticas en el contexto valenciano	15
2.3	Discursos sobre la lengua	18
3	Antecedentes sociohistóricos del valenciano frente al catalán	21
3.1	Historia sociolingüística del valenciano	21
3.2	Historia sociolingüística del catalán	23
3.3	Hipótesis acerca de la génesis del valenciano	24
4	Situación sociolingüística actual	27
4.1	Comunidad Valenciana	27
4.2	Cataluña	30
4.3	Diferencias entre el valenciano y el catalán	33
5	Lingüística de Corpus	35
5.1	Estudio del Discurso Asistido por Corpus	35
5.2	Modelado de tópicos	36
5.3	Análisis de colocaciones	38
6	Metodología	40
6.1	Consideraciones legales y éticas	40
6.2	Corpus	41
6.3	Métodos de investigación	42
7	Resultados del análisis	44
7.1	Modelado de tópicos	44
7.2	Análisis de colocaciones	48
7.3	Discursos	52

7.3.1	Jerarquización de las lenguas	52
7.3.2	Aspectos históricos	57
7.3.3	Identidad nacional	61
8	Conclusiones	66
	Bibliografía	70
	Apéndices	78
	Apéndice 1. Suomenkielinen lyhennelmä	78

1 Introducción

La realidad sociolingüística de una lengua se puede observar a través de las actitudes e ideologías lingüísticas reflejadas en los discursos de los hablantes. Hasta el día de hoy, el debate sobre si el valenciano y el catalán son la misma lengua o lenguas separadas divide las opiniones. Según la hipótesis repoblacionista, la lengua catalana llegó al Reino de Valencia cuando las familias procedentes de Cataluña empezaron a ocupar y repoblar el territorio anteriormente árabe después de la reconquista de Valencia del año 1238, lo que marcó el inicio de un cambio definitivo en la situación lingüística de la ciudad y su entorno (Calatayud, 2011). La hipótesis autoctonista sugiere que la lengua valenciana ya se hablaba en el territorio actual antes de la reconquista. De todas formas, recientemente y más precisamente en el año 1982, en el Estatuto de Autonomía se aceptó el nombre “Comunitat Valenciana” en vez de “País Valenciano”, así como la denominación de “valenciano” como lengua propia de la *Comunitat*. Se puede afirmar que desde entonces hasta ahora, este conflicto no solo trata de lo lingüístico, sino que divide la sociedad, y se ha arraigado en ella.

En este Trabajo de Fin de Máster voy a analizar los discursos sobre el valenciano y el catalán en los comentarios publicados en ForoCoche, la mayor plataforma de foros de España. El propósito es hallar qué tipo de discursos se encuentran de las lenguas y cómo se manifiestan diversas ideologías lingüísticas en ellos, sobre todo para interpretar el estado actual del conflicto sobre el estatus del valenciano desde la perspectiva de los ciudadanos. Cabe aclarar que en este trabajo se emplea el término “lengua” únicamente como categoría formal para referirse a las denominaciones “valenciano” y “catalán”, sin tomar postura sobre su consideración filológica o sociopolítica.

A pesar de que, en la actualidad, la propuesta repoblacionista se considera la única científica en las universidades españolas, el debate todavía persiste, tanto en el ámbito académico como en el político, institucional y social (Ballester, 2021: 10). Asimismo, aunque la Academia Valenciana de la Lengua, oficialmente *Acadèmia Valenciana de la Llengua* (AVL), en 2005 declaró públicamente que el valenciano era una variedad del catalán que podía llamarse “valenciano”, la ambigüedad sobre la denominación de la lengua no se ha solucionado. Esto es, en el año 2006 el nuevo Estatuto de Autonomía consideraba el valenciano como lengua autóctona y no solo cooficial, pero sin especificar si es una variedad del catalán o una lengua propia. Cabe investigar este debate, ya que la declaración de la AVL del año 2005 se recibió con reacciones de indignación por algunos grupos que la percibían como una demostración de deslealtad a la

región y su lengua por parte del Gobierno (Lado, 2011). Con el análisis de las publicaciones en las redes sociales, este estudio puede aportar un enfoque distinto a la situación actual de la polémica que sigue sin resolverse a pesar de que ya hayan transcurrido dos décadas desde la declaración de la AVL.

En efecto, el corpus consiste en 2 164 comentarios publicados entre los años 2015 y 2024 en nueve hilos de discusión que tratan sobre el valenciano y el catalán. Se ha recopilado a través de *web scraping* con el programa Octoparse (versión 8.7.2), que permite la extracción automática de datos. Para realzar el análisis cualitativo de menor escala, llevo a cabo un análisis del discurso basado en corpus, utilizando métodos de la Lingüística de Corpus como el modelado de tópicos con el programa Orange (versión 3.39.0), y el análisis de colocaciones con AntConc (versión 4.3.1). Cabe mencionar que como el análisis de colocaciones posibilita la inspección de las palabras clave en su contexto, su función como fase intermedia será, además de confirmar los tópicos predominantes, facilitar la recopilación del subcorpus para el Análisis Crítico del Discurso (ACD). Dado que se combinan métodos cuantitativos y cualitativos, se trata de una investigación mixta. Así pues, las preguntas de investigación son las siguientes:

- 1) ¿Cuáles son los temas predominantes en los hilos de discusión que tratan sobre el valenciano y el catalán?
- 2) ¿Qué tipo de discursos se encuentran sobre el valenciano y el catalán en el corpus?
- 3) ¿Cómo se construyen y reflejan las ideologías lingüísticas en los discursos?

Los factores que pueden contribuir a los resultados son, entre otros, la educación, las ideologías políticas y la identidad nacional. En la Comunidad Valenciana, la opinión difiere entre los que aceptan la unidad lingüística y los que defienden la singularidad del valenciano, mientras que en Cataluña, predomina el enfoque de que son la misma lengua con variedades regionales (Davidson, 2020). Efectivamente, aunque en algunos casos se pueda deducir, por ejemplo, la edad, la procedencia o los conocimientos lingüísticos del informante a través de la interpretación de su emisión, a causa de que se conserva la anonimidad de los informantes, no será posible determinar con certeza cómo influyen los estratos sociales y las localidades de residencia en sus actitudes. En cualquier caso, propongo las siguientes hipótesis:

- a) Prevaldrán los temas como la comparación de las lenguas, sus orígenes y la política.

- b) Los discursos serán de índole argumentativa, centrandose en las diferencias léxicas y fonológicas de las lenguas, así como en los hechos históricos y aspectos políticos.
- c) Se podrán observar diferentes ideologías lingüísticas a través de la expresión de opiniones así como la manera de referirse a cuestiones relacionadas con la situación sociolingüística.

En lo que concierne a la organización del presente trabajo, primero, presento el marco teórico, centrándome en cuestiones relacionadas con la lengua en la sociedad, como los conceptos de políticas, ideologías y actitudes lingüísticas, así como la relación de lengua e identidad. Además, introduzco algunas consideraciones sobre discursos sobre la lengua y qué es lo que las lenguas nos pueden contar sobre la sociedad. Después, sintetizo los antecedentes sociohistóricos de las lenguas valenciana y catalana, presento la situación sociolingüística actual de la Comunidad Valenciana y Cataluña, y exhibo algunas diferencias fonológicas y léxicas entre las lenguas. A continuación, antes de presentar los métodos de análisis, trato sobre unas consideraciones legales y éticas con respecto al *web scraping*, y, presento el corpus recopilado de los comentarios. Después, expongo los principales resultados del modelado de tópicos, el análisis de colocaciones y el ACD. Por último, cabe sintetizar los puntos más relevantes de esta investigación, así como reflexionar sobre, por una parte, qué es lo que se hubiera podido mejorar en el presente estudio y, por otra, algunas propuestas para posibles líneas de investigación para el futuro.

2 Lengua y sociedad

La sociolingüística se ocupa del estudio del lenguaje en su entorno social, examinando la realidad social a través de los medios de la lingüística. En efecto, como estudia la relación entre el uso del lenguaje y la estructura social, el núcleo de la mayoría de las definiciones se centra en el hecho de que la lengua refleja la sociedad. La macrosociolingüística, la perspectiva aplicable al presente estudio, examina lo que las comunidades hacen con sus lenguas, analizando su cambio, preservación y desaparición, las actitudes lingüísticas y las funciones de las diferentes formas lingüísticas en una comunidad lingüística (Nuolijärvi, 2000: 13-15).

Al revelar diferencias entre los grupos lingüísticos, las discusiones sociolingüísticas han tratado sobre la relación entre las lenguas estándar y otras formas lingüísticas, la enseñanza y la educación lingüísticas, las relaciones entre las lenguas mayoritarias y minoritarias, la socialización lingüística y la discriminación lingüística. De cara a la selección del método sociolingüístico, cabe destacar la necesidad de evaluar el contexto relacionado con la estructura social y la historia de la comunidad lingüística. Además, es importante considerar la percepción de que la variación lingüística describe el mundo y las estructuras sociales subyacen al lenguaje, puesto que, en realidad, se trata de una cuestión más compleja: la lengua forma parte de la realidad social, y es un elemento igual de importante que cualquier otro que la construye (Nuolijärvi, 2000: 19-25).

Para exponer algunas teorías relevantes con respecto a esta investigación, primero me centro en la planificación y política lingüísticas, fundamentales al enfrentarse con cuestiones sobre las lenguas minoritarias, ya que muchas veces tienen como objetivo modificar el papel y las funciones de las lenguas. Después, trato sobre el estudio de las ideologías y actitudes lingüísticas, que pueden surgir, por ejemplo, en las conversaciones sobre la denominación de las lenguas. Además, el estudio de las actitudes lingüísticas y la aplicación de sus hallazgos puede resultar beneficioso para ámbitos como la política, la educación y la planificación lingüística (Rojas, 2012). Por último, trato sobre el papel que desempeña el Análisis Crítico del Discurso (ACD) en la interpretación de cuestiones sociolingüísticas, por ejemplo facilitando la observación de las ideologías y actitudes que reflejan la sociedad en los discursos sobre la lengua.

2.1 Planificación y política lingüísticas

Una lengua viva debe responder a los cambios que suceden dentro de una sociedad. El motivo principal de la política y planificación lingüísticas en una sociedad democrática es la

necesidad de una lengua común para los asuntos administrativos, la educación, la comunicación y la vida económica y cultural. Además, este tipo de desarrollo deliberado de un idioma puede contribuir a objetivos políticos. Como caso típico, cuando una nación pequeña construye su entidad estatal, el fomento planificado de una propia lengua nacional como símbolo central de la identidad nacional suele formar parte de la aspiración ideológica para adquirir una voluntad unificada entre la población. Entonces, el estatus sólido de la lengua propia de una comunidad pequeña refuerza tanto la identidad individual como la colectiva (Sajavaara, 2000: 67). Además, se ha acentuado la función de la lengua en la formación de una identidad colectiva al examinar las lenguas de los grupos étnicos, la conexión entre la lengua y la identidad, el nacionalismo lingüístico y la sustitución lingüística (*language shift*). No obstante, el estatus de la lengua se ha constituido como un aspecto clave en los estudios con enfoque sociológico sobre las condiciones lingüísticas de diferentes comunidades (Nuolijärvi, 2000: 18).

Según Ricento (2005: 5), las metas de la política lingüística muchas veces pretenden, de algún modo, mejorar o modificar el papel social y las funciones de la lengua, pero no pueden invalidar los efectos de la cultura lingüística de una sociedad. Schiffman (2012) argumenta que, primordialmente, la política lingüística es una construcción social que se fundamenta en la cultura lingüística, es decir, otros elementos conceptuales como los sistemas de creencias, las actitudes y los mitos. Aunque no es común que las lenguas provoquen guerras, se puede afirmar que han sido un factor significativo en disturbios étnicos que han sido hasta mortales (Horowitz, 2001: 1). Estos conflictos pueden ser de índole comunitaria, racial, religiosa, tribal o lingüística. En este tipo de disturbios lingüísticos la política lingüística nacional, entre otras cuestiones, es un tema regular (Spolsky, 2004: 2).

Pese a que, a primera vista, el concepto de la política lingüística puede parecer sencillo – se determinan las normas a través de la planificación lingüística y, llevándolas a la práctica, se establecen las normas del uso del idioma – no lo es. Schiffman (2012: 2) hasta cuestiona si la política lingüística es algo que, de hecho, puede ser gestionada por el ser humano de forma explícitamente intervencionista. Para ilustrar más, generalmente hay una diferencia entre la política tal y como se declara, *de iure*, y la política tal y como funciona en la práctica, *de facto*. Esto se puede explicar con la posibilidad de algún cambio histórico, por ejemplo, el aumento del número o del poder político de una minoría antes poco significativa, como consecuencia de la inmigración, una mayor tasa de natalidad o la conquista de territorio (Schiffman, 2012: 2).

En cuanto a los problemas acerca del planteamiento tipológico a las políticas lingüísticas, hay que tener en cuenta que no se aplican solo en el nivel nacional, sino también en zonas administrativas o territoriales más pequeñas dentro de una nación. Sin embargo, cabe añadir que aunque se podrían garantizar las políticas abiertas en el nivel federal, el efecto concomitante de estas políticas puede ser mínimo a nivel local, o puede que los recursos para defender estos derechos sean inadecuados (Schiffman, 2012: 2).

Ferrer (2000) define tres niveles del poder entre los Gobiernos centrales y autonómicos para el contexto español: el primero corresponde a las zonas bajo la competencia exclusiva del Gobierno central, el segundo a las áreas bajo el control exclusivo del Gobierno autonómico, y el tercero incluye las zonas que dependen tanto del Gobierno central como del autonómico. El tercer nivel es el más complejo, ya que los campos de acción y autoridad no se delimitan claramente y, por lo tanto, se pueden interpretar pertenecientes tanto al Gobierno central como al Gobierno autonómico. Asimismo, es en el tercer nivel de autoridad donde el Gobierno central establece las normas generales que las comunidades autónomas deben seguir y concretar (Ferrer, 2000). Sin embargo, Schiffman (2012: 2) ha argumentado que aunque el multilingüismo es, en cierta medida, el producto de determinadas políticas lingüísticas, sería casi inútil intentar establecer tipos de políticas lingüísticas, puesto que los recursos para proteger los derechos establecidos en el nivel central pueden ser inadecuados en el nivel local. Esto es, en los previos intentos tipológicos se ha detectado una confusión entre el multilingüismo de la sociedad y las políticas sobre multilingüismo, suponiendo que lo que es *de iure* también lo es *de facto*, y viceversa (Schiffman, 2012: 2).

2.2 Ideologías y actitudes lingüísticas

La relación entre la lengua y el pensamiento ha recibido mucha atención académica y popular, pero las creencias sobre la lengua por parte de sus hablantes han sido descartadas como objetos de estudio hasta hace relativamente poco. La ideología lingüística alude al carácter situado, parcial e interesado de las concepciones y los usos de la lengua. Estas concepciones representan intentos incompletos de justificar el uso de una lengua y, estas justificaciones, por su parte, suelen ser múltiples, dependientes del contexto y necesariamente construidas a partir de la experiencia sociocultural del hablante (Kroskrity, 2004: 496). La actitud lingüística, a su vez, hace referencia a la evaluación diferencial de las variedades lingüísticas (Carranza, 1982: 64).

2.2.1 Aproximación general a las ideologías y actitudes lingüísticas

La ideología está en interrelación dialéctica con las prácticas sociales, discursivas y lingüísticas, y, consecuentemente, influye significativamente en ellas (Woolard, 1992). Las ideologías lingüísticas se pueden definir como conjuntos compartidos de nociones de sentido común sobre la naturaleza del lenguaje en el mundo (Rumsey, 1990). Mientras que Rumsey (1990) busca posibles enlaces entre la estructura lingüística y la ideología lingüística, Woolard (1992) subraya que la estructura establece categorías formales que conducen especialmente al “reconocimiento erróneo”. Además, Woolard (1992) destaca que las ideologías lingüísticas no son solo un reflejo de las estructuras sociales, sino que las configuran y son configuradas por ellas. Por consiguiente, resulta provechoso analizar su papel de cara al mantenimiento o el cuestionamiento de las dinámicas de poder dentro de las sociedades, a condición de que se comprendan los contextos culturales y políticos en los que se arraigan. La ideología lingüística puede constituir un vínculo tanto entre la estructura de la lengua y la política lingüística como entre la lingüística y las ciencias sociales y, por ende, puede servir para redefinir problemas analíticos, como cuestiones de índole política (Woolard, 1992).

Swiggers (2019) ha definido tres “macro-planos” de ideología lingüística: ideología del lenguaje, ideología de la lengua/una lengua e ideología de la lingüística. El primer macro-plano refiere a la relación entre lenguaje, la identidad del individuo, de la sociedad y cultura, y la percepción de la realidad. El segundo macro-plano, por su parte, trata sobre una lengua o un grupo de lenguas, o sobre las propiedades de una lengua, distinguiendo entre dos niveles: la ideología de características de lenguas, como claridad, seriedad y riqueza, y la ideología de ciertos usos de la lengua, como el habla políticamente correcta. El tercer macro-plano tiene que ver con aspectos de ideología científica acerca de las normas de científicidad, de innovación y de modelización sistemática (Swiggers, 2019). Cabe mencionar que para el presente estudio, parecen relevantes el primer y el segundo macro-plano, sobre todo el primero.

Swiggers (2019) ha presentado la siguiente tipología de siete tipos de ideologías de la lengua:

1. Ideología de la lengua “territorial” o “regional”
2. Ideología lingüístico-cultural europea, de índole religiosa
3. Ideología de la lengua de índole “social”
4. Ideología de la lengua “nativista”, en clave xenófoba y defensiva
5. Ideología de la lengua “nativista”
6. Ideología de la lengua como instrumento sociopolítico

7. Ideología de la lengua político-religiosa, a nivel transnacional.

Para aclararlos más, el primer tipo, la ideología de la lengua “territorial” o “regional”, se basa en la concepción de que la lengua se vincula intrínsecamente con su territorio “original”. El segundo, la ideología lingüístico-cultural europea, defiende la superioridad de una lengua y de la cultura europeas. El tercer tipo de ideología, de índole social, discute sobre el buen uso de una lengua nacional, centrándose en los distintos usos, como el uso de los cuerpos políticos y el uso popular. La ideología de la lengua “nativista”, en clave xenófoba y defensiva está vinculada con el primer tipo, y surge cuando la lengua de un territorio se ve amenazada por una lengua de intrusión. Semejante al anterior, para la ideología “nativista” es característico alabar la lengua materna y desarrollar un discurso acerca de las cualidades de la lengua, por ejemplo, recalcando las características positivas de una lengua y menospreciando otras. La ideología de la lengua como instrumento sociopolítico se relaciona netamente a hechos políticos, y se puede observar en varios diccionarios de coloración ideológica, por ejemplo durante la Revolución francesa. Por último, para caracterizar la ideología de la lengua político-religiosa, a nivel transnacional, trata de la alabanza y defensa de una lengua en un contexto de enfrentamiento político y/o religioso (Swiggers, 2019).

Un aspecto central referente a las ideologías lingüísticas es el concepto de prestigio, ya que las variedades que se perciben como menos prestigiosas suelen ser menospreciadas en favor de las otras. Se ha distinguido entre lenguaje ‘alto’ y lenguaje ‘bajo’, de modo que el alto se corresponde con estatus, alta cultura y aspiraciones fuertes hacia la movilidad social, mientras que el lenguaje bajo se asocia a solidaridad, camaradería e intimidad por parte de sus hablantes. Además, factores esenciales que influyen considerablemente en el nivel de prestigio conferido a una lengua son la estructura social y los sistemas de valores culturales (Carranza, 1982: 63-64).

En lo que atañe a la estructura social, es fundamental cómo los miembros de una sociedad consideran un idioma. Esto es, una variedad lingüística – normalmente la estándar – se asocia más a menudo con un grupo de estatus socioeconómico alto que con clases más bajas (Carranza, 1982: 63-64). Los valores culturales, por su parte, desempeñan un papel fundamental en la perpetuación de las diferencias lingüísticas, especialmente en el caso de las variedades no estándares. En particular, para que una lengua menos prestigiosa se mantenga, debe asociarse a valores que sus hablantes consideran positivos y con los que desean identificarse. Por

ejemplo, la posición del catalán como lengua de prestigio se reforzó gracias a una creciente conciencia étnica, aunque a escala nacional es una lengua minoritaria (Carranza, 1982: 64-66).

Aunque desarrollado dentro de diferentes tradiciones intelectuales, un concepto que se asocia a las ideologías lingüísticas es el de las actitudes lingüísticas. Este vínculo se debe al hecho de que los dos conceptos reflejan y dan forma al manejo de los problemas del mundo real asociados con el lenguaje (Becker, 2025: 309-310). Como se ha mencionado antes, en el estudio de las ideologías lingüísticas es necesario tener en cuenta la relación entre las ideologías lingüísticas y otras ideologías sociopolíticas o culturales. La investigación de las actitudes lingüísticas, a su vez, se centra en la evaluación diferencial de las variedades lingüísticas, y es fundamental para comprender la relación entre el lenguaje y la sociedad (Carranza, 1982: 64; Garrett, 2001: 630).

Un factor fundamental en el estudio de las actitudes lingüísticas es el concepto de la utilidad lingüística, definido por Weinrich (1963) como la necesidad de una lengua para la comunicación social en una determinada comunidad de habla. Es decir, si los miembros de una sociedad consideran que deben conocer demasiadas lenguas, es probable que rechacen las que perciben como menos útiles desde el punto de vista de la comunicación, siempre que no se identifiquen fuertemente con ellas (Bastardas, 2007: 132).

En conjunto, mientras que las ideologías lingüísticas aluden a las creencias, emociones y percepciones sobre el lenguaje compartidas por una sociedad, las que conectan el lenguaje y la sociedad de forma dialéctica, las actitudes lingüísticas se entienden como el procesamiento y la predisposición de los individuos a diferentes situaciones lingüísticas y comunicativas y, como consecuencia, sus actitudes hacia las personas que utilizan tales formas lingüísticas (Becker, 2025: 309-310).

2.2.2 Lengua e identidad

Como señalado en el epígrafe 2.1, el estatus sólido de la lengua propia de una comunidad pequeña puede reforzar tanto la identidad individual como la colectiva. Con respecto a otros factores que influyen en la construcción de una identidad, a juicio de Kroskrity (2004: 509), se pueden utilizar las ideologías lingüísticas de forma productiva en la creación y representación de diversas identidades sociales y culturales, por ejemplo, la nacionalidad y etnia. Las ideologías de los hablantes de una lengua fusionan su experiencia sociocultural y sus recursos

lingüísticos y discursivos, constituyendo esas formas lingüísticas y discursivas como índices vinculados a las características de su experiencia sociocultural. Durante mucho tiempo, la lengua, especialmente una compartida, ha funcionado como la clave para naturalizar las fronteras entre diversos grupos sociales. Por ende, el lenguaje comúnmente se percibe como atributo esencial en los estudios sobre nacionalismo y etnicidad, los que muchas veces apelan al papel de la lengua y las formas discursivas en los procesos de construcción nacional. Los tropos contemporáneos, tales como la invención de tradiciones nacionales y la producción de noticias y ficción popular, y las teorías asociadas a ellos plantean como condición previa la existencia y eficiencia de formas lingüísticas compartidas como base para crear géneros discursivos que, a su vez, crean la nación (Kroskrity, 2004: 509).

En lo que concierne a la producción de discursos similares dentro de las naciones, cabe presentar el concepto de “comunidades imaginadas”, elaborado por Anderson (1983), según el cual los miembros de una nación, por muy pequeña que sea, nunca conocerán a la mayoría de sus conciudadanos, ni los verán u oirán hablar de ellos, pero en la mente de cada miembro existe la imagen de su comunidad. Además de que resulta útil considerar estas comunidades imaginadas como productos de la semiosis, también conviene contemplar el proceso mediante el cual los miembros de estas comunidades aprenden a producir discursos similares por sí mismos. Un factor que se ha sugerido al respecto es la educación pública (Tann, 2010: 149-150).

En lo concerniente al contexto del presente estudio, el sistema político democrático y la estructura territorial autonómica de España se podría calificar de casi federal, lo que ha formado el caso español como paradigma de Estado democrático multilingüe y multinacional. El origen del proceso de federalización se puede situar en un reforzamiento de las identidades colectivas periféricas como consecuencia de las políticas de las instituciones de autogobierno regional. Aunque este reforzamiento es más marcado en las Comunidades Autónomas con lengua propia diferente al castellano y con cierta experiencia de autogobierno reciente o histórica, como Galicia, el País Vasco y Cataluña, el caso valenciano es diferente. Esto es, en comparación con otras comunidades con una lengua propia, los valencianos se sienten más españoles, más orgullosos de ser españoles, y han desarrollado un sentimiento de identidad dual más fuerte que en cualquier otra región (Coller y Castelló, 1999).

