



Universidad
Zaragoza

Trabajo Fin de Grado

La aplicación de los sistemas complejos al estudio de la variación lingüística:

El caso de la pérdida de la F- inicial latina

Autor/es

Lucía Pitarch Ballesteros

Director/es

Elena Albesa Pedrola

Facultad / Escuela

Facultad de Filosofía y Letras. Filología Hispánica

Septiembre 2020

Resumen

Este proyecto adapta el modelo computacional de Pardo-Gordó (2020° “Cultural Hitchhiking” para indagar en las aplicaciones del marco teórico de los sistemas complejos a la variación lingüística. La investigación simula distintos aspectos de la pérdida de la F- inicial latina relevantes en la teoría sustratística de Menéndez Pidal y evalúa hasta qué punto este tipo de modelos se adaptan a la realidad lingüística y pueden ser útiles para su investigación.

Agradecimientos

La naturaleza interdisciplinar de este trabajo me ha dado la oportunidad de aprender de muchas personas, por eso me gustaría darles las gracias. Le quiero agradecer a Pedro que me descubriera los sistemas complejos, que me ayudase a entenderlos y, sobre todo, su continuo apoyo en todo este trabajo. A mi madre por su cariño, a mi padre por la curiosidad y la inquietud científica que me ha transmitido desde pequeña. A Salvador Pardo-Gordó le quiero agradecer su generosidad, su programa, el ánimo de compartirlo y el ayudarme a usarlo y adaptarlo. A Ricardo Badía le quiero agradecer sus correos y su paciencia, ayudando a una total desconocida a comprender los Sistemas de Información Geográfica. Y finalmente a Elena, quien sabiendo que nos adentrábamos en un campo desconocido decidió aceptar el reto y apoyarme en toda la investigación y en las dudas que se abren en un último año de carrera.

Índice

Índice	4
1. Introducción.....	5
2. Aplicaciones del marco teórico de los sistemas complejos a la lingüística 7	
2.1. Sistemas complejos: una intuición muy presente en la lingüística	7
2.2. ¿Qué es la complejidad?	8
2.3. Sistemas complejos adaptativos y modelos basados en agentes: NetLogo 10	
3. F- inicial latina.....	12
3.1. El sustrato vasco como inductor de su pérdida.....	13
3.2. Evolución de la F- inicial latina	15
3.3. F- inicial en Aragón	17
4. Modelización de la pérdida de la F- en la península	19
4.1. El modelo base: Cultural Hitchhiking	19
4.2. Adaptación del modelo a la lingüística.....	21
4.3. Resultados de las simulaciones	24
5. Conclusiones.....	30
6. Bibliografía:.....	32

1. Introducción

En el libro *Orígenes del español* (1964), Menéndez Pidal atribuyó la pérdida de la F- en posición inicial de palabra a la influencia del sustrato vasco, sin embargo, esta teoría ha sido muy debatida, y aún a día de hoy, no se ha encontrado ninguna respuesta concluyente. La falta de una decisión unánime se debe, sobre todo, a que los datos que existen sobre la desaparición de dicha consonante son muy escasos y dispersos –hecho que suele suceder en la lingüística histórica–. Por ello, aquí se plantea abordar la hipótesis desde la aplicación de un modelo computacional, basado en la teoría de los sistemas complejos, que simula la dispersión del fenómeno lingüístico.

Este trabajo pretende introducir una nueva visión desde la que acercarse a esta hipótesis y demostrar con ella la viabilidad y utilidad de este tipo de simulaciones en el campo de la lingüística diacrónica y de contacto.

El programa de simulación utilizado se basa en el código creado por Salvador Pardo Gordó: “Cultural Hitchhiking” (2020). El programa original busca simular la dispersión cultural en el neolítico y ver la relación que tuvo con la agricultura. La base teórica que tiene detrás es, por un lado, la teoría del autoestop cultural y, por otro, las herramientas y marco teórico proporcionado por el campo de los sistemas complejos, y en concreto, los modelos basados en agentes. En este trabajo se ha adaptado dicho programa a la dispersión lingüística¹; –se contempla simultáneamente la dispersión diacrónica y la diatópica–.

Las ventajas de la aplicación de este modelo (y de otros basados en la teoría de los sistemas complejos) a la lingüística de contacto y a la histórica son varias:

- permite mostrar tanto el fenómeno lingüístico de manera individual, como su evolución;
- rellena los huecos de la información perdida;
- presenta una alta capacidad para trabajar con gran cantidad de datos, sintetizarlos y representarlos de manera visual;

¹ Para consultar/usar el programa escribir a: pitarchlucia@gmail.com

- permite trabajar la lengua como *continuum* frente a una visión de las lenguas como unidades lingüísticas discretas, escapando la separación tradicional entre lenguas para facilitar el estudio;
- representa al mismo tiempo una visión macroescalar (desde las lenguas) y una microescalar (como puede ser el hablante, o las apariciones individuales de un determinado fenómeno lingüístico, en el caso de este trabajo, las desapariciones de la F- latina);
- da gran importancia a las relaciones e interacciones entre los hablantes
- y, finalmente, es una teoría que tiene como concepto clave y axial el cambio y la evolución, que son rasgos inherentes a las lenguas.

1. Aplicaciones del marco teórico de los sistemas complejos a la lingüística ²

1.1. Sistemas complejos: una intuición muy presente en la lingüística

Quizá, en un comienzo, mezclar física y lingüística parezca algo arriesgado, pero, al pensar en algunos ejemplos, puede verse que ya se ha estado haciendo desde hace un tiempo, como por ejemplo para medir los sonidos en fonética.

El hecho de concebir las lenguas –y también el lenguaje– como sistemas complejos, también se ha hecho ya. Hay quienes se han acercado a este enfoque de manera intuitiva: Saussure, por ejemplo, cuando dijo “la lengua es un *sistema* de signos arbitrarios”³ ya anticipó esa cuestión de que hay elementos que interactúan entre sí y que se dan sentido los unos a los otros; no solo como suma de todos, sino también por cómo se relacionan y la forma en la que se posicionan en el sistema. De su cita se deduce, que, aunque dos frases estén compuestas por los mismos elementos, su orden y estructura hacen que puedan tener distintos significados, pues no es lo mismo: “El bocadillo se comió a María” que “María se comió al bocadillo”.

Más adelante, Chomsky (por tomar otro de los autores claves de la lingüística moderna) en la teoría de Principios y Parámetros ya vio que con unas reglas limitadas y, en principio, simples (lo que él llamó *Principios*) los hablantes somos capaces de crear mensajes tremendamente complejos e ilimitados. Esta capacidad de “hacer emerger”, crear, desde unas reglas muy simples formas muy complejas es una cuestión que ha interesado mucho a autores como Stephen Wolfram o John Conway y a quienes desde la física y la teoría de sistemas han desarrollado teorías como la de “The New Science” (“La nueva ciencia”) o “The Game of Life” (“El juego de la vida”). Sus teorías prueban esa intuición chomskiana en el sentido en el que demuestran que, partiendo de reglas muy básicas y limitadas, se generan, al mismo tiempo, ciertos patrones repetitivos que proveen al sistema de regularidad y ciertas variaciones, a través de las cuales, estos evolucionan y se adaptan creando nuevas reglas del juego. En una charla, Larsen-Freeman (2016) describía esto parafraseando a Gleick de la siguiente manera: “The act

² Para profundizar en este tema se recomienda: Larsen-Freeman (2008) y Massip Bonet et al. (2019)

³ El subrayado es nuestro

of playing the game has a way of changing the rules” es decir, desde el momento en el que jugamos a un juego, empezamos a cambiarle las reglas. Como observaba Larsen en su charla, esto es lo que hacemos los hablantes con la lengua: desde el momento en el que la hablamos, ya estamos cambiándola y generando nuevas variantes semánticas, fonéticas, sintácticas...

Como ejemplo más cercano a nuestra lengua destaca el diccionario *Redes* (2004) de Ignacio Bosque, similar al concepto que utiliza *Wordnet* (2010). El funcionamiento de estos diccionarios se fija en las relaciones que existen entre diversos elementos y cómo se influyen unos a otros, o se llaman los unos a los otros, siguiendo la teoría de “mundo pequeño”, fenómeno que describe que la máxima distancia que hay entre personas es de otras 6 personas. Según esta teoría, por ejemplo, entre Chomsky y el lector hay un máximo de 6 personas que los separan y, si siguiera la red que forman, en tan solo 6 correos podría estar hablando con él. Esto mismo sucede con las palabras, desde *manzana* es posible llegar a *fútbol* en tan solo 6 palabras que las relacionan, ventaja que aprovechan los autores citados en sus diccionarios.

Estas intuiciones llevan ya unas cuantas décadas en funcionamiento, sin embargo, es a partir de los comienzos del siglo XXI cuando realmente se han empezado a estudiar las conexiones entre los sistemas complejos y la lingüística. Una cosa que me sorprendió y me maravilló al empezar a realizar este estudio es que toda la gente a la que leía estaba viva. Y esto es genial: los estudios de este campo están en pleno momento de descubrimiento y debate. Así, este es aún un campo de estudio muy joven y como se pretende demostrar en este trabajo, merece la pena explorarlo, pues puede aportar herramientas muy útiles para el estudio de la lingüística. El marco teórico de los sistemas complejos puede ayudar a entender desde cómo se organizan e interactúan nuestras neuronas para permitirnos entender y emitir mensajes, hasta cómo se extiende y triunfa una innovación lingüística, pasando a formar parte del propio sistema.

1.2. ¿Qué es la complejidad?

Hay muchas perspectivas desde las que se pueden interpretar los sistemas complejos, y por ello son muchas las definiciones que se les han dado. Aquí se ha querido dar una definición que sintetiza las ideas de varios autores que los trabajan desde diferentes perspectivas. Un sistema complejo puede definirse como una red

formada por un gran número de elementos que, al interactuar siguiendo reglas sencillas, crean comportamientos en los que se ve un equilibrio entre lo regular y lo aleatorio. Los organismos que forman el sistema no tienen un punto de control central que les diga cómo comportarse o que domine el modo de comportamiento. Además, esos sistemas tienen en ellos la información necesaria para evolucionar y hacer emerger desde esas reglas sencillas una estructura compleja, de forma que los individuos crean y modifican esta estructura que los une, y al mismo tiempo esa estructura los modifica a ellos.

Para entender a qué se refiere “equilibrio entre aleatorio y complejo” conviene tomar el ejemplo de Mitchell (2009: 97-99) del ADN. ¿Qué es más complejo: un texto generado de manera aleatoria con las letras ATGC o una cadena de ADN? Si *complejidad* equivale a *aleatoriedad* entonces la respuesta sería la primera. Pero, entonces, ¿qué pasa con la segunda? El ADN de alguna forma también está formado de manera aleatoria, pues cada individuo es distinto, pero al mismo tiempo tiene también unas cuantas regularidades que le permiten crear organismos funcionales; esta mezcla en el ADN de lo aleatorio y lo regular requiere mucha más información que la de la cadena que es únicamente aleatoria, y por ello puede decirse que el ADN es más complejo. Este mismo fenómeno es observable en las lenguas: si se le introduce a un ordenador un diccionario y se le pide que una palabras de manera aleatoria generará un texto incomprensible. El cerebro, sin embargo, es capaz de generar unas reglas para, de ese mismo número limitado de palabras, poder generar mensajes que comuniquen, como este texto que estás leyendo ahora.