El hecho de que la Comunidad Valenciana presenta tendencias más españolistas se ha explicado, por ejemplo, con la existencia de un conflicto acerca de la propia identidad. Este, por su

parte, deriva de la inexistencia de una identidad central, sino dos concebidas antagónicamente, la catalana y la valenciana (Coller y Castelló, 1999). Esa situación ha hecho que la mayoría de los valencianos opte por la identidad dominante: el españolismo. Resulta útil el concepto presentado líneas arriba sobre las comunidades imaginadas en el caso de la identidad colectiva, la que se puede caracterizar como una identidad construida y conflictiva, puesto que cada miembro puede tener una concepción diferente acerca de la identidad de la comunidad. Además, el término “identidad colectiva dual” alude a un doble fenómeno que entre marcos de referencia regional-nacional compatibiliza identidades que se pueden concebir como excluyentes, por ejemplo, sentirse español y valenciano al mismo tiempo, y, de este modo, define el ámbito de pertenencia de un individuo a una comunidad (Coller y Castelló, 1999).

En la investigación de Coller y Castelló (1999) sobre la identidad dual de los valencianos tomaron en consideración, por ejemplo, la ideología política (izquierda, centro-izquierda, centro-derecha y derecha) y la lengua (castellano, valenciano y bilingüe). Aunque las ideologías políticas no resultaron tan significativas, se observó que las cifras más altas de identidad fundamentalmente valenciana y las más bajas de identidad fundamentalmente española correspondían a la izquierda. Asimismo, los que optaron con mayor frecuencia por la identidad dual pertenecían al centro-derecha. En cambio, la lengua planteó un panorama más complejo: los castellanohablantes se decantaron fundamentalmente por la identidad española o por la dual, mientras que las personas con el valenciano como lengua materna optaron mayoritariamente por la identidad dual. Además, la mayoría se identificaba fundamentalmente valenciano en comparación con la identidad española, y los bilingües optaron mayoritariamente por la identidad dual, seguida por la identidad fundamentalmente española (Coller y Castelló, 1999). Esto pone de manifiesto la necesidad del estudio de las relaciones entre lengua e identidad en el caso valenciano.

2.2.3 Ideologías lingüísticas en el contexto valenciano

Históricamente, la imposición de las autoridades españolas y regionales forzaba a los hablantes del valenciano a una forma de diglosia que encarnaba la desigualdad. Un aspecto que diferenciaba los valencianos de los catalanes era que muchos valencianos habían aceptado el menosprecio de su propia lengua, mientras que catalanes no. Para caracterizar estas políticas del afecto, se ha presentado el concepto de “autoodio” (véase Ninyoles, 1971¹): la condición

¹ Ninyoles, R. L. (1971). *Idioma i prejudice*. Palma de Mallorca, Editorial Moll.

específica por la cual los miembros de una comunidad subalterna se devalúan o menosprecian a sí mismos porque adoptan las formas de juicio de la comunidad dominante que ejerce su discriminación sobre la minoría (Pujolar, 2021). Estos prejuicios suelen aparecer asociados a una percepción de imposición en la promoción del uso del valenciano, formando parte de los prejuicios lingüísticos en los que consisten las ideologías diglósicas (Colom i Ortiz, 2003: 20).

Para familiarizarse con un ejemplo concreto de tempranas ideologías lingüísticas en el contexto del presente trabajo, cabe mencionar Antoni M. Alcover y su obra *Diccionari català-valencià-balear*, que intentaba recopilar la lengua viva de los diversos territorios de habla catalana en la primera mitad del siglo XX. Esto demuestra su ideología lingüística, culminada en el cuestionamiento del regionalismo de las naciones catalanoparlantes. Además, su modo de ver la lengua era que, en vez de un hecho cultural, la percibía más bien como un hecho folclórico. Cabe destacar que consideraba el futuro del catalán escrito como un panorama donde conviven diversas hablas locales sin que ninguna prevalezca totalmente: mantenía la idea de que el valenciano, balear, alguerés, catalán de Francia, catalán oriental y catalán occidental deben ser autónomas e independientes en el ámbito literario, sin que ninguna prepondere u oprima a las demás, refiriendo a ciertos “rebeldes” de Barcelona (Perea, 2024).

En los estudios anteriores, se han sugerido cinco tipos ideales de lo que podrían ser las ideologías lingüísticas en el territorio valenciano, situadas en diferentes puntos en el mismo continuum: la asimilacionista españolista, la anticatalanista, la bilingüista españolista, la bilingüista valencianista y la normalizadora (véase Xambó, 1997²). En la aplicación de estas por Colom i Ortiz (2003), la ideología asimilacionista españolista se basa en la idea de que la única lengua correcta o útil en la vida pública es el castellano, mientras que el valenciano se considera como un mero dialecto local inútil en la sociedad moderna. Por ende, es característico oponer al uso del valenciano y negar a que se les hable en esa lengua. Además, esta ideología se puede asociar a la anticatalanista o la bilingüista. El principio de la ideología anticatalanista, a su vez, es negar la unidad del catalán, proponiendo detener la normalización con la excusa de la insuficiencia de la variedad estándar utilizada y afirmar un bilingüismo con la significativa prevalencia del castellano. Colom i Ortiz (2003) no diferencia entre la bilingüista españolista y valencianista, sino que define la ideología bilingüista como pragmáticamente indeterminada

² Xambó, R. (1997). Ideologies lingüístiques i mitjans de comunicació al País Valencià. En: Mollà, T (ed.): *Política i planificació lingüístiques*. Valencia: Bromera.

con una función de enlace, pudiendo aparecer asociada tanto a la ideología normalizadora como a la anticatalanista. Además, trata sobre la idea de que en la Comunidad Valenciana se hablan dos lenguas pero, en caso de conflicto, debe prevalecer el español, mientras que el valenciano se utiliza en ámbitos sociales no problemáticos de una manera simbólica, y se promueve con moderación, sin crear conflictos. La ideología normalizadora, por su parte, cuestiona la bilingüista y la anticatalanista definiéndolas como discursos que favorecen la regresión social de la lengua vernácula. Según esta ideología, en el territorio de “habla catalana” la lengua prioritaria y propia debe ser “el catalán”, defendiendo la extensión del uso de la lengua en todos los ámbitos mediante normas de uso expansivas (Colom i Ortiz, 2003).

Castañer (2018) ha cuestionado si las ideologías propuestas por Xambó (1997) se pueden situar en el mismo continuum, puesto que la posición anticatalanista está ligada a la relación del valenciano con el catalán, mientras que las posiciones españolistas y bilingüistas tienen que ver con la relación con el castellano. Por ende, sugiere que los asuntos lingüísticos más importantes en el contexto del valenciano son la relación entre el castellano y el valenciano, el continuum que se da entre el liberalismo y el intervencionismo lingüísticos, la relación entre el catalán y el valenciano, y la variabilidad lingüística y su relación con la norma. Tienen que ver, por ejemplo, con el sentimiento de identidad nacional, la consideración de los fenómenos lingüísticos como asuntos individuales o sociales, la difusión del secesionismo o del unionismo lingüístico entre ambas variedades (blaverismo-catalanismo), y con el hecho de que el valenciano es una lengua bastante fragmentada y con una norma poco establecida, de forma que pueden surgir conflictos entre las posiciones que favorecen la estandarización y las que defienden el respeto de los usos de la población (Castañer, 2018).

Lledó-Guillem (2023b), por su parte, ha sugerido las siguientes ideologías lingüísticas para el análisis de fenómenos político-ideológicos, también relevantes para la Comunidad Valenciana: nacionalismo cultural, continuidad lingüística y naturalismo sociolingüístico. El nacionalismo cultural es producto del colonialismo, y se puede entender como la ideología que postula una relación orgánica e inseparable entre una comunidad y su lengua autóctona y, consecuentemente, la misma entre una lengua y la identidad nacional que representa. La continuidad lingüística trata sobre la idea de que la lengua y la identidad nacional mantienen una relación constante de continuidad desde los orígenes mismos de la lengua, pudiendo ser aceptada o rechazada según la interpretación de las diferencias lingüísticas y el elemento político. Efectivamente, esta ideología puede ser cuestionada, puesto que se podría estudiar la historia de la lengua desde la perspectiva de la discontinuidad, inspeccionando los momentos de ruptura en

los que la relación entre la lengua y la identidad política que representa se oscurece. Un caso ejemplar sería la cuestión del secesionismo lingüístico valenciano, la que sigue constituyendo un debate ideológico tanto desde el punto de vista académico como desde uno más amplio que contiene todos los ámbitos sociales. Estudiar dicha cuestión desde la perspectiva histórica – creyendo en la continuidad de una lengua – implicaría la aceptación ideológica del secesionismo lingüístico valenciano (Lledó-Guillem, 2023b). Por último, el naturalismo sociolingüístico es la ideología lingüística que comprende la existencia de una relación natural e inseparable entre la lengua y la sociedad (Woolard, 2016: 30-32). Por un lado, puede resultar a que algunas lenguas se consideren auténticas porque funcionan principalmente como marcadores de la identidad de un grupo específico. Por otro, otras lenguas serían anónimas por naturaleza, ya que su valor principal reside en ser instrumentos de comunicación. Sin embargo, como contraposición a esta idea, se ha sugerido que no existen lenguas intrínsecamente auténticas o anónimas, sino que la extensión de algunas lenguas se debe a acontecimientos históricos específicos que podrían no haber ocurrido. Es decir, una lengua puede pasar de ser auténtica a anónima y viceversa dependiendo de las circunstancias políticas e históricas (Lledó-Guillem, 2023b).

2.3 Discursos sobre la lengua

El objetivo del Análisis del Discurso (AD) es identificar cómo se utiliza el lenguaje para construir cierto tipo de representación del mundo en relación con las ideologías, actitudes o relaciones de poder. El término “discurso” tiene varios significados, uno siendo que se considera como lenguaje producido más allá de la oración (Stubbs, 1991: 1). Foucault (1972: 49) lo define como “prácticas que forman sistemáticamente los objetos de los que hablan”. Según Mills (1997: 17), una estructura discursiva se puede detectar debido a la sistematicidad de las ideas, opiniones, conceptos, formas de pensar y comportamientos que se forman dentro de un contexto particular. Burr (2015: 74-75) continúa la definición de Foucault (1972), describiéndolo como “un conjunto de significados, metáforas, representaciones, imágenes, historias, afirmaciones, etc. que de alguna manera producen conjuntamente una versión particular de los acontecimientos”, y añade que “alrededor de cualquier objeto, acontecimiento, persona, etc. puede haber una variedad de discursos, cada uno con una historia diferente que contar sobre el mundo, una forma diferente de representarlo ante el mundo”. Asimismo, los discursos se pueden conceptualizar similares a las ideologías. Igualmente, el uso de los términos “discurso” e “ideología” sugiere que tienen significados semejantes (Baker, 2010: 121). Además, la

ideología se puede percibir como el antecedente materialista cultural del uso posestructuralista del discurso, y el discurso se puede considerar como portador de ideología (Sunderland, 2004: 6).

Con respecto al Análisis Crítico del Discurso (ACD), según Wodak y Meyer (2001: 9): “el adjetivo “crítico” debe interpretarse como distanciamiento de los datos, incorporación de los mismos a su contexto social o adopción de un enfoque autorreflexivo”. Además, Van Dijk (1993) afirma que la finalidad de la investigación del ACD consiste en desentrañar la reproducción de ideologías. Efectivamente, la utilización del ACD puede ser provechosa, ya que ha ofrecido recursos para descubrir supuestos ideológicos ocultos que sustentan los textos sobre temas particulares, y ha vinculado esos aspectos del lenguaje con los contextos sociopolíticos de su producción y uso (Fairclough, 2015: 106-108). Siguiendo los tres elementos del AD – texto, interacción y contexto social – Fairclough (2015: 128) ha definido un modelo tridimensional para el ACD que consiste en (i) la descripción del texto, (ii) la interpretación de la relación entre el texto y la interacción y (iii) la explicación de la relación entre la interacción y el contexto social.

Referente al análisis crítico de los medios de comunicación, según Fairclough (1995: 32-34), los procesos de cambio social y cultural a gran escala se manifiestan en los cambios en el discurso mediático, y cada discurso se debe vincular a un contexto social y cultural más amplio, teniendo en cuenta las relaciones entre el poder y las ideologías. Además, el discurso mediático debe interpretarse tanto como un producto de su entorno sociocultural como un factor que lo forma y transforma, lo que refuerza la tesis básica de la sociolingüística: el lenguaje refleja la sociedad (Nuolijärvi, 2000: 25-27).

Para continuar, a juicio de Tann (2004), la identidad como parte del discurso implica un intento de imponer una sensación de continuidad temporal a ciertos individuos o grupos, y establecer cierta discontinuidad con respecto a otros. Esto supone atribuir criterios particulares, como rasgos o ascendencia, a grupos de individuos con el fin de distinguirlos como individuos o grupos naturales. Por lo tanto, la “identidad” es tanto el proceso como el producto de una formación discursiva que implica un acto discursivo de “identificación” por parte de un actor social. Además, funciona dentro de un discurso situado para mantener una sensación de coherencia en el orden social construido a través del discurso, por ejemplo en los textos que sirven para hacer una crónica de una sociedad (Tann, 2004: 150-151).

Ahora bien, las identidades interpersonales se construyen, al menos en parte, y emergen en el discurso. El uso de un lenguaje evaluativo y emocional revela algo sobre la identidad interpersonal, ya que (i) construye el repertorio simbólico de una comunidad y, por lo tanto, una comunidad imaginada, (ii) realiza la variable del registro relacionada con la comprensión de las relaciones de roles entre los participantes y (iii) fomenta los vínculos y la afiliación (Bednarek, 2010: 229-230). De cara a los vínculos, se pueden mencionar experiencias compartidas en el texto, y algunas prácticas lingüísticas que parecen poner en riesgo los vínculos incluyen referencias a emociones, creencias (religión, política, etc.), repertorios lingüísticos (referencias intertextuales: uso de frases como "como sabes", presumir versus ser un experto: uso de lenguas extranjeras) y compartir experiencias personales, por ejemplo, a través de la narración de historias (Bednarek, 2010: 229-230).

Finalmente, el acercamiento de la identidad en el ámbito lingüístico puede llevarse a cabo a través de la combinación de diversas subdisciplinas de la Lingüística, como la Sociolingüística y el ACD (Bednarek, 2010: 226-227). Según Bednarek (2010: 227-228), la identidad se construye mediante el discurso de tres formas: filogenéticamente, estableciendo un sistema de identidad en la cultura; de forma ontogenética, creando en última instancia una identidad interior coherente y percibida en el individuo; y de forma logogenética, creando una interpretación particular de la identidad en un texto determinado. Además de las subdisciplinas antes mencionadas, cabe mencionar la utilidad de la Lingüística de Corpus, en la que se profundiza más en el capítulo 5, ya que los patrones repetidos a lo largo de los textos contribuyen tanto al desarrollo filogenético como al ontogenético de la identidad, afectando al sistema cultural e individual. Los corpus exhiben tanto lo real (la actuación) como lo típico (lo repetido), por lo que su utilización posibilita la formulación de afirmaciones sobre la probabilidad y la tipicidad. Concluyendo, mediante el análisis de patrones discursivos repetidos, la lingüística puede abordar la naturaleza performativa de la identidad (Bednarek, 2010: 227-228).

3 Antecedentes sociohistóricos del valenciano frente al catalán

Para entender la situación sociolingüística actual del valenciano y el catalán, es imprescindible empezar con la familiarización de los antecedentes sociohistóricos. Comienzo inspeccionando la historia sociolingüística del valenciano y del catalán a través de algunos documentos antiguos y declaraciones políticas. Después, trato sobre la existencia de esencialmente dos hipótesis sobre el origen de la lengua valenciana: la repoblacionista, que expresa un único hecho excepcional histórico, y la autoctonista, que aduce una plural sucesión de causas lingüísticas “banales” (Ballester, 2021: 10).

3.1 Historia sociolingüística del valenciano

De momento, el manuscrito en valenciano más antiguo conocido hasta la fecha data de 1253 (Torlà, 2024). Asimismo, el primer documento de la denominación “llengua valenciana” o “valencià” se encuentra en la traducción de Antoni Canal del *Dictorum Factorumque memorabilium* de Valerius Maximus del año 1395 (Lledó-Guillem, 2023a). Además, un ejemplo temprano de la existencia de diferentes variedades regionales se encuentra en *Regiment de la cosa pública* (1383) de Eiximenis: cuando se describe la región valenciana, se narra que la hermosura de esa tierra se debe al hecho de que su lengua se ha formado de las diferentes lenguas que la rodean, en la manera de tomar lo mejor de cada una y deshacer de las palabras más toscas y vulgares, quedándose con las mejores. De este ejemplo, se puede interpretar que la lengua valenciana se percibía como una lengua separada (Lledó-Guillem, 2023a).

Referente al uso del valenciano, se empezó a utilizar en la sociedad valenciana en 1458, fundamentalmente en contextos administrativos, en la correspondencia pública y privada, y en la literatura (Lledó-Guillem, 2023a). En lo concerniente al uso del término “valencià”, no se negaba la unidad del catalán, sino que en el reino de Valencia se utilizaba dicha denominación para subrayar la identidad jurídica de la zona, aunque hay que reconocer que al mismo tiempo su uso cuestionaba el papel central del Principado de Cataluña. Consecuentemente, la denominación “llengua catalana” se empezó a utilizar sobre la variedad hablada en el Principado de Cataluña (Lledó-Guillem, 2023a).

Además de que el valenciano se percibía como una lengua separada, también existen indicios tempranos de la clara jerarquización de las distintas variedades del catalán en *Los col·loquis de la insigne ciutat de Tortosa* de Despuig (1557), lo que constituye una interpretación ideológica de la diferencia lingüística (Lledó-Guillem, 2023a). Asimismo, aunque ya en esa época

se utilizaban términos como “llengua valenciana” (lengua valenciana), “mallorquí” (lengua mallorquina) y “llengua llemosina” (lengua lemosina), llama la atención la exclusiva utilización del término “llengua catalana” para referirse a la lengua hablada en las zonas catalanoparlantes, es decir, los otros términos no se mencionan en ninguna ocasión. De igual manera, Despuig hasta declara explícitamente que la variedad del Principado de Cataluña es superior a otras porque es donde la lengua encuentra sus raíces y, por eso, la lengua se denomina más apropiadamente “català” o “llengua catalana” (Lledó-Guillem, 2023a).

Al final de la Guerra Civil Catalana (1462-1472), el reino de Valencia se había convertido en el centro de la Corona de Aragón y las zonas catalanoparlantes, habiendo fortalecido su estatus política, cultural y económicamente (Lledó-Guillem, 2023a). Además, los valencianos comenzaron a desempeñar un papel significativo en la Cancillería Real Aragonesa, o *Cancelleria reial*, anteriormente gobernada por las élites urbanas de Barcelona, la que había determinado el uso estándar de la lengua catalana en los documentos oficiales durante ya aproximadamente dos siglos (Lledó-Guillem, 2023a). Por su estatus como el centro más importante de irradiación de la cultura catalana en el siglo XV, Valencia acabó convirtiéndose en el foco principal de castellanización, lo que levantó nuevas situaciones de diglosia, en las que el castellano empezó a adquirir funciones *B* cada día más relevantes y la estabilidad del valenciano como lengua nacional se comenzó a tambalear (Vallverdú, 1985).

En los principios del siglo XVIII, el decreto de Nueva Planta (1707) sojuzgó los fueros de Aragón y Valencia al gobierno y leyes de Castilla (Puchal, 1962). Como consecuencia, el valenciano dejó de ser la lengua administrativa del reino, y, a través de los años, se convirtió en solo la lengua del pueblo (Briz, 2004). Sin embargo, la época del prestigio de la variedad valenciana duró hasta la primera mitad del siglo XIX, período en que se podía utilizar el término “valencià” o “llengua valenciana” para aludir a la lengua común de la nación catalana, lo que cuestionaba la legitimidad del Principado de Cataluña como el origen y ubicación de la prestigiosa variedad de la lengua de la nación catalana (Lledó-Guillem, 2023a).

El siglo XX empezó con una aspiración de cultivar la lengua valenciana, del mismo modo que se había hecho con el provenzal y con el catalán. En la práctica, instituciones y partidos de izquierdas trabajaban para conseguir que el valenciano se estableciera como una lengua presente en el sistema educativo. Se intentaba normalizar el valenciano con la promulgación de las *Normes de Castelló*, pero como habría tenido lugar en 1936, la Guerra Civil (1936-1939) impidió que se llevara a cabo. Durante la dictadura franquista, se prohibió el uso del

valenciano bajo pena de multa. En los años 60, se emprendieron otros esfuerzos, como la obra *Nosaltres els valencians* (1962) de Joan Fuster, para la normalización del valenciano y la profundización de la comprensión en la realidad social valenciana (Calatayud, 2011).

La política lingüística del Gobierno valenciano ha sido notoriamente tímida en su promoción del valenciano, en contraste con las políticas de Cataluña y las Islas Baleares (Pujolar, 2021). Cabe introducir dos movimientos fundamentales relacionados con el valenciano: blaverismo y fusterianismo. El blaverismo se conceptualiza como el movimiento anticatalanista y catalanófobo, caracterizado por un fuerte populismo y conservadurismo con un discurso antimodernizador, regionalista y españolista. Denominado según Joan Fuster, el fusterianismo oponía el blaverismo, siendo de tendencia catalanista, racional, progresista y modernizador, antirregionalista y antiprovincialista (Mateu, 2023).

Después de la dictadura, se promulgó en la Constitución española de 1978 que además del castellano, “las demás lenguas españolas serán también oficiales en las respectivas Comunidades Autónomas de acuerdo con sus Estatutos” (art. 3.2). Asimismo, se expuso que “la riqueza de las distintas modalidades lingüísticas de España es un patrimonio cultural que será objeto de especial respeto y protección” (art. 3.3). El año siguiente, se dictó el Real Decreto 2003/1979, por el que se regula la incorporación de la Lengua Valenciana al sistema de enseñanza del País Valenciano para la situación transitoria hasta la promulgación del Estatuto de la Comunidad Valenciana.

3.2 Historia sociolingüística del catalán

En el siglo XII, el catalán era la lengua hablada por casi toda la comunidad que, además, estaba adquiriendo nuevos ámbitos de uso formal en la región de Cataluña (Vallverdú, 1985). En 1229 empezó la expansión catalana hacia el sur con la reconquista que terminó en Valencia en 1238 (Fradejas Rueda, 2010: 105-106). Además, el intento de normalizar el catalán como una lengua para todos los usos era un elemento importante de los siglos XIV y XV (Vallverdú, 1985).

El período desde el siglo XVI hasta el XIX se denomina la decadencia, ya que se disminuía el uso literario y cancilleresco del catalán en favor de castellano (Fradejas Rueda, 2010: 106). Además, como se ha mencionado antes, en el siglo XV, Valencia reemplazó a Barcelona como el principal centro de irradiación literaria. Como consecuencia del decreto de Nueva Planta (1707), el catalán sufrió un mayor abandono en las condiciones políticas, sociales y

económicas, solo manteniéndose entre la burguesía, mientras que la justicia y todos los organismos oficiales funcionaban obligatoriamente en castellano (Fradejas Rueda, 2010: 106; Vallverdú, 1985). A este respecto, el verdadero peligro para el catalán como lengua literaria fue la sumisión política de los Países Catalanes a la Corona de Castilla. Además, de este período entre los siglos XVI y XVIII brota el conflicto lingüístico contemporáneo que enfrenta a dos sectores de la sociedad: las clases populares – campesinos, trabajadores y artesanos – y sectores de la burguesía catalanoparlantes que no entendían el castellano, y las clases dominantes y los intelectuales que eran bilingües diglósicos (Vallverdú, 1985).

El proceso moderno de normalización lingüística en Cataluña empezó durante el período conocido como la “Renaixença” (1854-1939), que acabó teniendo una gran carga política y nacionalista (Vallverdú, 1985; Fradejas Rueda, 2010: 106). El objetivo de la normalización fue, por una parte, la extensión social de la lengua catalana y, por otra, la estandarización del catalán moderno. Como ejemplos de la aspiración de restaurar y reconstruir la lengua catalana normativizada, cabe destacar la fundación del *Institut d’Estudis Catalans* (IEC) en 1907, así como las publicaciones del lingüista Pompeu Fabra, quien redactó las Normas ortográficas (1913), la Gramática normativa (1918) y el *Diccionari General de la Llengua Catalana* (1932), obras que fueron adoptadas rápidamente por las instituciones catalanes (Vallverdú, 1985; Fradejas Rueda, 2010: 106).

La Guerra Civil (1936-1939) frenó los avances logrados durante la *Renaixença*, ya que se desautorizó el uso del catalán en todos los sectores durante la dictadura franquista. Esta etapa se determina por una política lingüística netamente asimilista, con una nueva situación de diglosia en la que el castellano se designa como lengua *A* (Vallverdú, 1985). Después de la dictadura, la promulgación de la Constitución española de 1978 (véase el epígrafe 3.1) posibilitó que el catalán volviera a adquirir el estatus de lengua oficial en Cataluña (Fradejas Rueda, 2010: 106). Tras la aprobación del Estatuto de Autonomía de Cataluña en 1979, la política lingüística del Gobierno autónomo empezó a difundir el conocimiento y el uso social del catalán, también entre los hispanohablantes (Veny, 1992).

3.3 Hipótesis acerca de la génesis del valenciano

Conforme señalado anteriormente, existen dos hipótesis fundamentales acerca de la génesis del valenciano. La hipótesis repoblacionista se basa en la teoría de que la lengua valenciana tiene sus raíces en el catalán que las familias procedentes de Cataluña llevaron al Reino de Valencia cuando empezaron a repoblar el territorio después de la reconquista del año 1238. La

autoctonista, relacionada con el secesionismo lingüístico valenciano, contrariamente, sugiere que la lengua valenciana ya se hablaba en el territorio actual antes de la reconquista.

Así pues, en lo que concierne a los tiempos antes de la reconquista, se ha confirmado el proceso de la romanización de la región valenciana durante el período de los siglos III y I a. C.. Además, debido a los diferentes determinantes geográficos y la influencia de los substratos de las lenguas prerromanas, probablemente había variación lingüística en el latín vulgar hablado en las diferentes provincias, e incluso en las diversas zonas de cada provincia. A este respecto, se ha afirmado que el latín valenciano poseía algunas particularidades en comparación con el latín hablado en los otros territorios que actualmente se conocen como los “Países Catalanes” (Mourelle de Lema, 1982).

Para continuar, por la incorporación de la región valenciana al poderío islámico en el año 718, el substrato árabe es un elemento importante en el vocabulario del valenciano. Esto es, durante sus cinco siglos bajo el poderío islámico, varios arabismos se transmitieron al léxico valenciano a través de los mozárabes, los cuales tuvieron que adoptar estas palabras durante el período de contacto con los árabes (Mourelle de Lema, 1982). Los documentos más tempranos del habla de la mozarabía levantina se sitúan en el siglo XI, y, es por la lengua mozárabe que se ha podido inspeccionar el grado de la transformación del latín vulgar a principios del siglo VIII. Además, se ha sugerido que el habla de los cristianos que convivían con los árabes formaba el primitivo valenciano, lo que, por añadidura, constituyó el principal determinante del peculiar aspecto del valenciano frente al catalán. También, como resultado de la inspección del desarrollo paralelo del valenciano y el catalán comparando algunos valencianismos a sus catalanismos correspondientes, se ha afirmado que muchos poseen etimologías distintas. Conste pues, la caracterización del léxico valenciano se ha explicado como una consecuencia de la influencia del romance mozárabe hablado en la región valenciana siglos antes de la reconquista. En conjunto, según la hipótesis autoctonista, el valenciano se origina, remotamente, en substratos prerrománicos de carácter particular y, próximamente, en un estilo de hablar el latín diferenciado, además de desarrollarse paralelamente con el catalán, siguiendo una diacronía propia (Mourelle de Lema, 1982).

Ahora bien, actualmente, la propuesta que se considera la única científica en las universidades españolas es la repoblacionista: la teoría general que presenta que las lenguas históricamente habladas en la Comunidad Valenciana se originan directa o indirectamente en la recuperación del territorio bajo el régimen islámico por parte de los cristianos (Ballester, 2021: 10-13). Un

aspecto que se relaciona a esta propuesta es que, como consecuencia de la reconquista, la lengua catalana sufrió alteraciones debidas a la influencia castellana en Valencia y Cerdeña, mientras que en las Islas Baleares no. Es decir, se ha sugerido que la variedad valenciana y su ideología lingüística concomitante, por ejemplo, la percepción del valenciano como el “otro” colonizado, evidencian un proceso de sustitución y colonización castellana (Lledó-Guillem, 2023a).

Existen varias objeciones a la hipótesis repoblacionista (véase Ballester, 2021). Una consideración presentada por Badia (1991) tiene que ver con la diferencia dialectal catalana explicada a través de hechos de la reconquista por los catalanes orientales. Aunque la notable participación barcelonesa en la ciudad de Valencia es un hecho probado en el *Llibre del Repartiment* (1239), en el que se registraron 630 casas de Valencia adjudicadas a catalanes de habla oriental y 388 a catalanes de habla occidental, las proporciones no explican por qué el valenciano pertenece al catalán occidental. Por ende, se ha cuestionado si esta explicación es aceptable en cuanto a su origen, expansión y consecuencias tardías (Badia, 1991). Otra objeción a la hipótesis repoblacionista es “la falta o escasez de análisis lingüísticos y la nula consideración de la existencia de buenos – y más que buenos – indicios de un romance afín al valenciano en época prejaimina” (Ballester, 2021: 21). En relación con la pronunciación distinta y clara del valenciano que se ha intentado explicar con el substrato ibérico de Valencia, se ha recalado que a causa de la temprana romanización antes mencionada, no ha sido posible determinar el substrato ibérico ni su importancia (Mourelle de Lema, 1982).

4 Situación sociolingüística actual

Aunque el País Valenciano es aproximadamente dos tercios del tamaño de Cataluña, tanto en territorio como en población, los catalanes solían considerarlo como una zona atrasada y poco interesante. Históricamente, los valencianos han tenido que basarse en la historia cultural y política que comparten con los catalanes, ya que el Principado de Cataluña y el Reino de Valencia habían sido tradicionalmente gobernados por los mismos líderes, compartían las mismas leyes civiles tradicionales y tenían estructuras institucionales separadas, pero en gran medida homólogas, diferentes de las de Castilla (Pujolar, 2021). De todos modos, aunque las dos comunidades han enfrentado conflictos similares a través de los años, han llegado a diferentes circunstancias sociolingüísticas. Así pues, para comprender el contexto sociolingüístico de esta investigación, presento las situaciones sociolingüísticas actuales de la Comunidad Valenciana y Cataluña. Además, trato sobre algunas diferencias fonológicas y léxicas del valenciano frente al catalán.

4.1 Comunidad Valenciana

Tras constituirse como gobierno legal en 1982, la Generalitat Valenciana adoptó acciones afirmativas en la política lingüística con el objetivo de resolver el conflicto de las dos variedades oficiales, el español y el valenciano (Gimeno-Menéndez y Gómez-Molina, 2007). En 1983, se promulgó la Ley 4/1983 de uso y enseñanza del Valenciano para fomentar el uso del valenciano e incorporarlo en el sistema educativo, salvo las regiones tradicionalmente castellanoparlantes. Además, en 1998 entró en vigor la Ley 7/1998 de Creación de la Academia Valenciana de la Lengua con la función de determinar y elaborar la normativa lingüística del valenciano. Más recientemente, se promulgó la Ley 4/2018 por la que se regula y promueve el plurilingüismo en el sistema educativo valenciano, unos de sus objetivos siendo que el alumnado logre el dominio oral y escrito de las lenguas oficiales y que se garantice la normalización del uso social e institucional del valenciano dentro del sistema educativo.

Actualmente, la Comunidad Valenciana se configura en tres áreas sociolingüísticamente diferentes: bilingüismo activo con predominio del valenciano, bilingüismo activo con predominio del castellano y bilingüismo pasivo con predominio del castellano. Además, se plantea un conflicto lingüístico en el que ocurre una sustitución lingüística del valenciano vernáculo al castellano estándar sin que se modifique la condición social. Este tipo de sustitución lingüística alude a un cambio que no implica una mejora o transformación en el estatus, poder o

condiciones de vida de la comunidad, sino que la nueva lengua se adopta por presiones externas o internas, como políticas lingüísticas, la planificación en lo tocante a la educación, el trabajo, y el prestigio asociado a la lengua dominante, manteniendo la misma posición en la estructura social (Gimeno-Menéndez y Gómez-Molina, 2007). De cara a los conocimientos del valenciano de la población, según los gráficos expuestos en la Imagen 1, en 2021, un 75,8 % lo entendía, un 59,3 % lo sabía leer, un 52,1 % lo sabía hablar y un 42,9 % lo sabía escribir, ya sea bastante bien o perfectamente (Generalitat Valenciana, 2021).

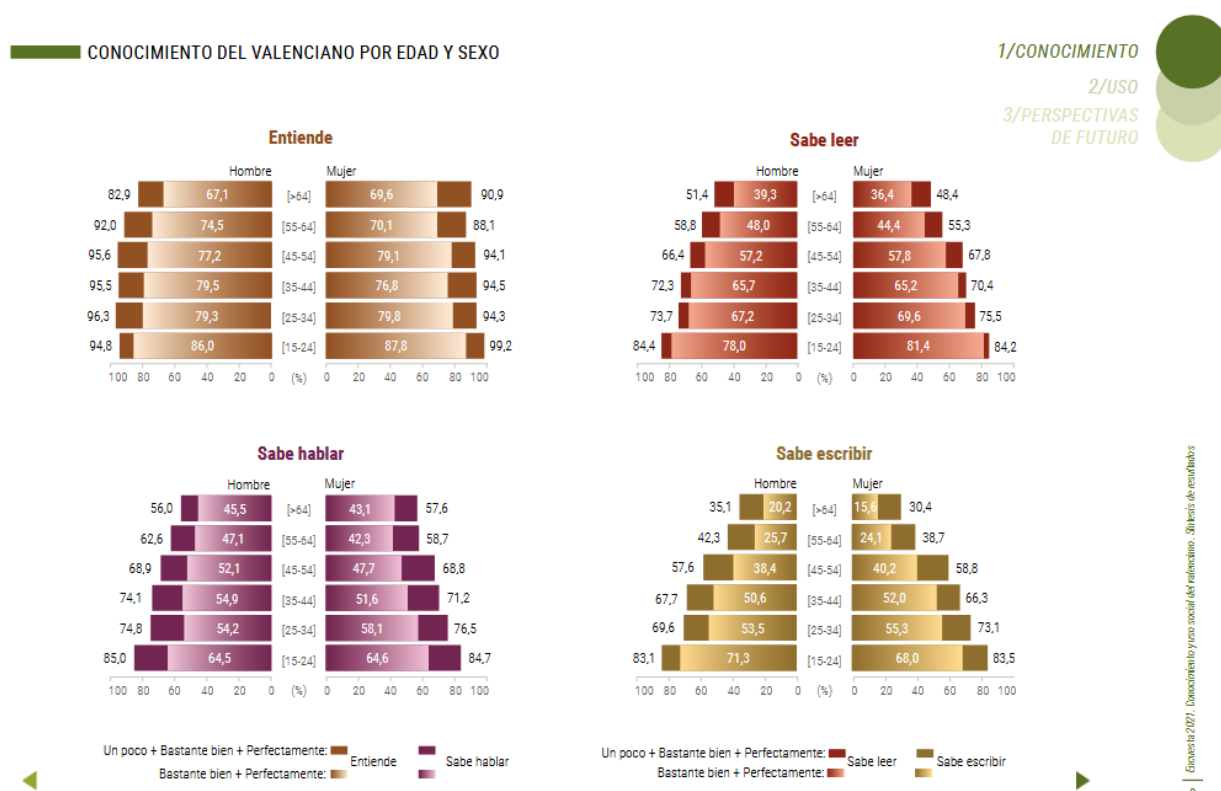


Imagen 1. Conocimientos del valenciano de la población valenciana por edad y sexo (Generalitat Valenciana, 2021).

Existen actualmente dos conflictos en la Comunidad Valenciana, uno lingüístico y otro ideológico (Gimeno-Menéndez y Gómez-Molina, 2007). Según Ferrer (2010), el verdadero origen del conflicto sobre las dos concepciones de la lengua valenciana se encuentra en la situación de tensión política y social durante los años de la transición de la dictadura de Franco a democracia desde 1975 hasta 1982. Este conflicto ideológico está motivado por un sector político blaverista que considera el valenciano y el catalán como dos lenguas diferentes, provocando

el enfrentamiento entre grupos de la sociedad valenciana (Ninyoles, 1992³, citado por Gimeno-Menéndez y Gómez-Molina, 2007). Las encuestas del Centro de Investigaciones Sociológicas (CIS) situadas en los principios del milenio revelaron que alrededor del 60 % de los valencianos encuestados tendía a optar por la posición secesionista, mientras que aproximadamente un 30 % optaba por la unitaria, y un 10 % no quería o sabía responder. En lo concerniente a este conflicto de identidad valenciano, todavía se debate sobre cómo plantear la pregunta sobre la unidad lingüística. En efecto, se ha observado que las instituciones no saben cómo describir este fenómeno lingüístico en su comunicación pública en términos suficientemente eufemísticos para no generar tensiones entre la población general, por lo que se considera como un tabú sociolingüístico en la Comunidad Valenciana (Vásquez-Sanchis, 2025).

De cara al estado actual, según el estudio de Vásquez-Sanchis (2025) sobre la aceptación de la unidad lingüística, el hecho de aceptar afirmaciones como la declaración de la AVL del año 2005 “La lengua propia e histórica de los valencianos *desde el punto de vista filológico, es también la que comparten* las comunidades autónomas de Cataluña y las Islas Baleares y el Principado de Andorra” y “El catalán, el valenciano y el mallorquín son *la misma lengua*”, publicada por la catalana *Plataforma per la Llengua* (2020), correlacionan con variables sociolingüísticas: positivamente con sentirse catalán, ser competente en valenciano y utilizarlo a menudo, y negativamente con sentirse español y situarse en la derecha política. Cabe destacar que en este tipo de encuestas, se ha ponderado sobre la influencia de la formulación de las afirmaciones, ya que un mayor grado de eufemismo puede desorientar y hacerles desconfiar a los informantes. En particular, los hablantes suelen aceptar de forma más extendida la formulación “El valenciano, el catalán y el mallorquín son variedades diferentes de la misma lengua”, puesto que evita las fricciones identitarias, utilizando formulaciones no jerárquicas que reconocen la diversidad de las variedades dentro de la lengua (Vásquez-Sanchis, 2025).

Se ha sugerido que los debates sobre el contraste tanto entre el valenciano y el español como entre el valenciano y el catalán brotaron de la necesidad de justificar la posición desigual del valenciano (Pujolar, 2021). Desde la perspectiva de las ideologías diglósicas, mediante la que las lenguas *A* y *B* se asociaban a “opuestos conceptuales” – el español a los valores de modernidad, racionalidad, progreso y futuro, y el catalán a los de tradición, emoción, estancamiento y pasado – las representaciones del valenciano solían idealizarlo para compensar las

³ Ninyoles, R. L. (1992). Sociología de la lengua. En: García, M. (ed.): *La sociedad valenciana de los 90*, 421-438. Valencia: Alfons el Magnànim.

denigraciones reales expresadas en la vida social. Estas representaciones idealizadoras implicaron que no había necesidad para una movilización colectiva, ya que la esencia del valenciano, supuestamente, nunca se podría considerar amenazada. Como consecuencia, influyó en la falsa conciencia de las clases subalternas al ocultar la verdadera naturaleza del conflicto de intereses de clase, excluyendo cualquier posibilidad o justificación para la movilización colectiva (Pujolar, 2021).

El proceso opuesto al de sustitución lingüística se conceptualiza como la normalización, la que intenta conseguir que la lengua minoritaria fuera “completa” en sus funciones, de modo que toda la comunidad pudiera recuperar de la lengua dominante los usos de los que había sido excluida (Pujolar, 2021). Como se comentaba en el principio de este epígrafe, las medidas que se han tomado en el proceso de la normalización del valenciano tienen que ver, entre otras, con los programas de enseñanza y la creación de un marco jurídico autonómico plasmado en el Estatuto de Autonomía de la Comunidad Valenciana (*Estatut d’Autonomia de la Comunitat Valenciana*) y la Ley de Uso y Enseñanza del Valenciano (*Llei d’Us i Ensenyament del Valencià*), así como en la legislación que se ha derivado de este marco jurídico (Blas Arroyo, 2002). En la educación, la elección de la lengua como medio de instrucción es un aspecto central (Spolsky, 2004: 2-3). Efectivamente, los logros más significativos en el proceso de la normalización del valenciano se han alcanzado en el ámbito de la enseñanza. Algo parecido ocurre con los documentos legales, ya que han sido un factor importante en la recuperación de la lengua valenciana, aunque comúnmente los textos legislativos no son los elementos más decisivos en las políticas sobre la normalización de lenguas (Blas Arroyo, 2002).

4.2 Cataluña

Actualmente, Cataluña se define como una región bilingüe donde coexisten el catalán y el castellano. Cabe mencionar que no hay catalanes monolingües, pero la proporción de monolingües en castellano, así como de bilingües de catalán/castellano ha variado a través de los últimos cien años como consecuencia de la represión franquista, la migración y una política lingüística exitosa a favor de la lengua minoritaria. En general, a diferencia de Valencia, la transmisión intergeneracional de la lengua no se ha interrumpido (Boix-Fuster y Sanz, 2008: 87). Para ilustrar la situación actual de los conocimientos del catalán de la población catalana, de edad de 15 años o más, se puede observar en la Imagen 2 que en 2023, un 93,4 % lo entendía, un 80,4 % lo sabía hablar, un 84,1 % lo sabía leer, y un 65,6 % lo sabía escribir (Generalitat de Catalunya, 2023).

Coneixements lingüístics

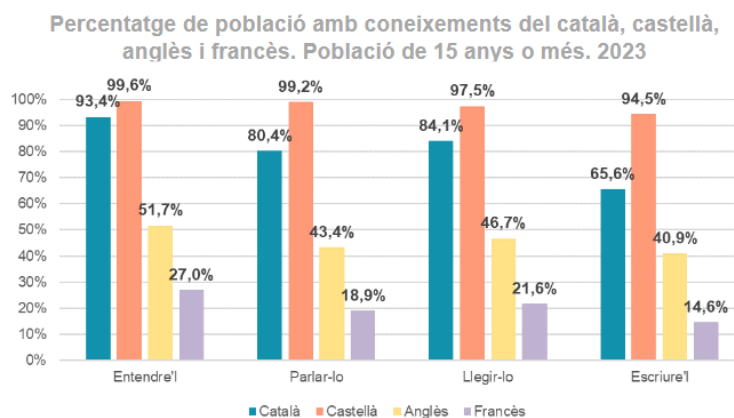


Imagen 2. Porcentajes de los conocimientos lingüísticos de la población catalana (Generalitat de Catalunya, 2023).

La lengua catalana ha sido un factor identitario, pero hoy en día esto está en contradicción con la ley, ya que la presencia de variedades bilingües tanto del castellano como del catalán ha obstaculizado que la lengua siga siendo un criterio determinante en la construcción de identidad. Efectivamente, el hecho de que los catalanes puedan identificarse como bilingües y como catalanes pero, al mismo tiempo, ser percibidos como no-catalanes constituye un conflicto entre la identidad autoconstruida y la identidad percibida (Boix-Fuster y Sanz, 2008: 87). Aunque el vínculo entre la lengua y la identidad étnica sigue fuerte en la sociedad catalana, simultáneamente está en una crisis que se debe, entre otros factores, a las nuevas olas de inmigración. En este sentido, la importancia de la prevalencia y la visibilidad en todos los ámbitos, incluso el político, de la primera y segunda generación de inmigrantes procedentes de otras regiones de España ha respaldado el crecimiento y la extensión de variedades bilingües tanto del español como del catalán. Cabe mencionar que entre los bilingües catalán/español no se encuentra tal cambio de código que se generalmente da en otras comunidades bilingües, sino que, como la elección de la lengua indica la identidad, cada grupo mantiene una clara diferencia entre lenguas y entre grupos. Sin embargo, quizás por los esfuerzos que han hecho los inmigrantes por integrarse, lo que requiere el aprendizaje de la lengua autóctona, ha crecido y se ha extendido una variedad bilingüe del catalán bien establecida en todos los grupos (Boix-Fuster y Sanz, 2008: 102-104). No solo eso, sino también que existe una interlengua catalana,

definida como una variedad no nativa que transfiere del español diferentes rasgos en distintas cantidades en comparación con la variedad bilingüe. Esta distinción entre una variedad bilingüe y una variedad no nativa ha empezado a enmarcar la identidad catalana (Boix-Fuster y Sanz, 2008: 104).

La distribución desigual del catalán y castellano en Cataluña se debe a dos factores fundamentales: por un lado, un proceso de subordinación política y, por otra, el proceso de inmigración destacado (Boix-Fuster y Sanz, 2008: 89). Además de la inmigración nacional, también la internacional ha reestructurado las características sociolingüísticas de la población, por lo que se puede afirmar que el siglo XXI empezó con una clara necesidad de una intervención más fuerte en las habilidades de lenguas. La política lingüística se enfrentó a nuevos retos en varios aspectos, como en la gestión de una sociedad lingüísticamente diversa, el trabajo por la inclusión lingüística de los inmigrantes, el incentivo a las personas para que dominen tanto las lenguas locales como las globales y, al mismo tiempo, el uso del catalán en sus interacciones locales (Pradilla, 2022: 172-178). En este sentido, se declaró en el Estatuto de Autonomía de Cataluña de 2006 que todos los ciudadanos de Cataluña tienen el derecho y el deber de conocer tanto el catalán como el castellano, recalcando que “el catalán es la lengua de uso normal *y preferente* de las Administraciones públicas y de los medios de comunicación públicos de Cataluña, y es también la lengua normalmente utilizada como vehicular y de aprendizaje en la enseñanza” (art. 6.1.).

Durante los años desde 2003 hasta 2010, no se quería relacionar política lingüística a la defensa de una identidad en el discurso oficial, posiblemente porque la identidad de Cataluña se estaba desarrollando y reconstruyendo. En 2010, el Tribunal Constitucional de España contestó en su Sentencia 31/2010 a los artículos del Estatuto de Autonomía de Cataluña de 2006, reinterpretando el concepto de “lengua propia” en el sentido de que el estatus preferencial del catalán sobre el castellano no fue justificado. Esta sentencia afectó a cuestiones lingüísticas y limitó significativamente el contenido del Estatuto de Autonomía de 2006. Fue el punto de partida de un creciente movimiento secesionista que repercutió vigorosamente en la política y las políticas de Cataluña durante la próxima década, cuya transformación más relevante se relacionaba con las ideologías, claramente impactadas por los orígenes y la dinámica del secesionismo (Pradilla, 2022: 176-186).

En lo que atañe a las áreas de uso de las lenguas oficiales de Cataluña, se utiliza el catalán en las entidades gubernamentales locales y regionales, mientras que el castellano se emplea en

los organismos gubernamentales controlados por el Gobierno central de España. Adicionalmente, el catalán se aplica como la lengua preferida en la educación primaria y secundaria, especialmente en la primaria. En el nivel universitario, por su parte, se utilizan las dos lenguas oficiales, lo que requiere que todos sean bilingües, aunque sea de modo pasivo (Boix-Fuster y Sanz, 2008: 89-90).

En la actualidad, hace falta una reflexión colectiva sosegada sobre las políticas lingüísticas con el fin de alcanzar un consenso social y político, la que debería centrarse en la relación entre las lenguas y las identidades de Cataluña, sobre todo si se tiene en cuenta la complejidad que plantea la gestión de las prácticas lingüísticas en una sociedad plurilingüe cuyos miembros atribuyen valores diferentes a las lenguas que hablan (Pradilla, 2022: 205). En cuanto a las opiniones acerca de la unidad lingüística de la familia lingüística catalana, según la encuesta publicada en el InformeCAT (2020), basada en los datos del *Gabinet d'Estudis Socials y Opinió Pública*, el 44,5 % de la población, y el 60,9 % entre los jóvenes, está en contra de la unidad de la lengua (Vásquez-Sanchis, 2025). Esto resultó sorprendente, puesto que la anterior Encuesta de Uso Lingüístico de la Población (2018) había señalado que el 70,6 % estaba totalmente o bastante a favor de la unidad lingüística, especialmente entre los jóvenes, ya que un 73,1 % de ellos se inclinaron hacia su aceptación (Vásquez-Sanchis, 2025). Para volver a la influencia de la formulación de las preguntas en los resultados, cabe mencionar que en la encuesta más reciente se utilizaba una oración interrogativa más directa, preguntando si el valenciano, el mallorquín y el catalán son la misma lengua, mientras que en la otra se preguntaba por el grado de acuerdo con la afirmación de que los dialectos de las Islas Baleares, la Comunidad Valenciana, Andorra y Cataluña son formas diferentes de hablar la misma lengua (Vásquez-Sanchis, 2025).

4.3 Diferencias entre el valenciano y el catalán

Actualmente, dos grandes variedades dialectales dividen el área lingüística catalana: la oriental y la occidental. El catalán oriental comprende el rosellonés, el central, el balear y el alguerés, mientras que el occidental se divide en el noroccidental y el valenciano. Además, algunos de estos se dividen en más subdialectos, como el balear a menorquín, mallorquín e ibicenco (Fradejas Rueda, 2010: 107-110). Cabe destacar que aunque se ha intentado explicar esta división dialectal, por ejemplo, con teorías de la reconquista y con la teoría del substrato prerromano, no se ha logrado una explicación suficiente ni aceptable (Badia, 1991: 5-8). De todos modos, en 2020, los presidentes del IEC y el AVL, así como el rector de la Universidad de las

Islas Baleares, firmaron un protocolo que manifestaba la voluntad de las tres instituciones de cooperar con el fin de establecer un modelo estándar común y unitario, que al mismo tiempo respetara la variación geográfica de la lengua (Pradilla, 2022: 124).