Visto todo esto, parece quedar claro que la capacidad del lenguaje es un sistema complejo. Se puede incluso ir más allá y concebir a una comunidad de habla y a su lengua como un sistema complejo. Pruebe a concebir a los hablantes como nodos (puntos en el espacio), cada hablante interactúa con los otros y, al hacerlo, genera una red de hablantes (representada por las líneas que los unen). Las relaciones entre estos hablantes van cambiando a lo largo del tiempo, modifican el sistema y, al mismo tiempo, este les modifica a ellos. Este sistema (lengua) y estos nodos (hablantes) a su vez están en contacto continuo con otros sistemas (culturales, históricos, ecológicos...). Son demasiadas variables para imaginarlas todas a la vez y en detalle y, además, observar cómo evolucionan, se influyen y modifican a lo largo del tiempo... Por ello,

las aplicaciones prácticas de las teorías de los sistemas complejos pueden muy relevantes e inspiradoras para un campo como la lingüística de contacto.

1.3. Sistemas complejos adaptativos y modelos basados en agentes: NetLogo

En este trabajo, en concreto, se concibe la lengua como un sistema complejo adaptativo, siguiendo la definición del grupo clave de investigación de los sistemas complejos aplicados a la lingüística: The five graces group. Los sistemas complejos adaptativos se oponen a los sistemas complejos no adaptativos. Mientras que los primeros hacen referencia a sistemas que tienen capacidad de cambio y aprendizaje, como la lengua o el comportamiento de un grupo de hormigas; los segundos, no tienen esa capacidad de evolución y son, por ejemplo, los huracanes, o el polen movido por el viento.

La explicación de enmarcar la lingüística en la primera clase es la siguiente:

The system is adaptive; that is, speakers' behaviour is based on their past interactions, and current and past interactions together feed forward into future behaviour. A speaker's behaviour is the consequence of competing factors ranging from perceptual constraints to social motivations. The structures of language emerge from interrelated patterns of experience, social interaction, and cognitive mechanisms (The Five Graces Group, 2009:2)⁴

Como se expone en la introducción de este trabajo, la lengua es inherentemente cambiante y dinámica, ya que se origina y modifica en y por las constantes interacciones entre los hablantes.⁵

Los modelos basados en agentes son una herramienta computacional que permite representar este tipo de sistemas, ya que permiten reproducir sus características básicas: autonomía, proactividad, reactividad (percepción y respuesta del entorno por

⁴ El sistema [la lengua] es adaptativo; es decir, el comportamiento de los hablantes se basa en sus interacciones pasadas, y, al mismo tiempo, las interacciones actuales y las pasadas son las que conformarán el comportamiento futuro. El comportamiento del hablante es la consecuencia de la sinergia de factores que van desde las restricciones perceptivas del hablante hasta su motivación social. La estructura lingüística emerge de los diferentes patrones que se interrelacionan en la experiencia comunicativa, en la interacción social y en sus procesos cognitivos. (Traducción propia)

⁵ Para una reflexión más extensa entorno a este tema se puede consultar: López Rivera, 2013.

parte de los agentes o unidades implicadas en el modelo), movilidad, emergencia, autoorganización y aprendizaje-adaptabilidad.

NetLogo es una plataforma de modelos de uso libre, desarrollada por Uri Wilensky que permite y facilita la programación de este tipo de simulaciones. El lenguaje computacional en el que se basa NetLogo es Python. En este trabajo se utiliza el código creado y accesible desde esta plataforma por Salvador Pardo-Gordó. Se ahondará en las características del código más adelante en este trabajo.

2. F- inicial latina

Las palabras que en latín comenzaban con F- evolucionaron de diferentes maneras en las lenguas romances. En la mayoría de casos, esta F- se ha mantenido (en el gallego, italiano, francés, catalán, aragonés...), sin embargo, en español, por lo general, lo que se observa es su pérdida.

En el paso del latín al español se cree que la F- relajó su pronunciación y se aspiró a [h-] y que esta relajación articulatoria fue incrementando hasta resultar en el cero fonético o pérdida de la F- inicial latina. Esto es lo que sucedió por ejemplo en la palabra *harina* proveniente del étimo latino FARINA. Sin embargo, en algunos contextos fónicos determinados la F- latina, en castellano, ha dado otros resultados (Alarcos, 1992: 37):

- Frente a *-ue* y *-r* la F- se ha mantenido: *fuego* < FOCUM, *frente* < FRONTIS.
- Ante *-i*: la evolución varía y da como resultado tanto la pérdida: *higo* < FICUM como el mantenimiento: *fin* < FINEM.
- Delante de la *-l* a veces palataliza: *llama* < FLAMA, y a veces se mantiene: *flor* < FLOR.
- En casos esporádicos, la F- velariza: *jamelgo* < FAMELICU, donde /f/ > [h] > /x/.

Además, ya que las lenguas no son entidades homogéneas, sino que, al contrario, tienen la variación como principio axial, algunas variedades diatópicas del castellano también presentan resultados diferentes al estándar. En el judeoespañol sirve para distinguir entre el grupo occidental, donde se mantiene, y el oriental, donde se pierde, debido a que los emigrantes a esas zonas viajaron desde las dos Castillas, donde ya había una pérdida de dicha F-. En cuanto a las hablas meridionales se aspira en el extremeño: [h]umo, [h]iguera; se velariza en el murciano popular: *jue* 'fue', *jurgar* 'hurgar', en el habla popular de Canarias: *jase* 'hace', igual que en gran parte de las hablas populares y de más bajo nivel de la Andalucía occidental, donde está socialmente estigmatizado. Y finalmente la aspiración de la F- en las hablas meridionales se extendió también a amplias zonas de América como Colombia, Venezuela, Chile..., o se produjo ya en el léxico castellano como hemos visto más arriba. (Lapesa, 2008; Moreno Fernández, 2009).

En cuanto a qué provocó los distintos resultados de la F- y de qué modo, existen diversas teorías. Algunas teorías profundizan en los procesos internos de la lengua (cómo el propio sistema lingüístico se fue reorganizando y potenciando los cambios dentro de sí mismo)⁶, autores como Penny (1972, en Quilis, 1996:338) elaboraron teorías basadas en la tendencia arcaizante y otros han seguido la hipótesis sustratística de Menéndez Pidal.

Este trabajo se centra en la última, sin embargo, en un estudio más extenso sería interesante simular las diferentes hipótesis a través del mismo programa o similares, y comparar los resultados.

2.1. El sustrato vasco como inductor de su pérdida

Menéndez Pidal (1985:122), el padre de esta teoría, anota que:

Desde el siglo XI se encuentran en la región setentrional de Burgos, en la Rioja y en el Alto Aragón ejemplos como hayuela, Rehoyo, Ormasa (...); estas regiones se encuentran inmediatas al país vasco, donde también la f fue siempre un sonido exótico; por esto debemos suponer que la sustitución de la f- por la h-, y subsiguiente pérdida en Castilla es un fenómeno primitivo hijo de la influencia ibérica, de los dialectos indígenas vecinos al vasco. También en Gascuña, colindante con el país vasco francés, se trueca la F por una aspiración.

Esta tesis fue apoyada por autores como Lapesa (2008), quien aporta además los siguientes argumentos a favor: que el vasco no parece tener una /f/ originaria, de hecho, en los latinismos la /f/ se sustituye por /p/ o /b/ y que, el vasco poseía una /h/ aspirada que pudo sustituir la /f/ con la que a veces alterna.

Las fronteras geográficas en las que se sitúa la isoglosa /f/h/, es decir, zonas en las que se divide por una línea imaginaria el territorio en el que se pronuncia /f/ y en los que se aspira, o no se pronuncia, también favorecen en gran medida la teoría sustratística. Estas fronteras Menéndez Pidal (1964:214) las sitúa en las orillas del río Sella.

⁶ Alarcos (1992:32-37), Naro (1972, en Calvo del Olmo, 2012: 132).



Ilustración 1: Isoglosa río Sella

Frente a estos datos, se han elaborado varios argumentos en contra (o por lo menos, críticos) con dicha teoría. Igartua (2011: 901-902) los resume de la siguiente manera:

1. Problemas factuales: que F- no cambie ni ante diptongo ni ante *-r* hace sospechar de la teoría sustratística. Pues, ¿por qué el vasco iba a reconocer la F- en estos contextos fonéticos y no en los otros?
2. Problemas contrafactuales: en navarro y en aragonés, a pesar de la influencia vasca que tuvieron, se mantiene F-.
3. Problemas argumentativos: el vasco reemplaza /f/ por /b/ o /p/, no por /h/, aunque esto Martinet (en Igartua, 2011:901) lo contraargumenta explicando que /f-/ evolucionó a /ph-/que más adelante terminó de aspirar totalmente hasta: /h-/
4. Problemas tipológicos: estos engloban tanto la idea de Penny de que la correspondencia entre /f/ y /h/ se dio diacrónicamente en otras áreas que no tenían influencia vasca, como la idea de Wallace y Joseph (1991: 90-91) de que /f/ > /h/ se debe a un cambio natural en la lengua por ser una relajación articulatoria.
5. Problemas cronológicos: la sonorización intervocálica /f/ > /b/, /v/ es anterior a /f/ > /h/.

Estas cuestiones, aunque ciertamente critican aspectos clave de la teoría sustratística no son totalmente incompatibles con ella. Y es que es no debemos olvidar que varios fenómenos e influencias debieron modificar la lengua simultáneamente. Así, aunque como afirman estas teorías, el sustrato vasco no pudo ser el único inductor de la pérdida de la F- inicial latina, esto no niega que sí pudiera influir en ella, junto con otros factores. De hecho, está claro que la pérdida de la F-, así como su posterior recuperación (en palabras concretas), sucedieron por un cúmulo de factores, (como sucede siempre en la lengua, y por lo que, se defiende en este trabajo la implementación de nuevas técnicas que permitan analizar la evolución tan compleja de la lengua).

En la línea de esta argumentación está la idea de Alarcos (1992), quien prefiere, sin negar la teoría del sustrato, hablar de que lo que se produjo fue una situación de lenguas en contacto donde, ya en las cabezas de los hablantes, convivían las distintas influencias y variantes lingüísticas, de manera que, el suceso lingüístico se entiende como algo que sucedió a nivel de los hablantes en primera instancia y no a nivel lengua, lo que facilita la comprensión de por qué, si bien los distintos resultados de F- se concentran en las distintas áreas geográficas, no son homogéneos en toda la extensión de cada lengua.

También, Igartua, desde una visión general, y explicando el fenómeno desde lo que llama una “común área lingüística” explica que:

vincular las similitudes diacrónicas de la fonología de estas lenguas a la presencia de un área lingüística y no a la acción inductora de un sustrato es compatible con la constatación de las diferencias en la distribución de la aspiración que presentan castellano, gascón y euskera (...) lo que tienen en común es el desarrollo de un fonema —además de otras características—, pero no comparten todos los detalles de su distribución. (2011: 913)

2.2. Evolución de la F- inicial latina

Menéndez Pidal (1985: 122) encuentra las primeras aspiraciones de /f-/ en el siglo XI en Burgos, Rioja y Alto Aragón. Lapesa (2008: 46) adelanta la fecha de esas primeras apariciones al siglo IX (en concreto, encuentra un documento en el monasterio de Silos y sospecha que este fue escrito por aragoneses o riojanos cuyos rasgos lingüísticos más particulares se propagaron desde San Millán de la Cogolla). Otros autores, como Igartua (2011: 914), sugieren un origen del fenómeno en la zona

circumpirenáica y Michelena (1977: 205) lo adscribe al norte de Navarra. En el siglo XII ya se observan algunas apariciones más de [h-] (Pidal, 1985: 123).

En el siglo XIII, aunque la [f-] y la [h-] ya se estaban generalizando como alófonos –sólo en algunos contextos como, por ejemplo, en el registro más popular–, la [f] se prefiere, y en algunos casos se usa, incluso, como ultracorrección. (Lapesa, 2008: 128). La [h-] se considera una marca lingüística del registro popular (Pla Colomer, 2012: 794). En esta época Pla Colomer (2012:794) observa un “empleo gráfico de la <h-> sin valor fónico para respetar la etimología latina”, Quilis (1996: 395) llega a los mismos resultados.