En los debates sobre los contrastes entre el catalán y el valenciano, se ha centrado, por ejemplo, en la pronunciación, describiendo el valenciano como suave, dulce y tranquilo, mientras que el catalán se ha percibido como fuerte, duro y violento (Pujolar, 2021). No obstante, el primer rasgo principal que diferencia el catalán occidental del oriental tiene que ver con las vocales e, ya que en su adquisición han seguido procesos diversos, ninguna de las dos siendo continuación directa de las latinas (Llorach, 1961). Para ilustrar mejor, la e larga latina (\bar{e}) se evolucionó a e cerrada (e) en el catalán occidental, mientras que en el catalán oriental llegó a ser una e abierta (e): *dēu/dēu* (Casanova, 1985). Según Sanchis Guarner (1985: 122), esta discrepancia es imputable al substrato mozárabe ibérica. Otro rasgo esencial es la existencia de dos incoativos en -ISC contra el catalán oriental en -ESC: *partisc/partesc* (Casanova, 1985).

Los demás rasgos, de segundo grado, son que en el valenciano (i) se mantiene los presentes de subjuntivo en *-a, -e, -o*, contra el oriental en *-i*: *tinga/tingui*, (ii) se forma la primera persona del presente de indicativo en *-e* u *-o* contra, por ejemplo, *u* en el catalán central, (iii) no se neutralizan las vocales átonas palatales y velares como en el oriental: *posaria/pusaria*, y, (iv) se conserva la fonema *i* del grupo *iš* a diferencia del catalán oriental: *caiša/caša* (Casanova, 1985). Además, se han destacado diferencias de base de articulación entre las variedades catalanes que se basan en la pronunciación de algunas consonantes en lugares especialmente anteriores, en el valenciano, o posteriores, en la variedad oriental (Recasens i Vives, 2014: 365).

En lo concerniente a las diferencias léxicas, algunas palabras valencianas como *aplegar, bes, bossar, comare, corder, escaló, esplill, follí, gamella, llavar, morella, poal, rabosa, rent, tort* y *xic* comparten la misma etimología con las castellanas *llegar, beso, rebosar, comadrona, cordero, escalón, espejo, hollín, gamella, lavar, molleja, pozal, raposa, reciente, tuerto* y *chico*. Por el contrario, en el catalán prevalecen sinónimos de otro origen, como *arribar, petó, perbo-car, llevadora, anyell, graó, mirall, sutge, obi, rentar, pedrer, galleda, guineu, llevat, borní* y *noi*. Algunas de estas distinciones pueden ser muy antiguas y estar basadas en las diferencias léxicas del latín provincial de la época romana, perpetuadas por el comportamiento de los mozárabes (Sanchis Guarner, 1985: 122-123).

5 Lingüística de Corpus

Los métodos de la Lingüística de Corpus (LC) pueden ayudar a encontrar unidades léxicas y combinaciones de palabras que se repiten a través del material de investigación. La LC se puede definir como una colección de métodos, es decir, como tal no se concibe como un método sino como una insistencia de trabajar solamente con datos de lenguaje real que se hayan extraído del discurso de manera sistemática y compilado en un corpus (Baker, 2010: 19; Teubert, 2005). Cabe añadir que se puede emplear “en todas las ramas o áreas de la lingüística, en todos los niveles de la lengua y desde enfoques teóricos diferentes” (Parodi, 2008: 95).

Al principio, los corpus se utilizaban para la lexicografía y pedagogía, pero durante las últimas décadas han empezado a desempeñar un papel significativo en una gama de propósitos más amplia, aunque su impacto en la sociolingüística ha sido relativamente pequeño. No obstante, metodología basada en la LC resulta útil para los sociolingüistas, ya que las dos disciplinas solapan en términos de epistemología, enfoque y alcance (Baker, 2010: 9-12). Además, debido al hecho de que los corpus constan de lenguaje natural, pueden contener información tanto sobre los valores de las sociedades como sobre la lengua: el lenguaje se utiliza para construir, mantener o cuestionar las actitudes, ideologías, los repertorios interpretativos o discursos (Baker, 2010: 132). A continuación, presento los métodos utilizados en esta investigación: el Estudio del Discurso Asistido por Corpus (EDAC), el modelado de tópicos y el análisis de colocaciones.

5.1 Estudio del Discurso Asistido por Corpus

Una aproximación alternativa al AD tradicional introducido en el epígrafe 2.3 es el intento de hallar patrones lingüísticos en corpus más grandes a través de métodos de análisis tanto automáticos como cualitativos, es decir, el Estudio del Discurso Asistido por Corpus (EDAC). Más específicamente, Baker (2010: 123) presenta la idea de que la investigación de frecuencias, colocaciones, palabras clave y concordancias puede ayudar a hallar pruebas de discursos. Cabe mencionar que existe el riesgo de solo contar lo que es fácil de contar relativo a varios fenómenos sintácticos y patrones discursivos. Sin embargo, según Hardt-Mautner (1995), este tipo de fenómenos contienen reflejos léxicos que pueden ser encontrados al examinar las concordancias, de ahí que puedan dar una dirección hacia los sitios del discurso en los que es posible encontrar una instancia de un patrón argumentativo particular.

Para ilustrar mejor, un programa de concordancias puede contribuir al análisis cualitativo, por ejemplo, facilitando la descripción de propiedades sintácticas y semánticas de unidades léxicas de una manera exhaustiva en vez de selectiva. Además, resulta útil aprovecharlo como un buscador, ya que permite que uno obtenga un control más firme del corpus que no sería posible en caso contrario. No solo eso, sino también que puede levantar cuestiones para un análisis cualitativo, así como producir resultados en sí, como las frecuencias de ciertas formas o la ocurrencia de ciertas colocaciones que pueden resultar relevantes desde un punto de vista crítico (Hardt-Mautner, 1995).

La utilización de la LC se ha demostrado un complemento extremadamente valioso para los estudios basados en textos. Los corpus aportan al descubrimiento de discursos, ya que son repositorios de lenguaje natural tan amplios que son capaces de revelar repeticiones o patrones que pueden ir en contra de la intuición y sugerir trazas discursivas (Baker, 2010: 124). Además, la combinación de los estudios cuantitativos y el *close reading* de ejemplos individuales es lo que hace que este tipo de estudios tengan éxito (Hunston, 2022: 224-225). Hay que añadir que si la tarea de encontrar y cuantificar todas las instancias de cierta palabra, frase o característica en una gran cantidad de texto con mucha exactitud pertenece a un buscador de corpus, la interpretación de los hallazgos es una tarea manual (Hunston, 2022: 187). Hardt-Mautner (1995: 24) sugiere el siguiente procedimiento para integrar el análisis cualitativo tradicional con el componente asistido por computadora: por un lado, el análisis cualitativo de textos individuales revela unidades “cargadas” cuyo comportamiento colocacional puede ser examinado utilizando un corpus más amplio. Por otro lado, la “itinerancia” en el corpus computarizado dirige la atención del analista hacia ciertas unidades o ciertos patrones colocacionales que pueden ser analizados cualitativamente (Hardt-Mautner, 1995: 24).

5.2 Modelado de tópicos

El modelado de tópicos es un método de investigación que se puede aplicar a grandes conjuntos de documentos para descubrir, anotar y revelar los conceptos clave o las características destacadas (Kherwa y Bansal, 2019). Más específicamente, trata de descubrir automáticamente los temas de una colección de documentos, utilizando varios algoritmos que aportan diversas técnicas para encontrar las significaciones ocultas y agrupar las materias como temas, por lo que resulta útil para el análisis de los contenidos en las redes sociales (Hong y Davison, 2010).

Los modelos se pueden clasificar a dos diferentes estrategias: modelos probabilísticos y no probabilísticos (Kherwa y Bansal, 2019). Algunos de los algoritmos más utilizados del modelado de tópicos son el Análisis Semántico Latente (LSA, *Latent Semantic Analysis*) y la Asignación Latente de Dirichlet (LDA, *Latent Dirichlet Allocation*). El LSA es un método no probabilístico, e inicialmente se introdujo como una solución para mejorar la recuperación de información, utilizando técnicas de la reducción de la dimensionalidad (Dumais, 2004). Luego, se empezó a utilizar en el procesamiento de discursos (Evangelopoulos, 2013). El LDA, por su parte, se basa en las probabilidades, captando la intercambiabilidad de tanto palabras como documentos. Según Blei et al. (2003), es un modelo bayesiano jerárquico de tres niveles, en el que cada elemento de una colección se modela como una mezcla finita sobre un conjunto subyacente de temas. Cada tema, por su parte, se modela como una mezcla infinita sobre un conjunto subyacente de probabilidades temáticas (Blei et al., 2003).

Entonces, el LDA supone que las palabras de un documento son intercambiables, y que cada documento se puede representar como una secuencia de palabras individuales (Blei et al., 2003). Por lo tanto, los temas se agrupan a través de un proceso de ignorar el orden de palabras y analizar la frecuencia de la ocurrencia de cada término, por lo que el orden de las palabras en un documento y el orden de los documentos en un corpus se consideran insignificantes a la hora de identificar los temas (Hong y Davison, 2010; Mifrah y Benlahmar, 2020). Por ende, se basa en el modelo de bolsa de palabras (BOW, *bag of words*), que descarta la relación secuencial entre el conjunto de palabras en el que consiste un documento. En el BOW, se puede utilizar la puntuación TF-IDF (*term frequency-inverse document frequency*) para calcular la relevancia de una palabra para un documento en un corpus (Jiang, 2023). Para ilustrar más, en pocas palabras, se multiplica la frecuencia del término (TF) y la frecuencia inversa del documento (IDF). La TF se calcula de la frecuencia de aparición de una palabra en los documentos, y la IDF, por su parte, indica la importancia de un término en un documento concreto, comparándolo a la frecuencia de aparición de ese término en todos los documentos (Qaiser y Ali, 2018).

A la hora de avanzar con el análisis, es importante evaluar la calidad de los modelos generados. Las medidas de evaluación que ofrece Orange para el LDA son la Coherencia (*Topic coherence*) y los valores de Perplejidad (*Log Perplexity*). La Coherencia mide la homogeneidad interna de los tópicos, mientras que la Perplejidad indica cómo el modelo se ajusta a los documentos de entrenamiento. Al realizar el modelado de tópicos, los valores de Perplejidad determinan la mejor manera de ejecutar el proceso. Es decir, conviene avanzar con los ajustes que

producen el valor de Perplejidad más bajo, ya que indica que el modelo es más eficaz con datos que no ha visto antes, y se asocia a menor cantidad de tópicos (Fernández, 2024).

5.3 Análisis de colocaciones

La colocación, identificada por Firth (1957), es una manera de demostrar relaciones relativamente exclusivas o frecuentes entre las palabras. Si se encuentra un par de colocaciones, esas palabras tienden a aparecer cerca o contiguas en el lenguaje natural. La colocación indica una relación, pero para identificar cómo exactamente se manifiesta esa relación en el lenguaje, puede que sea necesario examinar las concordancias con más detalle. De este modo, examinar las colocaciones puede ayudar a revelar patrones frecuentes en la conversación. En lo que concierne a los estudios sociolingüísticos, el análisis de colocaciones puede revelar patrones de colocación que tienen funciones especiales de las que no somos conscientes y, de este modo, contribuir al AD o al análisis de argumentación o de ideologías (Baker, 2010).

Existen varias modalidades de indicadores para calcular la frecuencia de las colocaciones y determinar el grado de asociación de una determinada combinación de palabras (Santana Suárez et al., 2011). El grado de asociación desempeña un papel crucial en la operacionalización de las colocaciones empíricas, junto con la definición formal de coocurrencia y el cálculo adecuado de los datos de frecuencia de coocurrencia (Evert, 2008). Según Bartsch y Evert (2014), examinar la colocación de palabras similares de alta frecuencia de forma comparativa puede facilitar la observación de la actitud discursiva, las ideas y la ideología del hablante. En este trabajo, cabe presentar dos medidas: la que indica la frecuencia relativa y la que mide la información mutua (*MI-score*).

La frecuencia relativa, o la frecuencia de aparición de x con y, calcula el porcentaje de veces que cierta palabra aparece con otra en comparación con el número total de veces que aparece (Santana Suárez et al., 2011). El *MI-score*, por su parte, toma en consideración la exclusividad al medir la fuerza de la asociación entre dos palabras (Baker, 2010). Para ilustrar mejor, calcula la cantidad de información que la coocurrencia contribuye sobre la información de las ocurrencias individuales de las dos palabras, utilizando directamente las frecuencias relativas para la calculación de probabilidades (Santana Suárez et al., 2011). Teniendo en cuenta las posiciones relativas de las dos palabras a través del corpus entero, un *MI-score* alto indica que suelen aparecer cerca la una a otra y raramente separadas. Asimismo, el índice desciende si aparecen juntas con frecuencia pero igual de frecuentemente separadas. Igualmente, los valores más bajos pertenecen a los pares de palabras que tienden a aparecer separados y raramente

juntos. Cualquier par de colocaciones con un *MI-score* superior a 3 se considera estadísticamente significativo al nivel del 5 %, esto es, la probabilidad de que la relación se haya producido por casualidad es del 5 % (Baker, 2010). Como afirmado por Santana Suárez et al. (2011), esta medida no se recomienda para cantidades pequeñas, por lo que se confirmará si resulta relevante para este estudio cuando se analicen las colocaciones.

6 Metodología

A continuación, presento algunas consideraciones legales y éticas en cuanto al método utilizado para recoger el material de investigación, el *web scraping*. Después, presento el corpus recopilado para el presente estudio. Para terminar, explico los métodos de investigación de los discursos.

6.1 Consideraciones legales y éticas

En la actualidad, no existe legislación sobre la utilización del *web scraping*, una técnica para extraer datos de la Web y guardarlos en una base de datos para un análisis subsiguiente (Zhao, 2022), de ahí que muchas veces se perciba como un terreno ambiguo. Aunque técnicamente no es ilegal, por causas relativas al Reglamento General de Protección de Datos (RGPD), los derechos de autor o los términos de servicio de las páginas web, es posible cometer una infracción de la ley (Szwed, 2021). Según Jones (2023), el *web scraping* es legal mientras (i) la legislación local y la página web permitan la extracción del tipo de datos en cuestión, (ii) los datos se utilicen para la investigación académica, (iii) no se recopilen datos personales y (iv) no se utilicen los datos con fines comerciales.

Según el RGPD y el artículo 49 de la ley finlandesa de derechos de autor (Tekijänoikeuslaki 404/1961, 49 §), la protección de datos se aplica solo en los casos de un plagio o una recolección de una parte fundamental del contenido de la base. Debido al hecho de que el corpus de este estudio consiste en alrededor de 2 000 comentarios extraídos de la plataforma, proporcionalmente no se trata de una parte fundamental del contenido de ForoCoche, la página web que contiene cientos de millones de comentarios (Rodríguez, 2016).

En las Condiciones Generales de Uso de ForoCoche (2023) se prohíbe la obtención de sus contenidos “empleando para ello medios o procedimientos distintos de los que, según los casos, se hayan puesto a su disposición a este efecto o se hayan indicado expresamente en las páginas web donde se encuentren los Contenidos o, en general, de los que se empleen habitualmente en Internet por no entrañar un riesgo de daño o inutilización del Website y/o de los Contenidos” (punto 3.2.8.). Aunque no se menciona explícitamente la recolección de datos, se alude a la obtención del contenido utilizando medios inhabituales, lo que puede estar sujeto a interpretación. Dado que el *web scraping* es un método comúnmente empleado en la investigación académica en la LC, lo considero como un medio habitualmente empleado en Internet.

Además, al recopilar datos en cantidades pequeñas como en este estudio, no existe el riesgo de que la extracción de datos sobrecargue el servidor.

Por otro lado, al estudiar los foros de Internet pueden surgir cuestiones éticas sobre el consentimiento informado de los objetos de investigación (Sugiura, 2016). Aun así, como las publicaciones se pueden acceder de forma libre, se puede argumentar que su investigación pública no requiere este tipo de consentimiento (Liu, 1999). En el caso de este estudio, como los comentarios son publicados bajo un nombre de usuario inventado, el anonimato de los informantes está asegurado. Lo que es más, para proteger la privacidad completa de los publicadores, los nombres de usuario se excluyen del corpus final. De este modo, la recolección de datos no pone en peligro o causa molestias a los informantes.

Para resumir, los comentarios que se recopilaron para el corpus de esta investigación se pueden acceder de forma libre sin ningún tipo de registración, el corpus extraído no es una parte fundamental de la plataforma, se puede asegurar el anonimato completo de los informantes y, finalmente, los datos se utilizan solo para la investigación académica. Por dichas razones, se puede afirmar que esta investigación cumple con las normas legales y éticas.

6.2 Corpus

El corpus está compuesto de 2 164 comentarios automáticamente recogidos de nueve hilos de discusión mediante *web scraping*. Más específicamente, el corpus consta de 80 674 tókenes y 9 428 *types*, es decir, formas únicas de cada token. Los hilos de discusión se eligieron manualmente de ForoCoches según su actualidad y contenido con respecto a los discursos sobre las lenguas valenciana y catalana. Los años de creación de los hilos se dividen como sigue: tres hilos del año 2022, dos del año 2018 y uno de los años 2024, 2017, 2016 y 2015.

Puesto que el corpus consta de comentarios anónimos, hay que omitir los estratos sociales de los informantes. Por ende, para obtener una idea de los usuarios generales de la plataforma, cabe observar sus datos demográficos. Analizando el tráfico de ForoCoches, según Similar-Web (2025), un 77.6 % de su público es masculino y un 22.4 % femenino, y los grupos de edad que más tráfico generan son los de 25 a 34 años y de 35 a 44 años, estando compuestos de un 30.56 % y un 22.18 % de la totalidad de los usuarios, respectivamente. No obstante, no se puede hacer la suposición de que el corpus corresponda con estas cifras.

Cabe mencionar que se utilizarán dos versiones del corpus en el análisis, una para el acercamiento cuantitativo y otra para el cualitativo. En lo concerniente a la primera, en el caso de los

comentarios que responden a un texto anterior, consideré necesario excluir los contenidos de los comentarios citados. Con esto quería impedir la posibilidad de que ciertos comentarios aparezcan varias veces en el corpus y alteren los resultados. La otra versión, que incluye los comentarios citados, permite la inspección de la correspondencia entre los usuarios, ya que se pueden ver los comentarios citados. Mejor dicho, en el análisis cuantitativo, en los procesos automáticos cada discurso se trata como una contribución individual aunque sea una respuesta a otra. En cambio, a la hora de analizar los discursos, se recurre a la versión extendida del corpus, buscando los comentarios que surgen del análisis de colocaciones.

6.3 Métodos de investigación

Esta investigación trata de un enfoque de métodos mixtos, formándose de tres partes: el modelado de tópicos, el análisis de colocaciones y el ACD. En efecto, la estructura de la investigación se puede interpretar como una combinación de niveles macro, meso y micro. Los dos primeros pertenecen al *distant reading*, o análisis cuantitativo, mientras que el último se conceptualiza como *close reading*, ya que se realiza para inspeccionar un muestreo más conciso de manera cualitativa.

En primer lugar, extraje los comentarios automáticamente con el programa Octoparse. Después de limpiar los datos utilizando técnicas básicas del lenguaje de programación Python, realicé el modelado de tópicos con el programa Orange (versión 3.39.0) para identificar los conceptos clave o las características destacadas de los comentarios. Ante todo, preprocesé los datos con las herramientas de Orange, esto es, borré los *stopwords*, las palabras frecuentes como pronombres, verbos y preposiciones comunes, y excluí los tókenes de menos de cinco apariciones en los documentos. Después, apliqué el BOW para ignorar el orden de palabras, y probé la puntuación TF-IDF⁴, pero al final avancé sin él. Probé los algoritmos LDA y LSA variando el número de tópicos. Finalmente, elegí el LDA con el número de tópicos establecido a tres. Los factores que determinaron los ajustes finales fueron la coherencia semántica y la interpretabilidad de los temas en cuanto a las preguntas de investigación.

Después, utilicé algunas palabras clave generadas en el modelado de tópicos para buscar sus colocaciones con el programa AntConc (versión 4.3.1). Analicé las colocaciones tomando en consideración cuatro colocaciones a la izquierda (4L) y cuatro a la derecha (4R), con la frecuencia mínima ajustada a dos. Además, examiné los contextos de las colocaciones más

⁴ Con el TF-IDF, el valor de Perplejidad era de 75.99, mientras que sin él, era de 46.71.

detalladamente utilizando la herramienta *Keyword in Context* (KWIC), que expone la palabra clave en su contexto. Esta inspección de las palabras destacadas facilitó el hallazgo de los discursos significativos que extracté para analizarlos en la siguiente fase.

A continuación, me familiaricé con unos cien comentarios extraídos. Al seleccionar el subcorpus, para examinar las diferentes maneras de manifestar ciertas ideologías, presté atención a características como actitudes hacia las lenguas desde la perspectiva de la polémica, argumentos relacionados con la política y la identidad nacional, así como a aspectos históricos, por ejemplo, los orígenes de las lenguas. Para analizar los discursos, utilicé el modelo de tres dimensiones de Fairclough (2015: 58-59) presentado en el epígrafe 2.3. Referente a las ideologías, intenté encontrar conexiones entre las ideologías lingüísticas presentadas en el subepígrafe 2.2.3 al interpretar los discursos y conectarlos a su contexto sociopolítico. Cabe mencionar que los comentarios se expondrán en su forma original, por lo que pueden contener faltas ortotipográficas y gramaticales, así como coloquialismos.

Antes que nada, se visualiza en la Imagen 3 un *Word Cloud*, o nube de palabras, sobre las palabras más utilizadas en el corpus, las que se relacionan con el valenciano, castellano y catalán, así como con conceptos como comunidad, cultura, política, lenguas y aprendizaje.

Tabla 1. Los tópicos predominantes y sus palabras clave

Tópico	Palabras clave
T1	catalán, valenciano, mismo, dialecto, lengua, valencia, español, palabra, sonar, político
T2	valenciano, lengua, castellano, idioma, mismo, valencia, catalán, aprender, inglés, gente
T3	valenciano, español, valencia, españa, sentir, gente, lengua, castellano, cultura, pueblo

Entonces, se pueden observar los tres tópicos destacados en la Tabla 1. Aunque predomina el tema de la distinción y comparación entre el valenciano y el catalán, los temas se podrían dividir de la manera siguiente:

- T1: la comparación entre el catalán y el valenciano
- T2: el valenciano en el sistema educativo
- T3: la identidad nacional de los valencianos.

En la Tabla 1, observando las palabras del T1 como “mismo”, “dialecto”, “palabra” y “sonar”, se puede deducir que el primer tema puede haber formado de conversaciones tanto sobre las similitudes que comparten el catalán y el valenciano como sobre sus diferencias léxicas y fonológicas. Además, el término “político” junto a “catalán”, “valenciano” y “lengua” puede referirse a cuestiones sobre la política lingüística. De todos modos, clasificaría este tema como la comparación entre el valenciano y el catalán.

Según se presenta en la Tabla 1, la mitad de las palabras del T2 corresponde con las del T1. Sin embargo, es posible que las palabras “aprender”, “inglés” y “gente” indiquen que además de tratar sobre la distinción entre el valenciano y el catalán, el segundo tópico contiene conversaciones sobre el sistema educativo y, más específicamente, el aprendizaje de las lenguas, tanto las españolas como las extranjeras. Deduzco que una parte significativa de los comentarios trata sobre el aprendizaje del valenciano y el valenciano en el sistema educativo.

Evaluando la última sección de la Tabla 1, advierto que las palabras clave del T3 también coinciden con los otros temas. Sin embargo, me parecen informativas las palabras “España”,

“sentir”, “gente”, “lengua”, “castellano”, “cultura” y “pueblo”, ya que las interpreto como referencias a la identidad nacional.