El siglo XIV representa una época de gran inestabilidad del fenómeno: coexisten los dos “alófonos sociolectales”: [f-] en el registro culto y [h-] en el popular (Pla Colomer, 2012: 794); sin embargo, en el registro culto comienza a haber discrepancias y, mientras que Alfonso X aún no asume el cambio de /f-/>/h-/ por ser considerado una marca dialectal de Burgos, la aspiración ya se integra en *El Libro del Buen Amor*. (Lapesa, 2008: 258)

A finales del siglo XIV en obras como *El Rimado de Palacio* o el *Cancionero de Baena* ya se observa una pérdida de la aspiración (Pla Colomer, 2012: 794). La omisión de la aspiración se extiende desde Castilla la Vieja (Lapesa, 2008:273), situada en la zona norte de la península, zona que había funcionado de foco para la extensión de la aspiración [h-].

Según Pidal (1985: 121), la F- se conserva en la lengua escrita hasta finales del siglo XV. Los Reyes Católicos muestran un empleo indistinto de las variantes (Pla Colomer, 2012). En 1492 Nebrija defiende la [h] como lo más normalizado (Lapesa, 2008: 280) y es que, entre 1472-1525 “la literatura conserva abundantes restos de *f* inicial (...) pero es muy general la *h* (...) que se impone por completo entre 1500 y 1520” (Lapesa, 2008: 280)

Entre los siglos XV y XVI lo más extendido es la aspiración en [h] o, a veces, también en [x] (Pidal, 1985: 121, 123). En esos siglos hay dos normas muy diferenciadas: en Castilla la Vieja el cero fonético y en la norma de Toledo la conservación, bien sea con [x] aspirada o con [h].

En la primera mitad del siglo XVI Lapesa (2008: 368) observa que aún hay restos de una *f* escrita arcaizante.

En 1578 fray Juan de Córdoba da una muestra clara de la convivencia de la variante aspirada y la total omisión de la /f-/: “Fray Juan de Córdoba, que había salido de España hacia 1540, afirma en su *Arte en lengua Zapoteca* (Méjico, 1578): ‘Los de Castilla la Vieja dizen (...) *alagar*, y en Toledo *halagar*⁷.’” (Lapesa, 2008:371).

Desde Madrid, con el cambio de la Corte, la omisión de la /h-/ se generaliza (Lapesa, 2008: 372)

En este trabajo se realizarán diferentes simulaciones variando el punto de origen del fenómeno para observar cuál se ajusta más al resultado actual de la pérdida de la F- en Aragón.

2.3. F- inicial en Aragón

Una de las objeciones que se le ha puesto a la teoría sustratística ha sido la aparente contradicción que existe en Aragón, pues, si realmente el vasco fue el que indujo el cambio, y Aragón estaba en estrecho contacto con hablantes vascófonos, ¿cómo es que la pérdida de la F- en el territorio aragonés fue mucho más tardía y lenta que en Castilla?

M. Alvar en su estudio “Sobre la pérdida de la F- en el aragonés del siglo XI” (1947) observa que, mientras en Castilla en el siglo XI ya se documentan abundantes palabras sin F-, en Aragón tan solo se registran unos 3 casos. Dicha tendencia continúa en el s. XII. Arnal y Enguita (1993) llegan a resultados parecidos, encontrando también escasas desapariciones de la F-, y proponen el final del siglo XV como el momento clave para el devenir de la pérdida de la F-, ya que es en este siglo cuando el proceso de castellanización se acelera. F. Lázaro Carreter (1951, en Arnal y Enguita, 1993: 59) coincide con Arnal y Enguita diciendo que en el siglo XV aún se documenta un fuerte mantenimiento de la F-; en 1478 se registra la primera H-, y en 1518 se regulariza la desaparición. En cuanto a estudios más profundos de determinados territorios de Aragón, Terrado (1991) no encuentra casos de pérdida de F- en Teruel y alrededores en

⁷ El subrayado es nuestro

el periodo medieval y Albesa (2017), registra también pocas pérdidas en el Bajo Aragón. A través de una rápida búsqueda en CODEA se han llegado a resultado similares en este trabajo (se profundizará en ellos en el apartado 3.4)

En conclusión, la pérdida de la F- en Aragón ha sido escasa, pese a lo que cabría esperar por su cercanía al área vascófona. De hecho, Alvar (1947) admite que muchos de los casos de pérdida de F- en Aragón posiblemente fueran vasquismos o castellanismos.

Este trabajo analizará la expansión del fenómeno desde el siglo X (momento de orígenes de las lenguas romances en la península) hasta el siglo XV (momento en que finaliza la expansión del castellano en la mayor parte de Aragón). Las simulaciones realizadas aportarán imágenes que representen la dispersión por siglos de la pérdida de la F- que tendría que haberse dado por influencia del vasco. Estas imágenes se contrastarán con los datos conocidos y registrados de la pérdida en documentos de los mismos siglos para ver hasta qué punto el modelo creado se ajusta a la realidad lingüística estudiada.

3. Modelización de la pérdida de la F- en la península

3.1. El modelo base: Cultural Hitchhiking

El modelo en el que se basa este trabajo fue creado por Salvador Pardo-Gordó (2020) para simular cómo la agricultura facilitó la dispersión cultural en el primer Neolítico en el Mediterráneo Occidental. Para ello se basa en la teoría del autoestop cultural⁸ (cultural hitchhiking) en la que un fenómeno cultural favorable—llamado *gen favorable* en la teoría— arrastra e influye el avance de un fenómeno neutral. En su modelo, el gen favorable es la agricultura y el neutral que se ve arrastrado es la cerámica, que sería la muestra o la evidencia de la cultura.

Dicha simulación se creó en NetLogo, una plataforma que permite crear entornos de modelado basados en agentes, aplicando distintas herramientas propuestas en la teoría de los sistemas complejos.

En el modelo de Pardo-Gordó los agentes con los que trabaja son el entorno ecológico del momento (recreado a través de la recopilación de distintos datos meteorológicos de la época), y los yacimientos arqueológicos. Estos dos agentes se representan de manera geográfica en mapas. Sobre estos mapas se realiza la simulación.

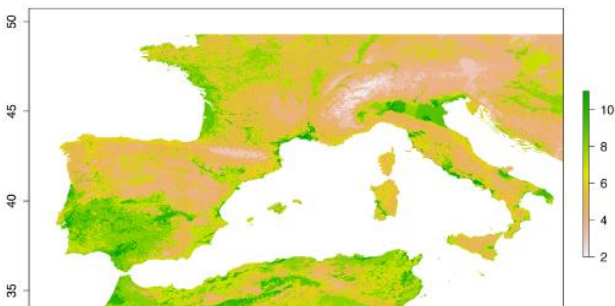


Ilustración 2: Mapa con la información ecológica
(Pardo-Gordó, 2015: 119)



Ilustración 3: Mapa con los yacimientos arqueológicos
(Pardo-Gordó, 2015: 47)

⁸ Más detalles sobre dicha teoría en Pardo-Gordó, 2015: 80.

La simulación comienza colocando el punto de origen de la dispersión en el modelo, que se representa con una x roja. Este punto de origen se puede modificar cambiando el valor de los botones de las coordenadas: X -Coord e Y -Coord.

Desde ese punto se realizan dos tipos de movimiento: el de ola de expansión de Fischer (Pardo Gordó, 2015: 122) y el denominado Leapfrog de Van Adel y Runnels (Pardo Gordó, 2015: 122). Nuestro trabajo incorpora ambos por pensar que la transmisión cultural debió de producirse de una manera similar a la que lo hace la dispersión lingüística.

La ola de expansión de Fischer consiste en que, desde una celda del mapa, el fenómeno estudiado se transmite a todas sus celdas vecinas.



Ilustración 4: Movimiento ola de expansión de Fischer (Pardo-Gordó, 2015: 122)

Por su parte, en el movimiento Leapfrog, desde una celda del mapa, el fenómeno se extiende a las celdas que tienen elementos favorables para su dispersión.⁹

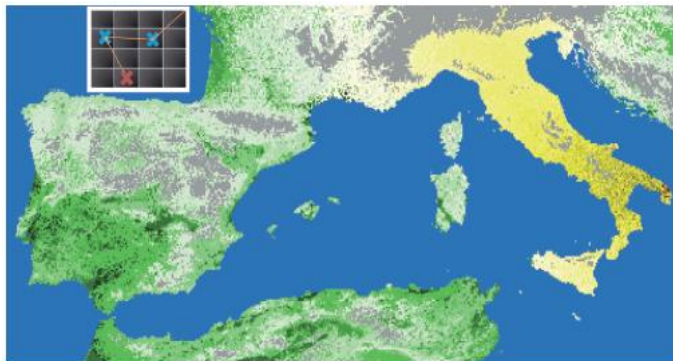


Ilustración 5: Movimiento Leap-frog (Pardo-Gordó, 2015: 122)

E

El modelo permite trabajar con un tipo de movimiento o con el otro y compararlos;

⁹ En el caso de Pardo-Gordó, la comunidad neolítica se dispersa desde una celda con características ecológicas favorables para el cultivo a otras con las mismas características.

además, permite modular la distancia del salto de celda que puede abarcar el movimiento Leap-frog con el botón *leap-distance*.

El siguiente par de variables, “biased/unbiased hitchhiking”, nos permite decidir qué tipo de autoestop cultural realizar. En el caso de nuestro programa, al elegir *biased* (“condicionado”) la expansión se realiza desde las celdas que primero han registrado la pérdida de la F-, y al elegir *unbiased* (“no condicionado”) la expansión se realiza desde todas las celdas que han registrado pérdida de F-, sin importar la antigüedad con la que lo hicieron.

El *slider* llamado *umbral-vascófono* permite modular los valores mínimos necesarios para que se realice la dispersión. En nuestro modelo, lo utilizamos para cambiar la zona de habla vascófono; correspondiendo el valor 0: a la zona no vascófono, 1: al área vascófono en el s. I, 2: área vascófono en el s. X y 3: área vascófono en el s. XV.

La *ratio de innovación* también se puede modificar para evaluar su influencia en el modelo, sin embargo, a través de sus estudios, Pardo-Gordo (2015) determina que la *ratio de innovación* más común en la expansión propuesta en la teoría del autoestop cultural es de 0.04 y 0.06, por lo que tampoco conviene salirse de esas cifras si se pretende obtener resultados ajustados a la realidad.¹⁰

3.2. Adaptación del modelo a la lingüística

En este trabajo se ha optado por simplificar el modelo, ya que pretende únicamente servir de base para futuras investigaciones y probar si este tipo de simulaciones pueden ofrecer resultados válidos en el campo de la lingüística diacrónica y de contacto.

En cuanto a los datos analizados en este trabajo, los agentes que se estudian como gen favorable, son aquellas casillas en el mapa en las que existe habla vasca y los agentes que representan el gen neutro que se relaciona con el habla vasca, son las

¹⁰ “Combinaciones” y “técnicas” son dos parámetros utilizados por Pardo-Gordó, sin embargo, en nuestro programa, debido a la simplificación de los datos realizada, han dejado de resultar útiles. Las mantenemos en el programa para evitar problemas en el código.

desapariciones de la F- inicial latina en los textos. De esta manera puede observarse cómo el habla vasca influyó en esa pérdida y de qué manera se extendió la pérdida de la F- en la península.