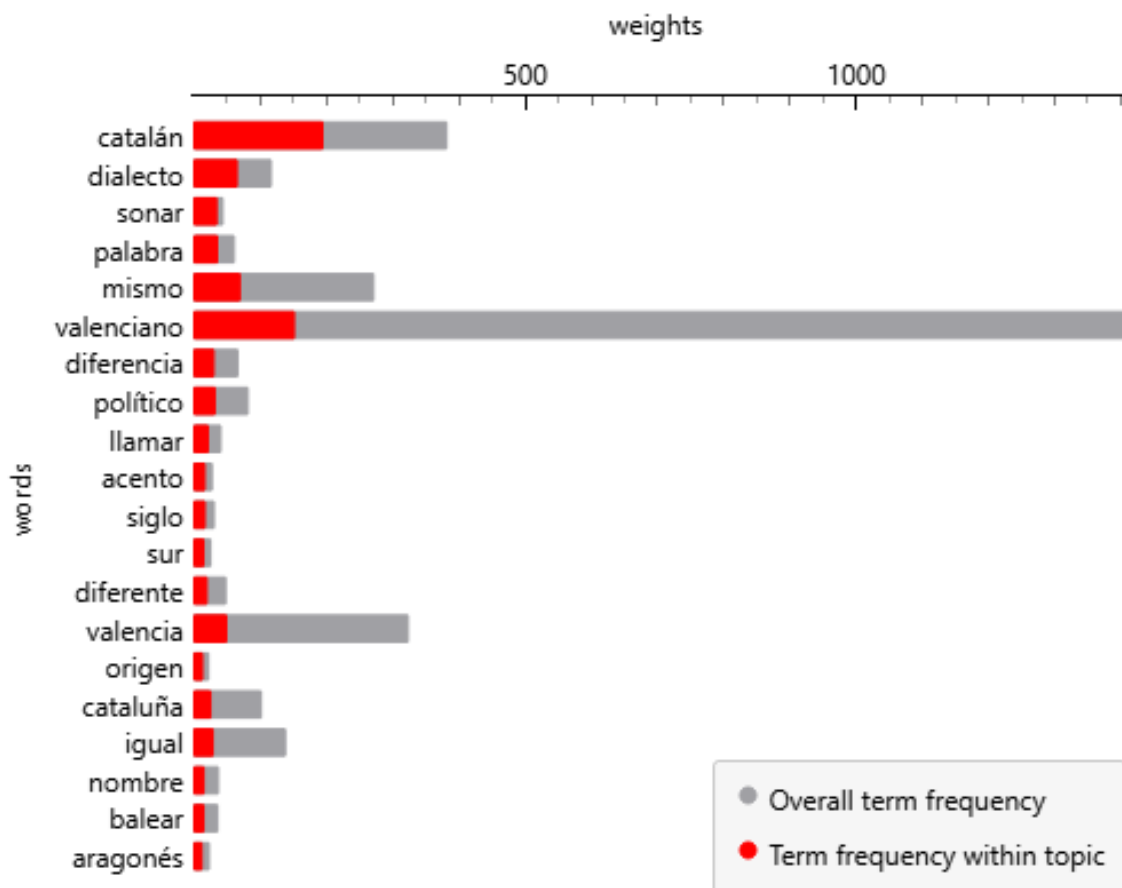


Gráfico 1. T1 visualizado con LDAvis (Orange)

Visualizado con LDAvis en el Gráfico 1, se puede comparar la frecuencia dentro del tópico de cada palabra con su frecuencia en el corpus entero. Efectivamente, el T1 trata un poco más sobre el catalán, mientras que la frecuencia del valenciano es relativamente pequeña en comparación con su frecuencia total. Debido al hecho de que las frecuencias de “catalán” y “valenciano” son casi iguales dentro del tópico y se puede inspeccionar más palabras frecuentes como “acento”, “diferente”, “Cataluña”, “origen” e “igual”, se puede confirmar que el T1 trata sobre la comparación de las lenguas, posiblemente reflexionando sobre sus orígenes, distinciones y áreas geográficas.

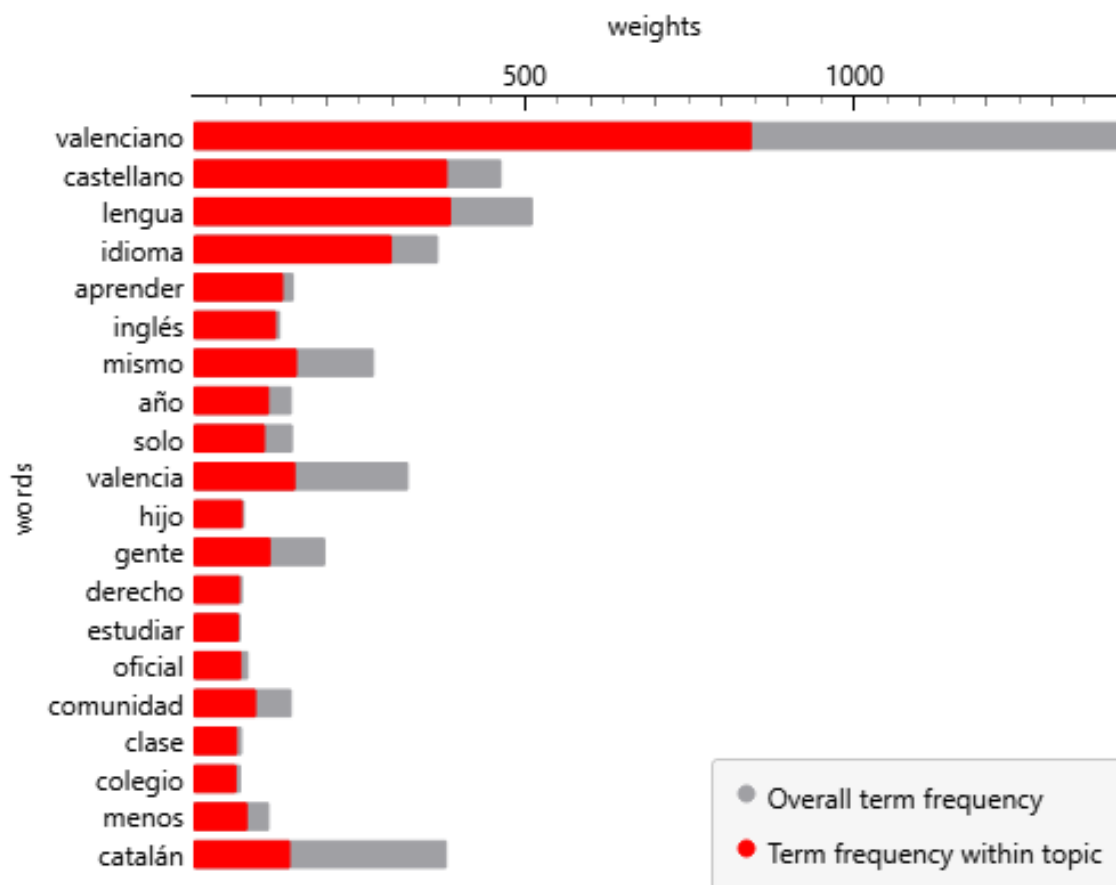


Gráfico 2. T2 visualizado con LDAvis (Orange)

En cuanto al T2, se puede observar en el Gráfico 2 que el nodo “valenciano” aparece frecuentemente en este tópic, esta vez junto a “castellano”, “aprender”, “inglés”, “estudiar”, “clase” y “colegio”, entre otros. También trata sobre el catalán, pero las cuestiones educativas parecen formar el núcleo de este tópic. Inspeccionar el tópic con más detalle, observando las frecuencias de los términos, refuerza la hipótesis de que posiblemente surjan cuestiones políticas sobre el valenciano en el sistema educativo, así como discursos sobre la enseñanza en castellano o valenciano.

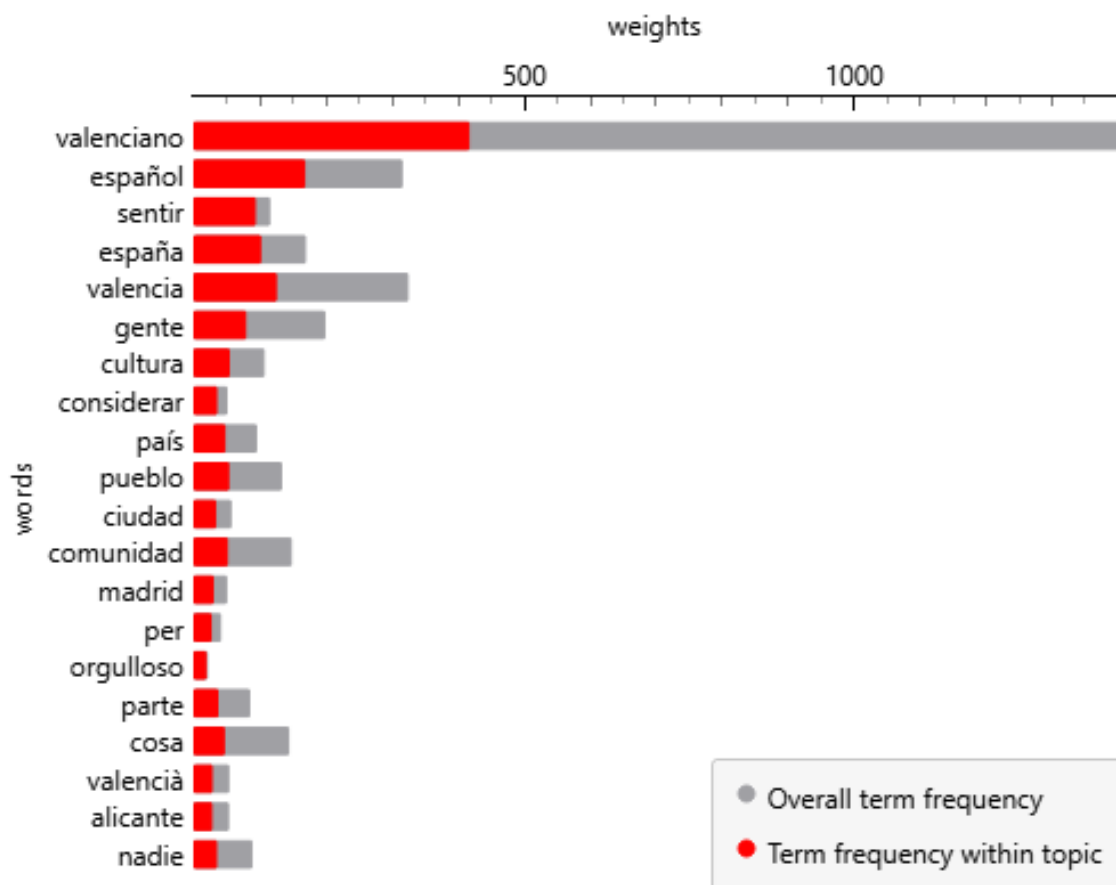


Gráfico 3. T3 visualizado con LDAvis (Orange)

Las frecuencias de las palabras del T3, ilustradas en el Gráfico 3, sugieren que este tópico trate sobre la identidad nacional de los valencianos. Casi todos los nodos, como “valenciano”, “español”, “sentir”, “España”, “Valencia”, “Alicante”, “gente”, “cultura”, “considerar” y “orgullosa”, indican que tiene que ver con los valencianos y su manera de identificarse. Además, como aparecen las palabras “pueblo” y “ciudad”, parece que también surjan cuestiones relativas a las localidades.

Para resumir, con el modelado de tópicos encontré tres temas significativos y útiles para avanzar a la fase siguiente, el análisis de colocaciones. Se puede interpretar que el corpus contiene discursos sobre la distinción entre el valenciano y el catalán, identidad nacional, así como cuestiones políticas, educativas y normativas.

7.2 Análisis de colocaciones

Como fase intermedia, a través del análisis de colocaciones quería comprobar si los comentarios que se encuentran con las palabras clave de los temas surgidos en la fase anterior se

corresponden con mis interpretaciones. Además, la inspección de las palabras clave en sus contextos con la herramienta KWIC facilitó el hallazgo de los discursos para el *close reading*.

Primero, intenté hallar las colocaciones significativas de los nodos “valenciano” y “catalán”, las que se expondrán a continuación. Para inspeccionar el corpus, también busqué algunas palabras clave que surgieron del modelado de tópicos, como “dialecto”, “lengua” y “español” del primer tópico T1. Del T2 destacaron los nodos “político/a”, “castellano”, “idioma” y “aprender”. Por último, además de las palabras solapadas, el T3 consistió en palabras informativas como “España”, “español”, “sentir” y “cultura”. Hay que mencionar que como la mayoría de las búsquedas se ejecutó para inspeccionar las palabras clave en sus contextos, en vez de analizar cada par de colocaciones, una gran parte de los frutos de este paso no se expondrá en esta parte, sino que se analizarán en el epígrafe 7.3, que trata del ACD.

Tabla 2. Las colocaciones de "valenciano"

Collocate	FreqLR	FreqL	FreqR	Likelihood	Effect
catalán	93	49	44	66.531	1.397
habla	64	45	19	59.766	1.634
valenciano	60	30	30	56.104	-1.190
hablo	32	23	9	41.618	1.993
hablan	41	37	4	38.439	1.639
castellano	100	48	52	35.097	0.935
línea	18	14	4	31.957	2.413
asignatura	19	12	7	31.326	2.305
clases	19	16	3	27.761	2.140
daba	11	10	1	25.564	2.860

En la Tabla 2, se pueden observar las primeras diez colocaciones para el nodo “valenciano” en el orden de probabilidad de aparecer juntas (*Likelihood*). Se podía esperar que en este corpus la primera colocación para “valenciano” fuera “catalán”. Hay que notar que la lista contiene tres formas del verbo “hablar”, lo que puede indicar que varias personas mencionen el uso y conocimiento del valenciano. Además, las colocaciones “asignatura” y “clases” pueden hacer referencia a la posibilidad de que algunas cuestiones relacionadas con la enseñanza sean un tema predominante en el corpus.

De cara a la fuerza de asociación entre las colocaciones, se puede comprobar que no hay muchas colocaciones fuertes. Es decir, inspeccionando los *MI-scores* en la última columna (*Effect*), predominan los valores menores de 3, y, como antes mencionado, tendrían que superar

el valor de 3 para considerarse estadísticamente significativos. Esto puede explicarse con el hecho de que esta medida de información mutua no se recomienda para cantidades pequeñas. Otra consideración es que quizás la palabra “valenciano” no puede obtener colocaciones fuertes por ser un término tan prevalente en el corpus. Es decir, aunque aparezca junto a ciertas palabras, aparece tantas veces con otras que no se destaca ningún par de colocaciones significativo.

Tabla 3. Las colocaciones de "catalán"

Collocate	FreqLR	FreqL	FreqR	Likelihood	Effect
valenciano	93	44	49	66.531	1.397
dialecto	22	14	8	61.051	3.268
hablado	10	1	9	27.471	3.249
euskera	9	3	6	26.673	3.419
gallego	8	3	5	25.550	3.597
mallorquín	5	2	3	19.011	4.057
dicen	7	7	0	17.532	3.057
originó	3	3	0	15.583	5.057
deriva	4	1	3	14.297	3.887

Las colocaciones del nodo “catalán”, presentadas en la Tabla 3, por su parte, poseen fuerzas de asociación más aumentadas, ya que todas, menos una, tienen un *MI-score* superior a 3. La que difiere del resto de las colocaciones es la primera según la probabilidad de *Likelihood*, “valenciano”. Puede parecer curioso que la palabra que más probablemente aparece junto a “catalán” al mismo tiempo es la que no se considere estadísticamente significativa. Aquí debo repetir la consideración del corpus consistiendo en tantos discursos sobre el valenciano y el catalán que aparecen igual de frecuentemente juntos que separados, por lo que no se consideran colocaciones fuertes.

Para continuar, la segunda colocación, “dialecto”, especialmente por aparecer con más frecuencia en el lado izquierdo, indicado por “FreqL”, posiblemente alude a discursos que reflexionan sobre si el valenciano es un dialecto del catalán. De las colocaciones “euskera”, “gallego” y “mallorquín”, se puede deducir que en los discursos sobre el catalán se suelen mencionar otras lenguas minoritarias de España, posiblemente comparando diferentes casos y estados sociolingüísticos. También, me parecen informativas las unidades léxicas “originó” y “deriva”, en lo que concierne a las consideraciones sobre los orígenes de las lenguas.

Tabla 4. Las colocaciones de "lengua"

Collocate	FreqLR	FreqL	FreqR	Likelihood	Effect
oficial	31	0	31	105.070	3.731
materna	21	0	21	93.362	4.503
vehicular	16	1	15	74.088	4.634
misma	20	20	0	53.961	3.202
propia	15	1	14	44.502	3.412
castellana	11	0	11	36.799	3.702
franca	6	1	5	30.039	4.898
única	8	4	4	26.046	3.634
cooficial	9	0	9	25.908	3.345
hablada	5	0	5	21.281	4.371
catalana	6	0	6	18.621	3.519
estado	12	1	11	18.580	2.242
minoritaria	4	0	4	18.508	4.634

Como última visualización de los datos de las colocaciones de una palabra particular, según la Tabla 4, el nodo “lengua” adquirió colocaciones fuertes, como “oficial”, “materna”, “vehicular” y “misma”. Además de esas, se puede interpretar que otras colocaciones como “cooficial”, “minoritaria” y “estado” se refieren a cuestiones sobre políticas lingüísticas.

	Hit	Right Context
podría llamar catañol . +1 el	valenciano	y el catalán son el mismo idioma , exactamente igual que el andaluz y el c
estrictamente lingüístico , el	valenciano	y el catalán comparten una gramática , sintaxis y léxico muy similares . los
o así que alguien diga que el	valenciano	y el catalán son lo mismo . tu estas en ambos grupos a mí me parece mu
e que el idioma original es el	valenciano	y que el catalán es un dialecto más ? es curioso este caso . mismo idioma
o que el catalán proviene del	valenciano	y que el catalán es un dialecto del valenciano , pero no se puede decir qu
subnormales que digan que	valenciano	y catalán son idiomas diferentes , se le llama valenciano no por ningún te
es la primera parte suficiente	valenciano	y catalán , nunca español pues claro que me siento español , y orgulloso
a ver . . . el catalán surge del	valenciano	y luego evoluciona de forma distinta . el valenciano es más armonioso er
· se les diga que el catalán es	valenciano	y viceversa es gente que no habla valenciano nunca totalmente deacuerd
si hablas catalán entiendes el	valenciano	y viceversa . del mismo modo que si hablas español “estándar” entiendes
toda la mierda que mueve el	valenciano ,	catalán y gilipolladas habidas y por haber . llamáis cultura a algo que no l
que da pena . dejad de puto	valenciano ,	catalán y mierdas que solo sirven en nuestras putas comunidades autónc

Imagen 4. El contexto del par de colocaciones “valenciano” y “catalán” (KWIC, AntConc).

Así pues, inspeccioné los contextos de varias colocaciones que surgieron de las búsquedas de las palabras clave de los tópicos en la manera presentada en la Imagen 4, en la cual se ve un ejemplo de la búsqueda de “valenciano” y los contextos en los que aparece junto a su colocación “catalán”. De este modo, pude recopilar el subcorpus de los comentarios que, además de reflejar, en su mayor parte, los temas generados en el modelado de tópicos, manifestaban explícitamente ideologías lingüísticas en el contexto valenciano.

7.3 Discursos

En esta parte, se analizó cómo se construyen y cómo actúan los discursos del subcorpus de 22 comentarios. Esto es, se contempló cómo se manifiestan los temas generados en el modelado de tópicos y cómo se reflejan diferentes ideologías en los discursos. No obstante, referente a los factores decisivos en la selección del muestreo, para alcanzar cierta profundidad analítica, al fin y al cabo se centró más en la variabilidad discursiva que en la correspondencia temática. A este respecto, debo advertir que la clasificación de los discursos según su característica principal no se corresponde totalmente con el modelado de tópicos: aunque se menciona el sistema educativo en algunos ejemplos, no llegó a formar un apartado propio. Esto se debe al hecho de que, aunque las ideologías se entran en todos los discursos, el material para el ACD se limitó a aquellos textos en los que la construcción del significado ideológico era más explícita. Para ilustrar mejor, a diferencia de los comentarios en los que las ideologías se expresaban, por ejemplo, a través de oposiciones, clasificaciones y lenguaje emocional, en las expresiones sobre cuestiones educativas estos rasgos eran menos frecuentes. A continuación, expongo el subcorpus y presento los principales resultados del ACD.

7.3.1 Jerarquización de las lenguas

Para empezar, expongo algunos ejemplos que tratan sobre la polémica sobre la denominación del valenciano y la jerarquización de las lenguas. Los ejemplos 1, 2 y 3 contemplan si son la misma lengua o si una es un dialecto de otra, sosteniendo una idea de jerarquización de las lenguas, más específicamente sobre la inferioridad del valenciano frente al catalán:

- (1) Valenciano es un idioma? Yo pensaba que era un catalán mal hablado...
- (2) El valenciano es un DIALECTO del catalán a ver si os entra de una vez en la cabeza.
- (3) el valenciano es un insulto hacia el catalán

El ejemplo 1 menosprecia el estatus del valenciano, definiéndolo solamente como una versión deficiente del catalán, no como una lengua propia. Asimismo, el ejemplo 2 demuestra una reacción fuerte al tema de la discusión, posiblemente porque se trata de un conflicto persistente del que algunos ya se cansan de discutir. Efectivamente, los ejemplos 1 y 2 se pueden concebir como demostraciones del catalanismo y, además, aceptaciones de la unidad lingüística, puesto que opinan que el valenciano es o una versión inferior del catalán o su dialecto. El ejemplo 3 demuestra una opinión aún más despreciadora, llamando metafóricamente el valenciano un insulto hacia el catalán, como si una lengua, una entidad abstracta, pudiera producir un acto ofensivo hacia otra lengua. Se puede suponer que considera el catalán como la forma correcta, y el valenciano como la incorrecta por sus diferencias en comparación con el catalán. En lo tocante a los macro-planos de ideología lingüística de Swiggers (2019), la demostración de este tipo de actitudes acerca de las cualidades de una lengua refleja el macro-plano de ideología de la lengua. Más específicamente, si es que opinan que el territorio original del valenciano es Cataluña, pueden representar ideologías de la lengua “territorial” o “regional”. También, como los ejemplos 1 y 3 menosprecian las características del valenciano, es posible que se vea reflejada una ideología de la lengua “nativista”. Se puede concluir que los tres ejemplos ponen de manifiesto posibles formas de juicio de una comunidad dominante, las que una comunidad subalterna puede adoptar y, empezar a menospreciar a sí mismos.

En el ejemplo 4 se manifiestan actitudes negativas hacia el valenciano y las personas que lo hablan y defienden, apelando a cuestiones sobre las posturas políticas:

(4) Que te pires a un foro de habla valenciana. Ah no, que no hay. Bueno sí, en skyscrapercity hay maricones rojos que hablan en valenciano. Los únicos que defienden esa basura de idioma son rojos maricones, no hay más.

En el ejemplo 4 se exhibe un comentario que manda que otra persona se vaya a un foro en el que se hable en valenciano, llamándolo “esa basura de idioma”, claramente manifestando actitudes negativas hacia la lengua y sus hablantes. Refiriéndose a creencias, en este caso a las creencias políticas, se puede observar prácticas lingüísticas que ponen en riesgo la creación de los vínculos al construir identidades interpersonales (Bednarek, 2010: 229-230). Además, afirmando que no existen ese tipo de foros, destaca la posición del valenciano como lengua minoritaria. Continúa su emisión mencionando uno de los foros más grandes del mundo, SkyscraperCity, en el que hay personas que hablan valenciano, utilizando lenguaje despectivo, hasta lenguaje de odio, al llamarlas “maricones rojos”, aludiendo a la postura política de las

personas que lo hablan y defienden. Como se ha señalado, la defensa de las lenguas regionales, como el valenciano y el catalán, se ha asociado a la política de izquierdas, mientras que el fomento del castellano a los partidos de derecha. Por ejemplo, como indicado en el epígrafe 3.1, eran instituciones y partidos de izquierdas que trabajaban para conseguir que el valenciano se estableciera como una lengua presente en el sistema educativo. Además, en el contexto valenciano, con la expresión “los únicos que defienden esa basura de idioma son rojos maricones, no hay más” hace referencia al enfrentamiento entre los grupos de sociedad valenciana provocado por el sector político blaverista que considera el valenciano y el catalán como dos lenguas diferentes. Por ende, siguiendo la tipología de Swiggers (2019), como se relaciona netamente con hechos políticos, este ejemplo demuestra una referencia a la ideología de la lengua como instrumento sociopolítico. Además, según los tipos ideales de ideologías lingüísticas en el territorio valenciano, por el menosprecio del valenciano, en este ejemplo se reflejan características de la ideología de la asimilacionista españolista, según la cual el valenciano se considera como un mero dialecto local inútil y el castellano como la única lengua correcta (Colom i Ortiz, 2003).

Seguidamente, expongo el ejemplo 5 que, manifestando ideologías anticatalanistas, toma una postura sobre las características y la promoción del catalán:

(5) El catalán es valenciano mal hablado, además de ser una aberración de dialecto del Español mantenido de manera artificial por los nacionalistas y que, cuando dejen de subvencionarlo, morirá como hace muchos años debió haberlo hecho.

Como se puede ver en el ejemplo 5, además de demostrar aversión hacia el catalán, el comentarista recurre a argumentos políticos al referirse a “los nacionalistas” que mantienen su uso. Desde luego, hace alusión a la exitosa política lingüística a favor del catalán. Además, sostiene que sin apoyo monetario se extinguiría, expresando actitudes claramente negativas hacia la lengua al desear que ya se hubiera extinguido. Supuestamente, desde la perspectiva de utilidad lingüística (Weinrich, 1963), lo considera como una lengua inútil. Sin embargo, con la expresión “el catalán es valenciano mal hablado” simultáneamente afirma que considera el valenciano como la forma correcta, por lo que evidentemente reconoce la lengua autóctona de la región valenciana. Aunque se podría interpretar que con esa comparación, de cierta manera, acepte la unidad lingüística, hace una clara distinción al atribuirle criterios particulares al catalán, como rasgos y ascendencia (Tann, 2004), llamándolo “una aberración de dialecto del español”. No hace explícita su opinión del valenciano, pero, por un lado, si posee actitudes

similares hacia todas las lenguas minoritarias, podemos estar en frente de un caso de ideología asimilacionista españolista. Por otro, es evidente que valora el valenciano más que el catalán, pero debido al hecho de que es imposible averiguar a cuánta escala, la interpretación de sus percepciones y valores resulta una tarea compleja.