Para conseguir estudiar estos datos lingüísticos se ha realizado una simplificación de los datos de entrada al programa, con sus respectivas modificaciones en el código. Frente al mapa con la información ecológica de Pardo-Gordó que aunaba datos de precipitación, temperatura y desnivel, este trabajo utiliza un mapa en el que se distingue únicamente área de habla vasca de la no vasca. Este mapa fue creado a partir de la sintetización e informaticación de los mapas recopilados por Cid Abasolo (2002) a través de ArcGis (un programa de Sistemas de Información Geográfica).

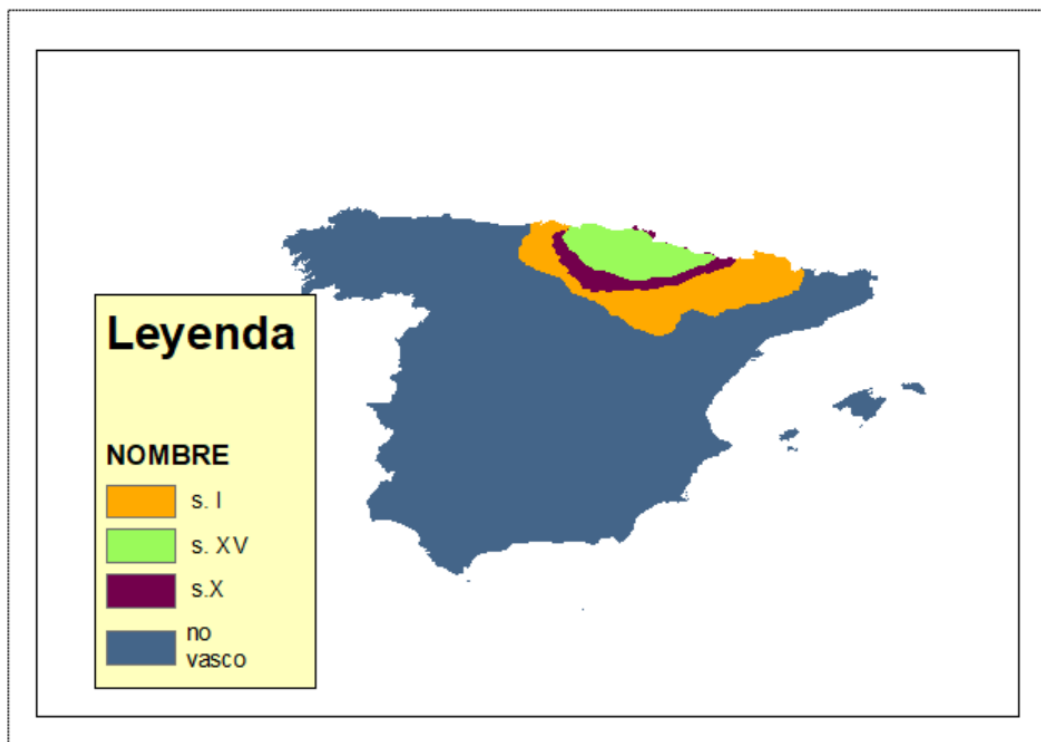


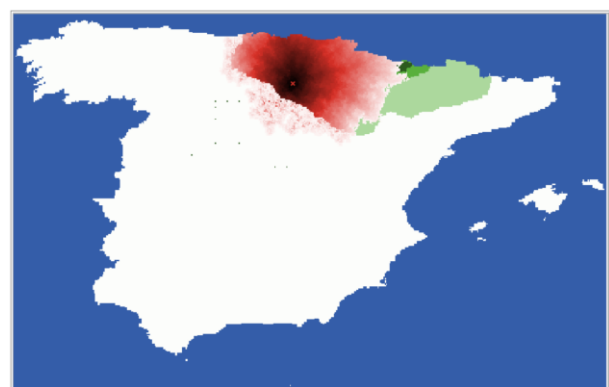
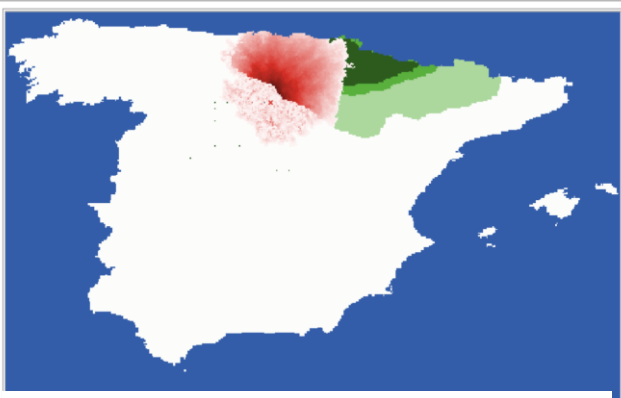
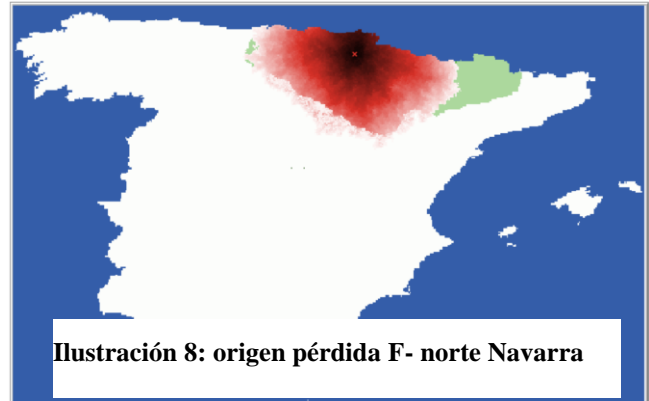
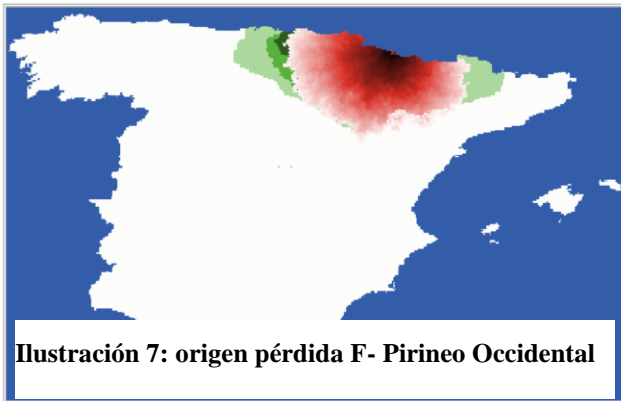
Ilustración 6: mapa zonas vascófonas