En el ejemplo 6, se puede observar un discurso anticatalanista que recurre a experiencias propias situadas en el ámbito educativo:

(6) Yo tuve que sufrir en la ESO a una catalufa de profesora en Valenciano que si no decia cada día la suerte que teniamos los Valencianos de hablar la misma lengua que los Catalanes, ya que el Valenciano era el Catalan mal hablado, no disfrutaba la hija de puta. Se me hincharon tanto las pelotas que acabe expulsado 1 semana después de decirle las cosas claras a la tontopollas de la feminazicatalana que teniamos por aquel entonces. (Te hablo del 2010 mas o menos).

En lo tocante al ejemplo 6, se puede percibir las tendencias anticatalanistas que posee la persona que publicó el comentario observando su manera de expresar las emociones al recordar que su antigua profesora les había dicho que el valenciano era el catalán mal hablado y, de este modo, la misma lengua. Esto es, quizás por su identidad nacional, lo describe como sufrimiento. Como consecuencia, se puede sacar la conclusión de que niega la unidad lingüística. A este respecto, cabe recordar que se ha sugerido que la formulación de las afirmaciones sobre la unidad lingüística desempeña un papel notable en su aceptación, es decir, es más probable que los hablantes acepten formulaciones no jerárquicas que evitan las fricciones identitarias y reconocen la diversidad de las variedades dentro de la lengua (véase el epígrafe 4.1). En este sentido, la afirmación “el valenciano es el catalán mal hablado” expresada por la profesora, desde luego, creó estas fricciones identitarias, aportando a la irritación del comentarista. Efectivamente, expresa su frustración mediante el uso de apodos despectivos, como “tontopollas”, “hija de puta”, “feminazicatalana”, el prefijo derivado de “feminista” y “nazi”, y “catalufa”, un nombre despectivo utilizado sobre los catalanes. Esto revela que, además de transmitir actitudes negativas hacia la lengua y las personas catalanas, la opinión divergente de la profesora afectaba a la impresión que este comentarista tenía sobre ella.

Con respecto a la construcción de las identidades interpersonales, además del uso de lenguaje emocional, el hecho de compartir experiencias personales a través de narrar historias es una práctica que pone en riesgo la creación de vínculos (véase el epígrafe 2.3). Además, se puede observar la manera en que construye una identidad diferenciada de los catalanes, estableciendo cierta discontinuidad tanto a un individuo, la antigua profesora, como a un grupo particular, los catalanes. Mediante el uso de lenguaje emocional, despectivo y narrativo, además de

un tono agresivo, impide la creación de vínculos para la construcción de identidades interpersonales, por lo que representa una identidad distinguida de los catalanes y refuerza su identidad valenciana.

El ejemplo 7 trata sobre las cualidades de las dos lenguas, reconociendo que comparten similitudes, pero con el fin de menospreciar el catalán:

(7) Suenan igual que un catalán, pero quitándole la polla de la boca⁵

En el ejemplo 7, se burla, de una manera grosera, de la pronunciación del catalán, manteniendo la opinión de que las dos lenguas suenan semejantes, pero que la pronunciación del valenciano es más clara, pura y sencilla. Volviendo a los contrastes entre el catalán y el valenciano, se ha descrito el valenciano como suave, dulce y tranquilo, y, contrariamente el catalán como fuerte, duro y violento (Pujolar, 2021). Asimismo, algunas consonantes se pronuncian en lugares especialmente posteriores en la variedad oriental, en comparación con la pronunciación en lugares especialmente anteriores en valenciano (Recasens i Vives, 2014). Se puede ver que la asociación de la pronunciación del catalán a aspectos más negativos que la del valenciano se ha mantenido hasta la actualidad. Aportando un discurso acerca de las cualidades de la lengua que menosprecia las características del catalán, se lo puede percibir como una manifestación de una ideología de la lengua “nativista” (Swiggers, 2019). Aunque se trata de una aportación centrada en las cualidades que diferencian las lenguas, al mismo tiempo, si la característica mencionada es el único factor que las separa, respalda cierto unionismo lingüístico, por lo que no se puede afirmar que trate de una postura plenamente anticatalanista. Como en el ejemplo anterior, en este también se obstaculiza la creación de vínculos, esta vez entre los valenciano y castellanoparlantes mediante el uso de repertorios lingüísticos, ya que el comentario original se había publicado en valenciano.

A continuación, se puede observar una interacción en la que se discute sobre la definición del término “dialecto”, así como si el catalán y el valenciano son igualmente legítimos:

(8) ¿Sabes lo que es un dialecto? Son las distintas variedades de un idioma. No se puede decir que el valenciano es un dialecto del catalán ni que el catalán sea un dialecto del valenciano. Simplemente son el mismo idioma y todos los dialectos de un idioma son igual de legítimos. Intentar hacer como que uno es superior al otro es una gilipollez

⁵ Traducción producida por el traductor Salt de la Generalitat Valenciana del comentario original: “Sóna igual que un català, pero llevant-li la polla de la boca”

(9) Menuda gilipollez has soltado. Un dialecto no significa superior o inferior, significa variante. El español mejicano es un dialecto del español europeo porque es una variante del mismo, y ambas son mutuamente inteligibles, salvando jergas locales. Por la misma razón, el catalán puede ser perfectamente un dialecto del valenciano. Y claro que se puede decir que lo es, igual que se puede decir que el español de Méjico es dialecto del español de España. Si no lo fuera, serían idiomas distintos, como lo son el español y el valenciano

En los ejemplos 8 y 9 se puede detectar cierto tipo de confusión sobre la definición de las variedades de un idioma. Una lengua sí es un sistema más amplio que puede incluir diferentes dialectos, o variedades regionales o sociales. Se puede ver que es común la idea de inferioridad en el nivel social y político, aunque no existiera en la realidad lingüística. Quizás es común que se le dé importancia a la “lengua original” de la que una variedad es un “dialecto”, por lo que se puede percibir un dialecto como algo inferior, un mero subsistema del idioma “verdadero”. También, como se ha distinguido entre lenguaje ‘alto’ y ‘bajo’, es posible que distintas variedades, en este caso el valenciano y el catalán, tengan diferentes niveles de prestigio, influidos por la manera de considerar una lengua y las características asociadas a ella, así como con la perpetuación de las diferencias lingüísticas, especialmente en cuanto a las variedades no estándares (Carranza, 1982). Por ende, puede que esta discusión entre los publicadores de los comentarios expuestos en los ejemplos 8 y 9 haya brotado del hecho de que el catalán se considera como una lengua prestigiosa, mientras que el menosprecio del valenciano ha sido tan persistente que incluso los mismos valencianos han adoptado las discriminatorias formas de juicio hacia su propia lengua (Carranza, 1982; Pujolar, 2021). Sin embargo, como en el ejemplo 8 se expresa que “simplemente son el mismo idioma”, es incuestionable que este informante acepta la unidad lingüística. Además, ya que lo continúa con la afirmación de que “son igual de legítimos”, se puede percatar del mismo tipo de ideología que poseía Alcover, la que cuestiona el regionalismo de las naciones catalanoparlantes y mantiene la idea de que ninguna de las diversas hablas debe preponderar u oprimir a las demás (Perea, 2024).

7.3.2 Aspectos históricos

En este subepígrafe, expongo los comentarios que apelan a argumentos sobre la historia de las lenguas. Primero, se examina el ejemplo 10, en el que se recurre a los orígenes de las lenguas, así como a cuestiones sobre la identidad y las políticas lingüísticas:

(10) Ambas lenguas son derivaciones del Romance, por lo que no se puede decir que el valenciano derive del catalán, yo siempre seré firme defensor de que son la misma lengua, pero como a nivel territorial siempre ha habido disputas, me tienen un poco quemado con el tema de que la llamen catalán únicamente. Los catalanes fueron más listos, hicieron un diccionario de la lengua y han

defendido mejor ese signo identitario mientras que en valencia se ha renegado de él de forma sistemática.

En el ejemplo 10, se afirma que los dos idiomas comparten las mismas raíces, en el latín vulgar, lo que impide que uno pueda derivar del otro. Como indicado en el epígrafe 3.3, se ha teorizado, por un lado, que el valenciano se ha desarrollado paralelamente con el catalán, siguiendo una diacronía propia, y, por otro, que tiene sus raíces en el catalán que fue trasladado a la región valenciana por los repobladores catalanes (Mourelle de Lema, 1982; Ballester, 2021). Así pues, aunque opina que son la misma lengua, expresa que se debería dar más reconocimiento a la variedad valenciana, y no solo llamarla catalán. Esto se puede explicar con la aspiración de subrayar la identidad jurídica de la región valenciana (Lledó-Guillem, 2023a). Afirmando que siempre ha habido disputas a nivel territorial, se refiere a los debates en los que se discute sobre la unidad lingüística, estrechamente relacionada con la ideología de la lengua “territorial” (Swiggers, 2019). Además, ya en el siglo XVI existía el discurso sobre la superioridad de la variedad del Principado de Cataluña, recurriendo a teorías sobre los orígenes de la lengua y argumentando que, por sus raíces, se denomina más apropiadamente “català” o “llengua catalana” (Lledó-Guillem, 2023a). En lo que concierne a la normalización del catalán, menciona el *Diccionari General de la Llengua Catalana* y otros esfuerzos de defender el catalán como signo identitario, posiblemente refiriéndose a la política y planificación lingüísticas, expresando que “los catalanes fueron más listos” que los valencianos. En este sentido, se refiere a la timidez del Gobierno valenciano en la promoción del valenciano. Esto, por su parte, tiene que ver con la relación entre el castellano y el valenciano, influida por el sentimiento de identidad nacional. Además, puesto que se alude al hecho de que el valenciano no posee una norma bien establecida, cabe destacar que este comentario también indica hacia la consideración de Castañer (2018) sobre la variabilidad lingüística y su relación con la norma, que puede acabar en conflictos entre las posiciones que favorecen la estandarización y las que defienden el respeto de los usos de la población (véase el subepígrafe 2.2.3).

Para continuar, en los ejemplos 11 y 12 se comparan las lenguas apelando a la historia:

(11) El valenciano es más antiguo, pero el poder político de los catalanes ha hecho que hoy en día se imponga la enseñanza del catalán de Gerona. Hay muchas palabras diferentes. En muchas comarcas de Valencia se utilizaban muchas palabras del castellano, pero han ido adoctrinando a los niños para que usen las catalanas. Y cambia bastante la pronunciación, en muchos sitios el valenciano se parece más al castellano que al catalán.

(12) Son muy similares. El Catalán deriva del valenciano (el valenciano es más antiguo) y tiene también cosas del francés. Ahora, esto es meterse en una guerra.

Similarmente al ejemplo 10, también en el ejemplo 11 se menciona el poder político de los catalanes. Más específicamente, se afirma que se enseña el catalán procedente de Gerona, y que en varias comarcas valencianas se utilizan muchas palabras del castellano, aunque “han ido adoctrinando a los niños para que usen las catalanas”. Con esto puede referirse a los esfuerzos para la normalización del valenciano, por ejemplo en el ámbito de la enseñanza, donde se han alcanzado los logros más significativos al respecto (véase el epígrafe 4.1). Aquí debo repetir que en Cataluña, por motivos identitarios, no se encuentra tal cambio de código que en otras comunidades bilingües (Boix-Fuster y Sanz, 2008). Esto resulta relevante, puesto que en este comentario se contempla las diferencias léxicas entre el catalán y el valenciano, afirmando que en Valencia se utilizan muchas palabras del castellano. Efectivamente, varias palabras valencianas comparten la misma etimología con las castellanas, como *aplegar* con “llegar”, *bes* con “beso” y *bossar* con “rebosar” (véase el epígrafe 4.3). Igualmente, menciona las diferencias fonológicas, expresando que la pronunciación del valenciano comparte más similitudes con el castellano que con el catalán. Como señalado en el epígrafe 4.3, la pronunciación es uno de los rasgos principales en la que se han centrado los debates sobre los contrastes entre el catalán y el valenciano.

También comparando el valenciano con el catalán, en el ejemplo 12 se mencionan las similitudes de las lenguas, caracterizándolas como “muy similares”. Además, del mismo modo que en el ejemplo 11, también en el 12 se discute que el catalán se origine del valenciano y no viceversa, aunque ninguno de estos discursos contiene más explicaciones sobre esa afirmación. Tampoco expresan explícitamente sus posturas en la polémica sobre la denominación del valenciano, aunque por el hecho de designar el valenciano como la lengua más antigua, se puede suponer que ambos lo reconocen como una lengua separada. Recurriendo a una metáfora que compara esta polémica con una guerra, el autor del ejemplo 12 es consciente de que con su contribución adopta una postura en una conversación controvertida. En esta afirmación, se pueden ver características metacomunicativas cuando, consciente de la magnitud de la polémica, reflexiona sobre el impacto del comentario que estaba próximamente publicando.

A continuación, expongo el ejemplo 13 que compara la relación del catalán y el valenciano a la del español de España y el español de Argentina o la variedad de Andalucía:

(13) Decir que el catalán viene de valenciano, es como decir que en español viene del argentino o el andaluz; una puta subnormalidad.

Se puede deducir que el publicador del ejemplo 13 opina que el catalán es la lengua que fue trasladada a la región valenciana del mismo modo de que los colonizadores españoles llevaron el castellano a Argentina. Utiliza la expresión vulgar “una puta subnormalidad” para intensificar su opinión al describir cómo le parece la idea de que el valenciano es el idioma del que desciende el catalán. Efectivamente, en este discurso se ven reflejadas tanto ideologías de la lengua “territorial” como, de cierta manera, ideologías lingüístico-culturales europeas, ya que también enfatiza la relación entre el argentino y el español (Swiggers, 2019).

Antes de avanzar a los ejemplos sobre la identidad nacional, cabe observar unos comentarios más neutrales en los que se combinan características históricas y políticas:

(14) En la Comunidad Valenciana, muchas personas consideran el valenciano como una lengua distinta debido a su identidad regional y su historia específica. Esta perspectiva tiene un componente cultural y político significativo.

(15) Es muy duro de verdad que se intente manipular y hacer daño por política, no pasa nada por reconocer la historia, y eso no quita derecho, los aragoneses no son castellanos por hablar "castellano" ni los valencianos son catalanes por hablar "catalán", etcétera.

En el ejemplo 14, se apela a la identidad regional y la historia específica de la Comunidad Valenciana, reconociendo que la lingüística no es el componente central en este debate, sino que se trata de lo cultural y político. Para ilustrar mejor, se refiere al secesionismo valenciano, blaverismo, que ha motivado el conflicto ideológico presente en la sociedad valenciana. Asimismo, en cuanto a la historia específica, puede comprender aspectos como el funcionamiento de Valencia como el centro de las zonas catalanoparlantes durante la época del prestigio de la variedad valenciana, así como la utilización del término “valencià” o “llengua valenciana” sobre la lengua común de la nación catalana, y el papel significativo de los valencianos en la *Cancelleria reial* que determinaba el uso estándar de la lengua catalana (véase el epígrafe 3.1). Además, la alusión a la historia específica también puede estar relacionada con algunos antecedentes de la lengua sucedidos antes de la reconquista del año 1238. A este respecto, cabe mencionar la peculiaridad del léxico valenciano que se ha explicado por la influencia del romance mozárabe hablado en la región siglos antes de la reconquista. Consecuentemente, se ha considerado que el habla de los cristianos que convivían con los árabes formaba el primitivo valenciano (véase el epígrafe 3.3).

En el ejemplo 15, se menciona que no es incorrecto reconocer la historia, pero que no define la identidad de los miembros de una comunidad, haciendo alusión a la ideología de la lengua “territorial”, según la que la lengua se vincula intrínsecamente con su territorio “original”. Lo

ilustra con ejemplos sobre la denominación de las lenguas regionales, como que hablar “catalán” no hace que los valencianos fueran catalanes, cuestionando la idea del naturalismo sociolingüístico, según el cual existe una relación natural e inseparable entre, en el caso valenciano, el valenciano y la identidad valenciana. Asimismo, también cuestiona la ideología de la continuidad lingüística, opinando que no es necesario que la lengua y la identidad nacional mantengan una relación constante de continuidad desde los orígenes mismos de la lengua (Lledó-Guillem, 2023b).

7.3.3 Identidad nacional

Finalmente, expongo algunos comentarios que tratan particularmente sobre la identidad nacional. El ejemplo 16 es el comentario iniciador de un foro titulado “No considero valencianos a quienes no hablan valenciano”. Así pues, en los ejemplos 16, 17, 18 y 19 se pueden observar fragmentos de una interacción que contiene opiniones sobre la identidad nacional con referencias políticas:

(16) Por mucho que vivas en un sitio y tengas tu empadronamiento si no te adaptas a su tierra, su gente y sus costumbres no mereces que te reconozcan como uno de ellos.

(17) Los valencianos son dignos de estudio...sobre todo aquellos fascistas que se autodenominan valencianistas...acusando de catalanistas a los verdaderos valencianistas...

(18) El problema es que los catalanistas como tú meten las narices en esta comunidad Autónoma, creyéndose que formamos parte de su territorio inventado, y queriendo lobotomizarnos a todos, eso sí en nombre de la libertad. [...] Tenéis muuuucha faena si queréis colonizar la CV. Corre vete, es hora de rezarle a tus políticos, y esta tarde a atarte en un árbol vestido de amarillo, sé buen Catalán. Hablas ce Fascismo, un Indepe hablando de Fascismo!!!!

(19) yo os hacía un muro más alto que trump en méxico y os quemaba aprovechando las fallas todos los libros que teneis en polaco...y a mamarla campeón

El ejemplo 16 toma una postura sobre la identidad nacional, sosteniendo que para ser reconocido por una comunidad, hay que adaptarse a su tierra, su gente y sus costumbres aludiendo, en este contexto, al aprendizaje de las lenguas regionales. Dado que esta opinión se liga al caso valenciano, se puede observar la idea típica sobre la construcción de la entidad estatal de una nación pequeña, que trata sobre una propia lengua nacional como símbolo central de la identidad nacional, la que, por su parte, funciona como un componente de la aspiración ideológica para adquirir una voluntad unificada entre la población (véase el epígrafe 2.1). Según la clasificación de Swiggers (2019), tiene que ver con el macro-plano de ideología del lenguaje,

ya que refiere a la relación entre lenguaje, la identidad del individuo, de la sociedad y cultura. Enfatiza el papel de la lengua como una base central del nacionalismo cultural, según el cual existe una relación orgánica e inseparable entre una comunidad y su lengua autóctona. Esta afirmación también corresponde con el eje relativo a la relación entre el castellano y el valenciano, puesto que se destaca el sentimiento de identidad nacional (Castañer, 2018). Además, en este caso, se considera la lengua valenciana como la lengua auténtica y prestigiosa de la región, exigiendo que los verdaderos valencianos dominen su lengua autóctona. En ese sentido, esta toma de postura también se puede relacionar con la ideología del naturalismo sociolingüístico, el concepto que sugiere que algunas lenguas se consideren auténticas porque funcionan principalmente como marcadores de la identidad de un grupo específico (Lledó-Guillem, 2023b). Aunque he mencionado dos ideologías que comprenden la relación inseparable entre la lengua y la comunidad, hay que recordar que tienen fines diferentes: el nacionalismo fundamentalmente comprende acciones para crear la nación, mientras que el naturalismo se refiere a la idea de una evolución natural sin intervenciones, tales como la planificación, política y normalización lingüísticas. Sin embargo, es posible que en el ejemplo 16 se considere que el nacionalismo cultural se apoya en el naturalismo sociolingüístico, es decir, la lengua y la sociedad valenciana tienen una relación natural, la lengua funcionando como marcador de la identidad, lo que construye una relación orgánica entre la comunidad y su lengua, merecedora de protección. Por último, en este ejemplo se puede observar de una manera más explícita cómo el escritor posee cierta concepción sobre la identidad colectiva de la comunidad, o la “comunidad imaginada” (Anderson, 1983), puesto que afirma que las personas que no hablan la lengua autóctona de la región no van a ser reconocidas por la comunidad, como si supiera las concepciones de cada uno de sus miembros.

El ejemplo 17, respondiendo al comentario anterior, habla sobre los valencianistas, quizás refiriéndose a la ideología blaverista, que defiende el uso de la denominación “lengua valenciana” y una identidad valenciana diferenciada de la catalana, implicando que el iniciador de este hilo es un “fascista que se autodenomina valencianista”. Cabe observar que esta aportación gira la discusión hacia la relación entre el catalán y el valenciano, entrando en el tema de blaverismo vs. catalanismo, es decir, la difusión del secesionismo o del unionismo lingüístico entre ambas variedades (Castañer, 2018). El escritor transmite la impresión de haber sido acusado de catalanista, pero afirma que es un “verdadero valencianista”, enfatizando su identidad nacional y postura política. El ejemplo 18 responde a esto, expresando en tono agresivo que el escritor del ejemplo 17 es un catalanista independentista que quiere colonizar la Comunidad

Valenciana, haciendo alusión a los acontecimientos históricos, más específicamente a la reconquista del siglo XIII. Consecuentemente, se destacan características de la ideología de la lengua como instrumento sociopolítico (Swiggers, 2019). Evidentemente, con la emisión “creyéndose que formamos parte de su territorio inventado” acentúa la impresión de que la Comunidad Valenciana es una región independiente totalmente diferenciada de Cataluña, manifestando ideologías anticatalanistas (Colom i Ortiz, 2003). Asimismo, refuerza ciertos estereotipos de los catalanes, expresando que las características de un “buen catalán” se relacionan con el independentismo. En el ejemplo 19, se puede observar una referencia a los actuales temas de conversación de política internacional, planteando una situación hipotética en la que se construyera un muro “más alto que Trump en México” en la frontera entre Cataluña y la Comunidad Valenciana. Se puede deducir que es otro ciudadano valenciano demostrando ideologías anticatalanistas, ya que amenaza que quemara todos los libros en “polaco”, un término peyorativo sobre una persona catalanoparlante utilizado en este contexto para referirse a la lengua catalana, aprovechando los fuegos de las Fallas de Valencia.

Los argumentos políticos vinculados a la identidad nacional han sido una característica persistente en el corpus. Por ejemplo, se puede percatar de la manera de que algunos valencianos se diferencian de los catalanes afirmando que, contrariamente a los catalanes, son españoles y no quieren separarse de España:

(20) Soy valenciano y por eso, soy español. Lo que hay en la región de arriba y en menor grado, en la nuestra, son traidores a la patria.⁶

En el ejemplo 20 el comentarista manifiesta su identidad nacional y, al mismo tiempo, su actitud hacia los catalanes, haciendo alusión a las cuestiones políticas de Cataluña. El comentario original se había escrito en valenciano, por lo que se puede contemplar si la decisión de elegir el valenciano para expresar que se siente español en sí era un medio para acentuar su mensaje. Además, como ya se ha mencionado, el hecho de recurrir al uso de lenguas extranjeras puede ser una práctica lingüística para impedir la creación de vínculos necesaria para la construcción de identidades interpersonales (Bednarek, 2010). En este caso, se obstaculiza la creación de vínculos entre las personas que no hablan valenciano, pero, de cierta manera, se apela o a los catalanes o a las personas con las que forma su comunidad, posiblemente fortaleciendo los

⁶ Traducción producida por el traductor Salt de la Generalitat Valenciana del comentario original: “Sóc valencià i per això, sóc espanyol. Lo que hi ha en la regió de dalt i en menor grau, en la nostra, són traïdors a la patria.”

vínculos entre ellos. De todos modos, hace hincapié en su afirmación de que ser valenciano significa ser español, quizás por no querer dar la impresión de que los valencianos posean una identidad nacional similar a la que se asocia estereotípicamente a los catalanes. Esto es, establece cierta discontinuidad a los catalanes, atribuyéndoles criterios particulares, la identidad nacional, como un factor separador entre las dos poblaciones: los valencianos se identifican como españoles, mientras que los catalanes no. Este ejemplo pone de manifiesto cómo, en comparación con otras comunidades con una lengua propia, tales como la catalana, los valencianos se sienten más españoles, más orgullosos de ser españoles, y han construido un sentimiento de identidad dual más fuerte que en cualquier otra región (véase el subepígrafe 2.2.2). Esto se ha explicado con el conflicto acerca de la propia identidad, derivado de la inexistencia de una identidad central, sino dos concebidas antagónicamente, la catalana y la valenciana (Coller y Castelló, 1999). Seguidamente, expresa que los catalanes, e incluso algunos valencianos, son “traidores a la Patria”, posiblemente porque han reforzado la posición del catalán frente a la del castellano, fundamentando la identidad catalana en la lengua autóctona de su región. Asimismo, en lo tocante a su referencia a la Comunidad Valenciana, puede aludir a las personas que no se identifican como españolas, sino que optan por una identidad fundamentalmente valenciana en vez de una dual o española. Cabe destacar que no menciona explícitamente la lengua en la construcción de la identidad nacional, por lo que se dificulta la interpretación de una ideología lingüística específica para el caso del valenciano. De hecho, este ejemplo pertenecería más bien al primer macro-plano de Swiggers (2019), ideología del lenguaje. Por añadidura, la ausencia de una referencia a la lengua también se puede interpretar como un rechazo del nacionalismo cultural, sugiriendo que la identidad valenciana no requiere un elemento lingüístico. Como se comentaba líneas arriba, es posible que haya elegido el valenciano para acentuar su mensaje y para apelar a los catalanes, así como a los valencianos que adoptan una postura separatista y, por consiguiente, se encajan en su definición de “traidores a la Patria”.

Para terminar con más observaciones sobre la identidad nacional, en los ejemplos 21 y 22 se pueden detectar elementos de ideologías diglósicas, más específicamente en forma de autoodio:

(21) yo soy valenciano, con familia de pueblo cerrado valencianoparlante y yo me cago en el valenciano, si fuese por mi lo borraría de la tierra.

(22) Soy mas valenciano que tu y no hablo valenciano aunque lo he estudiado desde pequeño porque me parece una puta perdida de tiempo, es preferible que abandonen este dialecto en las escuelas y le dediquen mas tiempo a idiomas que te vas a servir mucho mas que el valenciano

El escritor del ejemplo 21 pone énfasis en el hecho de que es valenciano, además con “familia de pueblo cerrado valencianoparlante”, y aun así, desea que no existiera la lengua valenciana. De este modo, subraya su identidad nacional antes de expresar el autoodio hacia el propio idioma de su región (véase el subepígrafe 2.2.3). En efecto, se puede sacar la conclusión de que, al afirmar todo esto, descarta la idea de que una lengua nacional funcione como símbolo central de la identidad nacional. En el ejemplo 22, se puede observar un caso similar en relación con el autoodio, ya que acentúa su identidad nacional, y afirma que la lengua valenciana no le parece un factor identitario, sino un mero dialecto inútil. Como se ha señalado anteriormente, parece corroborar la teoría de utilidad lingüística (Weinrich, 1963): si uno no se identifica fuertemente con una lengua, es probable que la rechace si la percibe como menos útil en comparación con las otras que debe conocer desde el punto de vista de la comunicación. Además, afirmando que ha estudiado el valenciano desde pequeño, menciona el sistema educativo, ámbito en el que se han alcanzado los logros más significativos en el proceso de la normalización del valenciano (véase el epígrafe 4.1). Desde luego, es justo la expresión vulgar “una puta pérdida de tiempo” de la que se puede deducir que le parece una habilidad inútil, y que, en vez del valenciano, se debería estudiar idiomas más útiles. Por consiguiente, el ejemplo 22 parece consolidar el hecho de que estos prejuicios lingüísticos en los que consisten las ideologías diglósicas suelen asociarse a una percepción de imposición en la promoción del uso del valenciano (véase el subepígrafe 2.2.3). No solo eso, sino también que en ambos ejemplos se reflejan ideologías de asimilacionista españolista (Colom i Ortiz, 2003), ya que menospreciar el valenciano, considerarlo como un dialecto inútil y estar en contra de su incorporación en el sistema educativo es un elemento común en estos discursos.

8 Conclusiones

El propósito de esta investigación fue analizar discursos e ideologías lingüísticas sobre el valenciano y el catalán en los hilos de discusión publicados en ForoCoches. Más específicamente, intenté crear una imagen actual de la polémica sobre la posición de la lengua valenciana a través de utilizar métodos tanto cuantitativos, el modelado de tópicos y el análisis de colocaciones, como cualitativos, el Análisis Crítico del Discurso (ACD). Mis hipótesis fueron que la mayoría de los comentarios se centraría en la comparación de las lenguas, sus orígenes y la política, y que los discursos serían de índole argumentativa, con referencias a las diferencias léxicas y fonológicas de las lenguas, al igual que a algunos aspectos políticos y acontecimientos históricos. Además, estimé que se podrían observar diferentes ideologías lingüísticas en los discursos inspeccionando la expresión de opiniones así como la manera de referirse a cuestiones relacionadas con la situación sociolingüística.

A través del modelado de tópicos, se encontraron tres temas predominantes en el corpus: (T1) la comparación entre el catalán y el valenciano, (T2) el valenciano en el sistema educativo y (T3) la identidad nacional de los valencianos. Con respecto al T2, conviene observar que no pude estimar que el sistema educativo formaría uno de los temas. Sin embargo, no lo considero como un hallazgo imprevisto, puesto que las políticas lingüísticas comprenden las decisiones sobre la enseñanza. Así pues, como deducción del modelado de tópicos, interpreté que el corpus contiene temas como la distinción entre el valenciano y el catalán, algunas cuestiones políticas, cuestiones educativas y normativas, y la identidad nacional de los valencianos.

Como fase intermedia, a través del análisis de colocaciones se comprobaron los temas que surgieron en la fase anterior y, además, se eligieron discursos que reflejaban dichos temas y manifestaban ideologías lingüísticas para el subcorpus utilizado en el ACD. Valga como ilustración que con la apariencia de las colocaciones “asignatura” y “clases” del nodo “valenciano”, se pudo confirmar que el valenciano en el sistema educativo es un tema predominante en el corpus. Además, la colocación “dialecto” del nodo “catalán”, frecuente en el lado izquierdo, evidenció discursos que tratan sobre la denominación del valenciano, ya que alude a la expresión “dialecto del catalán”. Asimismo, las colocaciones “originó” y “deriva” del mismo nodo parecieron confirmar que el corpus contiene argumentos sobre la historia de las lenguas. Al inspeccionar el nodo “lengua”, destacaron las colocaciones fuertes “oficial”, “materna”, “vehicular” y “minoritaria”, ya que tienen que ver con las políticas lingüísticas.

En el *close reading* del subcorpus de 22 comentarios, presté atención a características como la jerarquización de las lenguas, argumentos que apelan a los orígenes y hechos históricos, así como a la identidad nacional. Cabe mencionar que aunque hice esta división según la característica que consideré más predominante en cada discurso, también había bastantes solapamientos. Es decir, algunos discursos clasificados bajo el título de, por ejemplo, jerarquización de las lenguas, también pudieron contener consideraciones sobre la identidad nacional. Sin embargo, para adquirir una visión general, clasifiqué que unos nueve comentarios hablaron expresamente sobre cierta jerarquización de las lenguas. La identidad nacional, por su parte, fue la característica principal de unos siete comentarios. Por último, en seis comentarios se centraban explícitamente en aspectos históricos, refiriéndose a la historia sociolingüística y los orígenes de las lenguas.

Se pudo detectar una conexión entre los discursos y los aspectos presentados en el marco teórico sobre, por ejemplo, la historia y la actualidad sociolingüística de las lenguas, así como los conceptos de políticas, ideologías y actitudes lingüísticas. Esto es, se percató de que ciertas ideologías lingüísticas afectaban a las actitudes lingüísticas de los hablantes, como la asimilacionista españolista en los discursos que menospreciaban el valenciano, o la ideología de la lengua “nativista” en los que trataban sobre las cualidades de una lengua (Colom i Ortiz, 2003; Swiggers, 2019). Los discursos sobre la jerarquización de las lenguas afirmaron, por un lado, que el valenciano es un catalán mal hablado, su dialecto y un insulto hacia él, y por otro, que el catalán es valenciano mal hablado y una aberración de dialecto del español. Asimismo, se expresó que el catalán suena más inarmónico y áspero que el valenciano, y que las personas que piensan que el valenciano y el catalán son la misma lengua son unas “feminazicatalanas”. Para dar otro ejemplo, algunos argumentos sobre los orígenes de las lenguas y la sociolingüística mencionaron que el catalán y el valenciano comparten las mismas raíces en el latín vulgar, así como el diccionario que promovió la normalización del catalán en comparación con la tímida promoción del valenciano por parte del Gobierno valenciano (Vallverdú, 1985; Fradejas Rueda, 2010; Pujolar, 2021). Cabe destacar algunos discursos en los que se abordó la identidad nacional mencionando, por ejemplo, (i) la identidad regional y la historia específica del valenciano, (ii) la opinión de que hablar catalán no define a los valencianos como catalanes, (iii) la suposición de que los catalanes quieran colonizar la Comunidad Valenciana y (iv) la amenaza de quemar los libros en catalán en las Fallas de Valencia. Además, un comentarista expresó que como es valenciano, consecuentemente también es español, a diferencia de los “traidores a la Patria”, es decir, los catalanes. Del mismo modo, pero además poseyendo

características de ideologías diglósicas en forma de autoodio (Pujolar, 2021), uno afirmó que es valenciano pero borraría la lengua valenciana de la tierra, y otro que es más valenciano que otro comentarista, pero que no habla valenciano a pesar de haberlo estudiado desde pequeño porque lo considera una lengua inútil que se debería quitar del currículo y reemplazar con lenguas más útiles. En estos ejemplos, también se pudieron detectar elementos de la ideología de asimilacionista españolista (Colom i Ortiz, 2003).

A través del ACD, pude contemplar cómo los ejemplos del subcorpus reflejaban los conceptos fundamentales para este estudio, como políticas lingüísticas, teorías sobre los orígenes de las lenguas, utilidad lingüística, orgullo y autenticidad. Además, una gran parte de los discursos poseían características argumentativas, pero en ocasiones, las posturas se manifestaban sin apelar a argumentos sólidos, a veces utilizando lenguaje despectivo, y, hasta lenguaje de odio. Se pudieron percatar de algunas ideologías lingüísticas asociadas al contexto valenciano, como de asimilacionista españolista y anticatalanista, así como ideologías de la lengua como instrumento sociopolítico, “nativista” y “territorial” (Castañer, 2018; Swiggers, 2019). Otras consideraciones lingüísticas presentes en el corpus fueron la variabilidad lingüística y su relación con la norma, la relación entre el castellano y el valenciano, la identidad nacional, el autoodio y la unidad lingüística (Castañer, 2018; Pujolar, 2021).

Así pues, se puede concluir que la manera de percibir una lengua como signo identitario influye significativamente en las ideologías y actitudes lingüísticas, las que, por su parte, dan forma a la cultura lingüística en la que se fundamentan la planificación y política lingüísticas. Además, se puede comprobar que este conflicto todavía despierta emociones y provoca reacciones fuertes, quizá debido al hecho de haberse arraigado en la sociedad durante, por lo menos, décadas, si no siglos. A pesar de que se inspeccionaron diversos fenómenos y argumentos, de los que surgieron tres temas predominantes, se pudo detectar la influencia de la identidad nacional en la mayoría de los comentarios, por lo que parece ser el meollo de la polémica que rodea la lengua valenciana. Igualmente, además de la índole sociopolítica, la incertidumbre y la imposibilidad de confirmar la génesis del valenciano hace el desenredo de las ambigüedades de esta polémica una tarea aún más difícil.

Algo que se hubiera podido mejorar en el presente estudio es que, aunque el corpus consistía mayoritariamente en comentarios en castellano, había algunos en valenciano o catalán. Es decir, debido al hecho de que en los análisis automáticos el corpus se procesó solamente en español, es posible que se haya omitido contenido que habría incluido contribuciones

beneficiosas para esta investigación. Asimismo, he reflexionado sobre si los análisis cuantitativos habrían producido resultados diferentes si en la versión del corpus utilizada se hubiera reconocido las interacciones entre los comentarios en vez de tratar cada comentario como aportación individual.

Para proponer algunas posibles líneas de investigación para el futuro, planteo la aplicación de los métodos utilizados en el presente estudio a discursos políticos que tratan del caso del valenciano y el catalán, o incluso de otros casos que comprenden cuestiones sobre el papel de la identidad nacional en la planificación lingüística. Como otra propuesta, se podría llevar a cabo un estudio comparativo sobre el caso del valenciano y el catalán y otro conflicto lingüístico, por ejemplo, a través de un análisis de palabras clave (*keyword analysis*), para destacar las características particulares de cada situación sociolingüística.

Concluyendo, para adquirir un punto de vista cercano a las personas afectadas por la polémica persistente sobre el valenciano y el catalán, he sacado a relucir algunas de sus opiniones expresadas anónimamente en las redes sociales. Según el modelado de tópicos, los temas predominantes en las discusiones se centraban en las diferencias de las lenguas, así como en algunos aspectos políticos, educativos e identitarios. En el análisis cualitativo, se pudo confirmar que la polémica sobre la denominación del valenciano es un conflicto ideológicamente cargado, en el que la identidad nacional desempeña un papel significativo. Resolver este conflicto no será una tarea fácil, y como conclusión del presente estudio, se puede afirmar que requeriría un análisis profundo de las actitudes e ideologías poseídas por los ciudadanos, ya que influyen significativamente en la planificación y política lingüísticas.

Bibliografía

- Anderson, B. (1983). *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. Verso.
- Badia Margarit, A. (1991). *La formació de la llengua catalana*, (Vol. 55). L'Abadia de Montserrat.
- Baker, P. (2010). *Sociolinguistics and Corpus Linguistics* (1.^a ed.). Edinburgh University Press. <https://doi.org/10.3366/j.ctv2f4vj51>
- Ballester, X. (2021). *Orígenes de la lengua valenciana: La hipótesis repoblacionista*. Prensa Universitaria Zaragoza.
- Bartsch, S., y Evert, S. (2014). Towards a Firthian Notion of Collocation. *Vernetzungsstrategien, Zugriffsstrukturen und automatisch ermittelte Angaben in Internetwörterbüchern*, 2(1), 48-61.
- Bastardas i Boada, A. (2007). *Language and identity policies in the 'glocal' age*. Generalitat de Catalunya. Institut d'Estudis Autònoms.
- Bednarek, M. (2010). Corpus Linguistics and Systemic Functional Linguistics: Interpersonal Meaning, Identity and Bonding in Popular Culture. *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation* (1.^a ed.), Continuum. 215-244.
- Becker, L. (2025). 16 Language Ideologies and Language Attitudes. *Manual of Applied Romance Linguistics*, 33, 309.
- Blas Arroyo, J. L. (2002). The Languages of the Valencian Educational System: The Results of Two Decades of Language Policy. *International Journal of Bilingual Education and Bilingualism* 5.6, 318-338.
- Blei, D. M., Ng, A. Y., y Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3, 993-1022.
- Boix-Fuster, E., y Sanz, C. (2008). Language and identity in Catalonia. *Bilingualism and Identity*, 87-106.
- Briz, A. (2004). El castellano en la Comunidad Valenciana. *Revista Internacional de Lingüística Iberoamericana*, 2(2 (4)), 119-129.

- Burr, V. (2015). *Social Constructionism* (3.^a ed.). Routledge.
- Calatayud, V. A. (2011). Análisis de la realidad sociolingüística del valenciano. *Papers. Revista de Sociología* 96.2, 501-514.
- Carranza, M. A. (1982). Attitudinal research on Hispanic language varieties. *Attitudes Towards Language Variation: Social and Applied Contexts*, 63-83. London: Arnold.
- Casanova, E. (1985). El valenciano dentro del diasistema lingüístico catalán. *Revista de Filología Románica*, vol. III-1985. Editorial de la Universidad Complutense de Madrid.
- Castañer, J. C. (2018). Las ideologías lingüísticas en Valencia y su área metropolitana. *Estudios interlingüísticos*, (6), 42-62.
- Coller, X., y Castelló, R. (1999). Las bases sociales de la identidad dual: el caso valenciano. *Reis*, 155-183.
- Colom i Ortiz, F. (2003). Repertori d'actituds i normes d'ús lingüístic entre els estudiants de la ciutat de València. *Llengua, societat i ensenyament*. Institut Interuniversitari de Filologia Valenciana.
- Condiciones Generales de Uso: FOROCOCHES.COM. (2023, 10 de agosto). Consultado el 20 de octubre de 2025 de Link World Network S.L., <https://www.lwnetwork.com/lwnet/legal.php?ident=forocoches.com#normas>
- Constitución Española. (1978). *Boletín Oficial del Estado*. núm. 311. pp. 29313-29424. [https://www.boe.es/eli/es/c/1978/12/27/\(1\)](https://www.boe.es/eli/es/c/1978/12/27/(1)), consultado el 3 de febrero de 2026.
- Corretger, M., y Salvador, V. (2019). *The Rise of Catalan Identity: Social Commitment and Political Engagement in the Twentieth Century*. (1.^a ed.). Cham: Springer International Publishing. Web.
- Davidson, J. (2020). Asymmetry and Directionality in Catalan-Spanish Contact: Intervocalic Fricatives in Barcelona and Valencia. *Languages*, 5(4), 60.
- Dumais, S. T. (2004). Latent semantic analysis. *Annual Review of Information Science and Technology (ARIST)*, 38, 189-230.
- Evangelopoulos, N. E. (2013). Latent semantic analysis. *Wiley Interdisciplinary Reviews: Cognitive Science*, 4(6), 683-692.

- Evert, S. (2008). Corpora and collocations. *Corpus Linguistics. An International Handbook*, 2, 1212-1248.
- Fairclough, N. (1995). *Media discourse*. Edward Arnold.
- Fairclough, N. (2015). *Language and Power* (3.^a ed.). Routledge.
- Fernández, M. F. D. S. (2024). Sistemas de pensión latinoamericanos vistos desde la academia: Modelado de tópicos a partir de títulos de artículos científicos. *Revista Oratores*, (21), 42-59.
- Ferrer, F. (2000). Languages, Minorities and Education in Spain: The Case of Catalonia. *Comparative Education*, 36(2), 187-197. <http://www.jstor.org/stable/3099867>
- Ferrer, R. C. (2010). Changing linguistic attitudes in Valencia: The effects of language planning measures. *Journal of Sociolinguistics*, 14(4), 477–500.
- Firth, J. R. (1957). *Papers in Linguistics 1934-1951*. Oxford University Press.
- Foucault, M. (1972). *The Archaeology of Knowledge*. Tavistock.
- Fradejas Rueda, J. M. (2010). *Las lenguas románicas*. Arco/ Libros.
- Garrett, P. (2001). Language attitudes and sociolinguistics. *Journal of sociolinguistics*, 5(4), 626-631.
- Generalitat de Catalunya. (2023). *Enquesta d'usos lingüístics de la població 2023*. Departament de Política Lingüística y l'Institut d'Estadística de Catalunya. https://llengua.gen-cat.cat/ca/serveis/dades_i_estudis/dades/enquesta-eulp/2023/, consultado el 5 de febrero de 2026.
- Generalitat Valenciana. (s. f.). *Traductor Salt*. <https://salt.gva.es/es/traductor>, consultado el 16 de diciembre de 2025.
- Generalitat Valenciana. Ley 4/1983, de 23 de noviembre, de uso y enseñanza del valenciano. *Boletín Oficial del Estado*, 24 de enero de 1984, núm. 20. <https://www.boe.es/eli/es-vc/l/1983/11/23/4/con>, consultado el 3 de febrero de 2026.

Generalitat Valenciana. Ley 7/1998, de 16 de septiembre, de Creación de la Academia Valenciana de la Lengua. *Boletín Oficial del Estado*, 21 de octubre de 1998, núm. 252.

<https://www.boe.es/eli/es-vc/l/1998/09/16/7/con>, consultado el 3 de febrero de 2026.

Generalitat Valenciana. Ley 4/2018, de 21 de febrero, por la que se regula y promueve el plurilingüismo en el sistema educativo valenciano. *Boletín Oficial del Estado*, 13 de marzo de 2018, núm. 63. <https://www.boe.es/eli/es-vc/l/2018/02/21/4>, consultado el 3 de febrero de 2026.

Generalitat Valenciana. (2021). *Conocimiento y uso social del valenciano*. [Síntesis de resultados]. Presidencia de la Generalitat. <https://presidencia.gva.es/es/web/politica-linguistica/enquestes-situacio-valencia>, consultado el 5 de febrero de 2026.

Gimeno-Menendez, F., y Gómez-Molina, J. R. (2007). Spanish and Catalan in the Community of Valencia. *International journal of the sociology of language* 184: 95-107. Print.

Hardt-Mautner, G. (1995). *'Only Connect'*. *Critical Discourse Analysis and Corpus Linguistics*. Lancaster: UCREL.

Hong, L., y Davison, B. D. (2010). Empirical Study of Topic Modeling in Twitter. *Proceedings of the First Workshop on Social Media Analytics*, 80-88.

Horowitz, D. L. (2001). *The Deadly Ethnic Riot*. University of California Press.

Hunston, S. (2022). *Corpora in Applied Linguistics* (2.^a ed.). Cambridge University Press.

Jiang, L. (2023). *Modelado de temas en documentos de texto: análisis comparativo de LSA, PLSA y LDA*. Valencia: Universitat Politècnica de València. Trabajo de Fin de Máster.

Kherwa, P., y Bansal, P. (2020). Topic Modeling: A Comprehensive Review. *EAI Endorsed Trans. Scalable Inf. Syst.*, 7(24).

Kroskrity, P. V. (2004). Language Ideologies. *A Companion to Linguistic Anthropology*, (1.^a ed.). Blackwell. 496-517.

Lado, B. (2011). Linguistic landscape as a reflection of the linguistic and ideological conflict in the Valencian Community. *International Journal of Multilingualism* 8.2: 135-150.

Ley Orgánica 6/2006, de 19 de julio, de reforma del Estatuto de Autonomía de Cataluña. *Boletín Oficial del Estado*. núm. 172, de 20 de julio de 2006.

<https://www.boe.es/eli/es/lo/2006/07/19/6/con>, consultado el 3 de febrero de 2026.

Liu, G. Z. (1999). Virtual Community Presence in Internet Relay Chatting. *Journal of Computer-Mediated Communication*, 5(1), JCMC514.

Lledó-Guillem, V. (2023a). Defending the Catalan Language: Cristòfol Despuig and the Description of the Valencian other. *Linguistica*, 63(1-2), 233-250.

Lledó-Guillem, V. (2023b). Ideologies lingüístiques a la Comunitat Valenciana: Un estudi introductor. *Quaderns d'Estudi 23*. València: Càtedra de Drets Lingüístics. Universitat de València.

Llorach, E. A. (1961). Historia y estructura en los sistemas vocálicos hispanos. *Estudios románicos*, 8, 105-116.

Mateu, J. À. C. (2023). A la caza del intelectual. *Pasajes*, (67), 110-124.

Mifrah, S., y Benlahmar, H. (2020). Topic modeling coherence: A comparative study between LDA and NMF models using COVID'19 corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4), 5756-5761.

Mills, S. (1997). *Discourse*. Routledge.

Mourelle de Lema, M. (1982). El valenciano, lengua autóctona. *Thesaurus: Boletín del Instituto Caro y Cuervo*, 37(2), 255-267.

Nuolijärvi, P. (2000). Sociolingvistiikka kielentutkimuksen kentässä. *Kieli, diskurssi & yhteisö*. Jyväskylän yliopisto, soveltavan kielentutkimuksen keskus, 13-38.

Parodi, G. (2008). Lingüística de Corpus: Una introducción al ámbito. *RLA. Revista de Lingüística Teórica y Aplicada*, 46(1), 93-119.

Perea, M. P. (2024). Una aproximación a la ideología lingüística de Antoni M. Alcover. *Estudios de Lingüística del Español*, 48, 162-177.

Pradilla, M. A. (Ed.). (2022). *Catalan sociolinguistics : state of the art and future challenges* (1.ª ed.). John Benjamins Publishing Company. <https://doi.org/10.1075/ivitra.32>

Puchal, P. P. (1962). La abolición de los fueros de Valencia y la Nueva Planta. *Saitabi*.

Qaiser, S., y Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International journal of computer applications*, 181(1), 25-29.

Real Decreto 2003/1979, de 3 de agosto, por el que se regula la incorporación de la Lengua Valenciana al sistema de enseñanza del País Valenciano. *Boletín Oficial del Estado*, 23 de agosto de 1979. núm. 202. pp. 19761-19762. <https://www.boe.es/eli/es/rd/1979/08/03/2003>, consultado el 3 de febrero de 2026.

Recasens i Vives, D. (2014). *Fonètica i fonologia experimentals del català: Vocals i consonants* (Vol. 74). Institut d'Estudis Catalans.

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (*Reglamento General de Protección de Datos*). *Diario Oficial de la Unión Europea*, L 119/1, de 4 de mayo de 2016.

Ricento, T. (2005). *An Introduction to Language Policy: Theory and Method*. John Wiley & Sons.

Rodríguez, P. (2016, 21 de octubre), Forocoches: el foro más influyente de España. *El País*. https://elpais.com/tecnologia/2016/10/21/actualidad/1477067074_754840.html?utm_source=chatgpt.com, consultado el 12 de marzo de 2026.

Rojas, D. (2012). Actitudes lingüísticas de hispanohablantes de Santiago de Chile: creencias sobre la corrección idiomática. *Onomázein*, (26), 69-93.

Rumsey, A. (1990). Wording, Meaning, and Linguistic Ideology. *American Anthropologist*, 92(2), 346-361.

Sajavaara, P. (2000). Kielenohjailu. *Kieli, diskurssi & yhteisö*. Jyväskylän yliopisto, soveltavan kielentutkimuksen keskus, 67-106.

Sanchis Guarner, M. (1985). *La llengua dels valencians*. Valencia.

Santana Suárez, O., Pérez Aguiar, J. R., Sánchez Berriel, I., y Gutiérrez Rodríguez, V. (2011). Extracción automática de colocaciones terminológicas en un corpus extenso de lengua general. *Procesamiento del lenguaje natural*, 47, 145-152.

Schieffelin, B. B., Woolard, K. A., y Kroskrity, P. V. (2023). *Language Ideologies: Practice and Theory*. Oxford University Press.

Schiffman, H. F. (2012). *Linguistic Culture and Language Policy*. Routledge.

Sentencia 31/2010, de 28 de junio de 2020. *Boletín Oficial del Estado*. núm. 172, de 16 de julio de 2010, pp. 1-491. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2010-11409, consultado el 3 de febrero de 2026.

SimilarWeb. (2025). Datos demográficos del tráfico de ForoCoches.com. <https://www.similarweb.com/es/website/forocoches.com/#geography>, consultado el 4 de septiembre de 2025.

Spolsky, B. (2004). *Language Policy*. Cambridge University Press.

Stubbs, M. (1991). *Discourse Analysis: The Sociolinguistic Analysis of Natural Language*. John Wiley & Sons.

Sugiura, L. (2016). Researching online forums: Ethics case study. *British Sociological Association*. Durham, UK. *British Sociological Association*.

Sunderland, J. (2004). *Gendered discourses* (Vol. 191). Basingstoke: Palgrave Macmillan.

Swiggers, P. (2019). Ideología lingüística: dimensiones metodológicas e históricas. *Confluência: Revista do Instituto de Língua Portuguesa*, (56), 9-40.

Szwed, P. (2021). Is web scraping legal? A short guide on scraping under EU law. *Discover Digital Law*.

Tann, K. (2010). Imagining Communities: A Multifunctional Approach to Identity Management in Texts. *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation* (1.ª ed.). Continuum. 149-177.

Tekijänoikeuslaki 404/1961. Helsinki 8.7.1961. <http://data.finlex.fi/eli/sd/1961/404/ajan-tasa/2025-06-27/fin> consultado el 20 de octubre de 2025.

- Teubert, W. (2005). My version of corpus linguistics. *International Journal of Corpus Linguistics* 10:1, 1-13.
- Torlà, L. (2024). El primer documento en valenciano es este manuscrito de hace casi mil años. *El Debate*, 28 de octubre 2024, https://www.eldebate.com/espana/comunidad-valenciana/20241028/primer-documento-valenciano-este-manuscrito-hace-casi-mil-anos_237860.html, consultado el 2 de diciembre de 2024.
- Vallverdú, F. (1985). La lengua catalana: su historia y su presente desde una perspectiva sociolingüística. *Iberoamericana (1977-2000)*, 9(1 (24), 24-37.
- Van Dijk, T. A. (1993). Principles of Critical Discourse Analysis. *Discourse & Society*, 4(2), 249-283. <https://doi.org/10.1177/0957926593004002006>
- Vázquez-Sanchis, M. (2025). Els valencians accepten la unitat de la llengua? Comparativa sistemàtica de formulacions. *Treballs de Sociolingüística Catalana*, (35), 91-109.
- Veny, J. (1992). El futuro de la dialectología catalana. *Revista de llengües i literatures catalana, gallega y vasca*, 2.
- Weinreich, U., y Martinet, A. (1963). *Languages in Contact: Findings and Problems* (1.^a ed.). Mouton.
- Wodak, R., y Meyer, M. (2001). *Methods of Critical Discourse Analysis: Vol. v.v. 113* (R. Wodak y M. Meyer, Eds.; 1.^a ed.). Sage. <https://doi.org/10.4135/9780857028020>
- Woolard, K. A. (1992). Language ideology: Issues and Approaches. *Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)*, 2(3), 235-249.
- Woolard, K. A. (2016). *Singular and Plural: Ideologies of Linguistic Authority in 21st Century Catalonia*. Oxford University Press.
- Zhao, B. (2022). Web scraping. In *Encyclopedia of Big Data* (pp. 951-953). Cham: Springer International Publishing.

Apéndices

Apéndice 1. Suomenkielinen lyhennelmä

JOHDANTO

Kielen sosiolingvististä todellisuutta voidaan tarkastella puhujien diskursseissa heijastuvien kielellisten asenteiden ja ideologioiden kautta. Vielä tänäkin päivänä kysymys valencian kielen yhteydestä katalaaniin jakaa mielipiteitä. Uudelleenasetusteorian mukaan katalaani tuotiin Valencian kuningaskuntaan, kun Kataloniasta tulleet perheet alkoivat valloittaa ja uudisasuttaa aiemmin arabien hallitsemaa aluetta Valencian takaisinvaltauksen jälkeen vuonna 1238. Kotoperäisyyteen perustuva teoria ehdottaa, että valencian kieltä puhuttiin nykyisellä alueella jo ennen takaisinvaltausta. Vuoden 1982 Valencian itsehallintolaissa alueen nimeksi hyväksyttiin valenciankielinen ”Comunitat Valenciana” aikaisemman espanjankielisen ”País Valenciano” -nimityksen sijaan, ja alueen omaksi kieleksi hyväksyttiin nimitys ”valenciano”. Siitä lähtien tämä konflikti ei ole koskenut pelkästään kielellisiä asioita, vaan se jakaa yhteiskuntaa ja on juurtunut sen rakenteisiin.

Tässä pro gradu -tutkielmassa analysoin valencian ja katalaanin kieliä koskevia diskursseja Espanjan suurimmalta keskustelufoorumilta (ForoCoches) kerätyissä kommentteissa. Tarkoituksena on selvittää, millaisia näihin kieliin liittyviä diskursseja löytyy, ja miten erilaiset kieli-ideologiat ilmenevät niissä. Erityisen kiinnostavaa on tutkia valencian kielen nimitykseen liittyvän konfliktin nykytilaa puhujien näkökulmasta. Tulee mainita, että tässä työssä käytän termiä ”kieli” ainoastaan muodollisena kategoriana viittaamaan nimityksiin ”valencian kieli” ja ”katalaani” ottamatta kantaa niiden filologiseen tai sosiopoliittiseen tulkintaan.

Vaikka uudelleenasetusteoria on ainut tieteellisesti hyväksytty teoria valencian kielen alkuperästä, keskustelu kielen synnystä jatkuu edelleen niin akateemisella, poliittisella, institutionaalilla kuin sosiaalisellakin tasolla (Ballester, 2021: 10). Vaikka Valencian kieliakatemia (AVL, *Acadèmia Valenciana de la Llengua*) vuoden 2005 lausunnon mukaan valencian kieli on katalaanin varieteetti, johon voidaan viitata nimellä ”valenciano”, nimitykseen liittyvä epäselvyys ei ole ratkennut. *Valenciano* sai virallisen kielen lisäksi myös paikallisen kielen tittein vuoden 2006 uudessa itsehallintolaissa, mutta siinä ei mainittu sen yhtenäisyydestä katalaaniin. Tätä keskustelua on syytä tutkia, sillä AVL:n vuoden 2005 lausunto närkästytti joitakin ryhmiä, jotka kokivat sen osoitukseksi hallituksen epälojalisuudesta aluetta ja sen kieltä

kohtaan (Lado, 2011). Analysoidessaan sosiaalisen median diskursseja tämä tutkimus tuo uuden näkökulman kiistaan, joka on vielä kahden vuosikymmenen jälkeen ratkaisematta.

Korpus koostuu vuosien 2015 ja 2024 välillä julkaistuista 2 164 kommentista, jotka on kerätty yhdeksästä valencian ja katalaanin kieliä käsittelevästä keskusteluketjusta. Kvalitatiivisen analyysin tukemiseksi suoritan korpusavusteisen diskurssianalyysin (*Corpus-Assisted Discourse Studies*, CADS) käyttäen korpuslingvistiikan menetelmiä, kuten topiikkimallinnusta ohjelmalla Orange (versio 3.39.0) ja kollokaatioanalyysiä ohjelmalla AntConc (versio 4.3.1). Koska AntConc mahdollistaa avainsanojen tarkastelun kontekstissaan, kollokaatioanalyysi toimii välivaiheena, joka vahvistaa mallinnetut topiikit ja osoittaa suunnan lähilukua varten valikoituville diskursseille. Tämä monimenetelmällinen tutkimus yhdistää kvantitatiivisia ja kvalitatiivisia menetelmiä, ja sen tutkimuskysymykset ovat seuraavat:

- 1) Mitkä ovat valencian ja katalaanin kieliä käsittelevien keskustelujen pääteemat?
- 2) Millaisia valencian ja katalaanin kieliin liittyviä diskursseja korpus sisältää?
- 3) Miten kieli-ideologiat rakentuvat ja heijastuvat näissä diskursseissa?

Tuloksiin vaikuttavia tekijöitä ovat muun muassa koulutus, poliittiset ideologiat ja kansallinen identiteetti. Valencian itsehallintoalueella mielipiteet vaihtelevat kielellisen yhtenäisyyden hyväksyjien ja valencian kielen ainutlaatuisuuden puolustajien välillä, kun taas Kataloniassa vallitsee näkemys, että ne ovat sama kieli alueellisine varieteetteineen (Davidson, 2020). Koska kommentit ovat anonyymejä, informanttien asenteisiin mahdollisesti vaikuttavia sosiaalisia tekijöitä tai asuinpaikkoja ei voida ottaa huomioon tämän tutkimuksen tuloksissa.

KIELI JA YHTEISKUNTA

Sosiolingvistiikka tarkastelee sosiaalista todellisuutta kielitieteen keinoin tutkimalla kieltä sen sosiaalisessa ympäristössä. Koska tarkastelun kohteena on kielen käytön ja sosiaalisen rakenteen välinen suhde, useiden määritelmien keskeinen ydin on, että kieli heijastaa yhteiskuntaa, mutta myös rakentaa sitä. Makrososiolingvistiikka keskittyy siihen, mitä yhteisöt tekevät kielilleen, jolloin tutkitaan kielen vaihtoa, säilymistä ja kuolemaa, kielellisiä asenteita sekä eri kielimuotojen tehtäviä kieliyhteisössä (Nuolijärvi, 2000: 13–15).

Elävän kielen tulee vastata yhteiskunnassa tapahtuviin muutoksiin. Kielipolitiikkaa ja -suunnittelua tarvitaan demokraattisessa yhteiskunnassa yhteisen kielen määrittämiseen hallinnollisissa asioissa, koulutuksessa, viestinnässä sekä taloudellisessa ja kulttuurisessa elämässä. Ne

voivat myös edistää poliittisia tavoitteita, kuten kansallisen identiteetin luomista. Valtiollista olemassaoloon rakentava pieni kansakunta voi oman kielen suunnitellun edistämisen kautta luoda siitä keskeisen symbolin kansalliselle identiteetille. Tällöin se on osa ideologista pyrkimystä luoda kansan yhtenäinen tahto väestön keskuudessa. Oman kansalliskielen vakaa asema voi myös vahvistaa sekä yksilön että yhteisön identiteettiä (Sajavaara, 2000: 67).

Kieli-ideologiat voidaan määritellä joukoiksi yleisesti hyväksytyjä käsityksiä kielen luonteesta maailmassa (Rumsey, 1990). Woolard (1992) korostaa, etteivät kieli-ideologiat ole vain sosiaalisten rakenteiden heijastumia, vaan ne sekä muovaavat niitä että muovautuvat niiden mukana. Siksi ideologioiden roolia yhteiskuntien valtdynamiikan ylläpitämisessä tai kyseenalaistamisessa kannattaa tutkia, mikä edellyttää niitä pohjaavien kulttuuristen ja poliittisten kontekstien ymmärrystä. Kieli-ideologia voi rakentaa sillan sekä kielen rakenteen ja kielipoliitiikan että kielitieteen ja yhteiskuntatieteiden välille, ja siten auttaa uudelleenmäärittelemään analyttisiä ongelmia, kuten esimerkiksi poliittisia kysymyksiä (Woolard, 1992).

Kieli-ideologiat vaikuttavat identiteetin rakentamiseen esimerkiksi erilaisten sosiaalisten ja kulttuuristen identiteettien, kuten kansallisuuden ja etnisyyden luomisen ja esittämisen kautta. Puhujien ideologioissa yhdistyvät kielelliset ja diskursiiviset resurssit sekä sosiokulttuurinen kokemus, sillä sen piirteet heijastuvat heidän käyttämiinsä kielellisiin ja diskursiivisiin muotoihin. Nationalismia ja etnisyyttä käsittelevissä tutkimuksissa kieltä pidetään usein keskeisenä tekijänä, ja painotetaan kielen ja diskursiivisten muotojen roolia kansakunnan rakentamisprosessissa (Kroskrity, 2004: 509).

Sekä kieli-ideologiat että -asenteet heijastavat ja muovaavat tosielämän kieleen liittyvien asioiden käsittelyä (Becker, 2025: 309–310). Kieli-ideologiat viittaavat yhteiskunnassa kieleen liittyviin, jaettuihin uskomuksiin, tunteisiin ja käsityksiin, jotka yhdistävät kielen ja yhteiskunnan dialektisella tavalla. Kielelliset asenteet taas ymmärretään sekä yksilöiden tapana käsitellä erilaisia kieleen ja viestintään liittyviä tilanteita että heidän taipumuksenaan suhtautua niihin, mikä kertoo myös heidän asenteistaan tiettyjä kielellisiä muotoja käyttäviä ihmisiä kohtaan (Becker, 2025: 309–310).

Viranomaisten määräykset ovat historiallisesti pakottaneet valencian kielen puhujat epätasaroiseen diglossian muotoon. Katalaaneista poiketen monet valencialaiset ovat hyväksyneet oman kielensä halveksunnan. Näiden tunnepoliitikkojen (*affective politics*) kuvaamiseksi esitetty, sisäistettyyn kielihierarkiaan viittaava käsite ”self-hate” viittaa tilanteeseen, jossa vähemmistöyhteisön jäsenet vähättelevät tai halveksivat itseään omaksuttuaan heitä syrjivän

enemmistöryhmän arvotuksia (Pujolar, 2021). Nämä ennakkoluulot rakentavat diglossisia ideologioita, ja ne usein yhdistetään käsitykseen siitä, että valencian kielen käyttöä edistetään pakottavasti (Colom i Ortiz, 2003: 20).

Valencian alueen kieli-ideoiksi on ehdotettu assimiloivaa espanjamielisyyttä, antikatalanistisuutta, sekä espanja- että valenciामीelistä kaksikielisyyden edistämistä, ja normalisoivaa ideologiaa (ks. Xambó, 1997). Valencialaisen kielikontekstin tärkeimmiksi kysymyksiksi on määritelty espanjan ja valencian kielen välinen suhde, kielellisen liberalismien ja interventio- nismien välinen jatkumo, katalaanin ja valencian kielen välinen suhde sekä kielellinen vaihtelu ja sen suhde normiin (Castañer, 2018). Lisäksi valencialaiseen kontekstiin voidaan soveltaa myös kulttuurisen nationalismien, kielellisen jatkuvuuden ja sosiolingvistisen naturalismin kieli-ideoita (Lledó-Guillem, 2023b).

NYKYTILANNE

Valencian itsehallintoalueella vallitsee sekä kielellinen että ideologinen konflikti. Valencian kieleen liittyvä konflikti on syntynyt poliittisesta ja sosiaalisesta jännitteestä, joka vallitsi Francon diktatuurista demokratisoitumisen aikana. Irtautumista kannattava poliittinen liike (*blaverisme*) pitää valencian kieltä ja katalaania kahtena eri kielenä, mikä aiheuttaa vastakainasettelua valencialaisen yhteiskunnan ryhmien välille (Gimeno-Menéndez ja Gómez-Molina, 2007). Vuosituhannen alussa tehdyt kyselytutkimukset paljastivat, että noin 60 % kyselyyn vastanneista valencialaisista kannatti separatistista kantaa, kun taas noin 30 % yhtenäisyyttä, ja loput 10 % ei halunnut tai osannut vastata. On todettu, että instituutiot eivät osaa kuvata ilmiötä aiheuttamatta jännitteitä väestön keskuudessa, minkä vuoksi sitä pidetäänkin sosiolingvistisenä tabuna Valencian itsehallintoalueella (Vásquez-Sanchis, 2025).

METODOLOGIA

Tässä tutkimuksessa käytetään monimenetelmällistä lähestymistapaa, joka koostuu kolmesta osasta: topiikkimallinnuksesta, kollokaatioanalyysistä ja kriittisestä diskurssianalyysistä. Korpus sisältää 2 164 verkkoharavoitua (*web scraping*) kommenttia ForoCoches-sivuston keskusteluketjuista, jotka käsittelevät valencian ja katalaanin kieliä. Aineistossa on kaiken kaikkiaan 80 674 sanaa (*token*) ja 9 428 tyyppiä (*type*). Haravoitavat keskusteluketjut valittiin manuaalisesti niiden ajankohtaisuuden ja sisällön merkittävyyden perusteella. Niistä kolme on vuodelta 2022, kaksi vuodelta 2018, ja yhdet vuosilta 2024, 2017, 2016 ja 2015.

Tutkimuksen ensimmäinen vaihe on topiikkimallinnus, jota sovelletaan suuriin asiakirjakoelmiin avainkäsitteiden tai merkittävien piirteiden löytämiseksi (Kherwa ja Bansal, 2019). Sen tarkoituksena on löytää tiedostokokoelman aiheet automaattisesti käyttämällä erilaisia algoritmeja, jotka tarjoavat monipuolisia tekniikoita piilevien merkitysten löytämiseen ja aiheiden ryhmittelemiseen (Hong ja Davison, 2010). Topiikkimallinnukseen käytettiin LDA-algoritmia ilman TF-IDF-menetelmää, ja aiheiden lukumääräksi asetettiin kolme.

Seuraavaksi tarkastellaan kollokaatioita, eli sanapareja, jotka voivat osoittaa suhteellisen ainutlaatuisia tai yleisiä suhteita sanojen välillä (Firth, 1957). Kollokaatio viittaa sanojen väliseen suhteeseen, mutta sen tarkka tunnistaminen vaatii konkordanssien yksityiskohtaisempaa tutkimista. Sosiolingvistikissa tutkimuksissa kollokaatioanalyysi voi paljastaa toistuvia malleja, joilla on erityisiä, meille tiedostamattomia tehtäviä, ja siten edistää argumentaation tai ideologioiden tutkimista (Baker, 2010). Kollokaatioanalyysi tehtiin käyttämällä topiikkimallinnuksessa esiin nousseita avainsanoja. Avainsanojen tarkastelu niiden konteksteissaan mahdollisti myös lähiluvussa analysoidavan alakorpuksen kokoamisen.

Viimeinen vaihe on kriittinen diskurssianalyysi (*Critical Discourse Analysis*, CDA), jolloin ideologioiden uudelleen tuottamista tutkitaan ottamalla dataan etäisyyttä ja sijoittamalla se sosiaaliseen kontekstiinsa (Wodak ja Meyer, 2001: 9; Van Dijk, 1993). CDA tarjoaa keinoja paljastaa tiettyjä aiheita käsittelevien tekstien taustalla piileviä ideologisia oletuksia sekä yhdistää nämä kielelliset ilmiöt niiden tuotannon ja käytön sosiopoliittisiin konteksteihin (Fairclough, 2015: 106–108). Lisäksi sen yhdistäminen sosiolingvistiikkaan edesauttaa identiteettiin liittyviä tutkimuksia kielitieteen alalla (Bednarek, 2010: 226–227). Tässä vaiheessa perehdyttiin ensiksi sataan kollokaatioanalyysissä poimittuun kommenttiin kiinnittäen huomiota topiikkimallinnuksessa nousseisiin aiheisiin ja kieli-ideologioiden heijastumiseen, minkä perusteella valittiin 22 kommentin alakorpus lähilukua varten. Analyysi pohjautuu kriittisen diskurssianalyysin kolmiulotteiseen malliin, joka koostuu (1) tekstin kuvauksesta, (2) tekstin ja vuorovaikutuksen välisen suhteen tulkinnasta ja (3) vuorovaikutuksen ja sosiaalisen kontekstin välisen suhteen selittämisestä (Fairclough, 2015: 128).

TUTKIMUSTULOKSET

Arvioin etukäteen, että korpuksessa toistuvia teemoja olisivat kielten vertailu, alkuperä ja politiikka, ja että diskurssit olisivat luonteeltaan argumentatiivisia ja viittaisivat kielten sanasto- ja ääntämiseroihin sekä joihinkin poliittisiin näkökohtiin ja historiallisiin tapahtumiin. Lisäksi

uskoin, että mielipiteitä ja viittauksia sosiolingvistiseen tilanteeseen tarkastelemalla diskursseissa voitaisiin havaita erilaisia kieli-ideologioita.

Topiikkimallinnus paljasti kolme pääteemaa: (T1) valencian ja katalaanin kielten vertailu, (T2) valencian kieli koulutusjärjestelmässä ja (T3) valencialaisten kansallinen identiteetti. On syytä huomata, etten osannut arvioida koulutusjärjestelmän muodostavan yhtä teemoista. Tämä ei kuitenkaan ole odottamaton tulos, sillä kielipolitiikkaan kuuluvat myös opetusta koskevat päätökset. Kollokaatioanalyysin avulla tarkistin muun muassa noodin ”valenciano” kollokaatioiden ”oppiaine” ja ”oppitunnit” esiintymisen perusteella, että koulutusjärjestelmä on toistuva aihe korpuksessa.

Kriittisessä diskurssianalyysissä kiinnitin huomiota kielten arvottamiseen, viittauksiin kielten alkuperästä ja historiallisista tapahtumista sekä kansalliseen identiteettiin. Kieliä arvottavissa diskursseissa väitettiin toisaalta, että valencian kieli on huonosti puhuttua katalaania, sen murre tai loukkaus sitä kohtaan, ja toisaalta, että katalaani on huonosti puhuttua valencian kieltä tai kieroutunut espanjan kielen murre. Lisäksi ilmaistiin, että katalaani kuulostaa valencian kieltä jurommalta ja karkeammalta, ja että kielten yhtenäisyyttä puoltavat ovat ”feminat-sikatalaaneja”. Joissakin kielten alkuperään tukeutuvissa argumenteissa viitattiin muun muassa kielten samoihin juuriin vulgaarilatinassa, sekä katalaania edistäneisiin toimiin – kuten sanakirjaan – verraten Katalonian kielipolitiikkaa Valencian hallituksen vähäiseen valencian kielen tukemiseen. Kansallista identiteettiä käsittelevissä diskursseissa esiin nousivat esimerkiksi (i) maininnat valencian kielen alueellisesta identiteetistä ja erityisestä historiasta, (ii) näkemys siitä, ettei katalaanin puhuminen määritä valencialaisia katalaaneiksi, (iii) olettaus, että katalaanit haluavat valloittaa Valencian itsehallintoalueen sekä (iv) uhka polttaa katalaankieliset kirjat Valencian Fallas-tapahtumassa. Lisäksi yhden kommentin kirjoittaja painotti oman valencialaisuutensa seurauksena olevansa täten myös espanjalainen, toisin kuin katalaanit, jotka ovat maanpettureita. Eräs kommentoija ilmaisi olevansa valencialainen, mutta halua vansa hävittää valencian kielen maan päältä, ja toinen, ettei puhu valencian kieltä, koska se on turha kieli, joka tulisi korvata opetussuunnitelmassa hyödyllisemmällä kielillä. Nämä viimeiset heijastivat sekä sisäistetyn kielihierarkian diglossisia ideologioita että assimiloivaa espanjamielisyyttä.

YHTEENVETO

Suuri osa alakorpuksen diskursseista oli argumentatiivisia, mutta toisinaan mielipiteitä ilmaistiin ilman vahvoja perusteluja, välillä halventavaa kieltä ja jopa vihapuhetta käyttäen. Tämä

osoitti, että valencian kieleen liittyvä poliittinen ja ideologinen konflikti aiheuttaa edelleen voimakkaita reaktioita, koska se on juurtunut yhteiskuntaan vuosikymmenten, ellei jopa vuosisatojen ajan. Sosiopoliittisen luonteen lisäksi myös epävarmuus valencian kielen alkuperästä tekee polemiikin epäselvyyksien ratkaisemisesta entistä vaikeampaa.

Teoreettisten ja historiallisten seikkojen – kuten kielipolitiikan, ideologioiden ja kielellisten asenteiden sekä kielten sosiolingvistisen historian ja nykytilanteen – ja diskurssien välillä voitiin havaita selviä yhteyksiä. Esimerkiksi tietyt kieli-ideologiat, kuten antikatalanismi ja ”nativistinen” kieli-ideologia vaikuttivat puhujien diskursseissa ilmeneviin kielellisiin asenteisiin. Muita näkökohtia olivat kielellinen variaatio ja sen suhde normiin, espanjan ja valencian kielen välinen suhde, kansallinen identiteetti, ”self-hate” ja kielellinen yhtenäisyys. Voidaan todeta, että se, miten kieli nähdään identiteetin merkkinä vaikuttaa huomattavasti kieli-ideologioihin ja -asenteisiin, jotka puolestaan muovaavat yhteiskunnan kielikulttuuria, johon kieli-suunnittelu ja -politiikka perustuvat. Tämän tutkimuksen johtopäätöksenä voidaan todeta, että vaikka esiin nousi kolme pääteemaa, valencian kieltä koskevan polemiikin ydinkysymys näyttää liittyvän kansalliseen identiteettiin. Tästä syystä sen ratkaiseminen edellyttää syvällistä perehtymistä puhujien kieli-ideologioihin ja -asenteisiin.