Sobre este programa ya modificado se realizan diferentes simulaciones para determinar de qué manera el vasco pudo influir a la pérdida de F- latina:

1. Simulaciones de la dispersión de la pérdida del s. I variando el punto de origen de la primera desaparición de la F-. Según las diferentes hipótesis de Michelena, Igartua, Pidal y Lapesa (ver apartado 2.2 para más información)
2. Simulaciones de la dispersión de la pérdida en las zonas vascófonas a lo largo del tiempo.
3. Variación de los parámetros existentes en el programa: ratio de innovación, tipo de movimiento y tipo de autoestop cultural.

3.3. Resultados de las simulaciones

Las primeras simulaciones realizadas tratan de ver qué punto de origen de la pérdida de la F-, de los propuestos por los distintos autores, se corresponde más con la realidad. Los resultados son los siguientes:



El punto de origen se muestra con una *x* roja; desde ahí se produce la dispersión. Cada uno de los cuadros rojos de la imagen representa una hipotética aparición en los textos de la pérdida de la F-. Cuanto más oscuros son los cuadros, más desapariciones de F- se representan en una misma área. Como puede verse, cuanto más cerca del punto de origen, más oscuros son los cuadros. Los parámetros utilizados para estas simulaciones han sido: 50 *ticks* (50 movimientos de expansión); el tipo de movimiento seleccionado es el de ola de expansión, pues pensamos que en un primer momento lo más factible era una expansión de la pérdida a los lugares vecinos; umbral-vasco 1, con lo que eliminamos como espacio favorable de expansión el área no vascófona; y tipo de autoestop *biased*.

La ilustración 7 muestra la simulación realizada colocando el punto de origen de la pérdida de la F- en los Pirineos Occidentales, siguiendo la hipótesis de Igartua (2011) y la ilustración 8, simula la hipótesis de Michelena (1977), de que la pérdida se originó al norte de Navarra. En estos dos primeros casos se observa una expansión que abarca gran parte del área vascofona, cosa que tendría sentido si se cumple la hipótesis sustratística de Pidal, sin embargo, al comparar las simulaciones con los datos referentes a la evolución de la pérdida de F- en el área de Aragón, no parecen ajustarse mucho a la realidad.

Mientras que las dos primeras simulaciones muestran ya cierta expansión de la pérdida por la región noroccidental de Aragón, Lázaro Carreter (1951, en Arnal y Enguita, 1993: 59) documenta las primeras pérdidas de F- en 1478; Alvar (1947) propone una fecha más temprana para la desaparición en Aragón, pero sigue sin concordar con los datos obtenidos de la simulación pues solo documenta 3 casos de pérdida en el s. XII (cuando en otras regiones de España la pérdida de F- ya estaba muy extendida); finalmente en la búsqueda realizada en CODEA para los primeros siglos no se ha encontrado ningún registro de la pérdida, siendo la primera aparición de H- en 1529 en Zaragoza.¹¹

En cuanto a la novena ilustración, la simulación se realizó situando el punto de origen en el monasterio de Silos. Este punto resulta sospechoso ya de primeras, puesto que se encuentra a las afueras de la región vascofona. Esto concuerda con la intuición de Lapesa (2008) de que, si bien el primer documento se encuentra en dicho monasterio, este no se redactó ahí, ni con las costumbres lingüísticas burgalesas de Silos, sino que este pudo ser traído desde San Millán de la Cogolla, en la Rioja (Ilustración 10).

Así, proponer como punto de origen para la pérdida San Millán de la Cogolla parece lo más coherente. Es un punto situado dentro del área vascofona, por lo que parece ser que el vasco pudo ser un inductor de la pérdida de F-, y además la dispersión

¹¹ Para llegar a estas observaciones se plantea de manera aproximada que cada paso de siglo equivale a 10 ticks en el programa. Estas simulaciones se han realizado durante 50 ticks por lo que tratan de demostrar la evolución del fenómeno a lo largo de 5 siglos. Para llegar a conclusiones más realistas convendría reslizar un estudio más detenido para averiguar a cuántos ticks equivale un siglo en esta simulación.

del fenómeno afecta más al área vascofona occidental dejando con desapariciones mucho más escasas (como muestran los textos) el área de Aragón.

Hasta aquí el modelo demostraría que un influjo vasco pudo ser posible y que este no es incompatible con que en Aragón se mantuviera la F-.

Lo siguiente por estudiar es cómo el modelo simula el paso del tiempo. Para estas simulaciones tomamos ya como punto de origen San Millán de la Cogolla, partiendo de ese punto, se han realizado 6 simulaciones, las tres primeras analizan el movimiento leapfrog y las 3 siguientes el de ola de expansión. Para cada movimiento se ha dejado actuar la simulación 10 ticks más que en la anterior.



Ilustración 11: Cogolla, 10 ticks, leapfrog

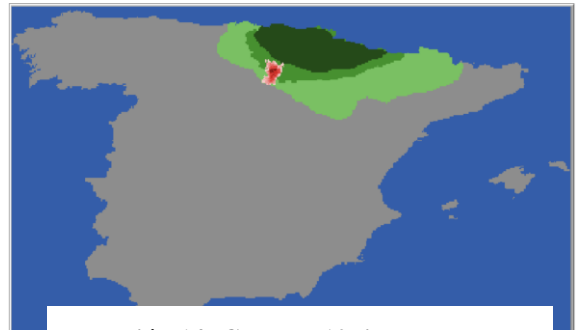


Ilustración 14: Cogolla, 10 ticks, ola

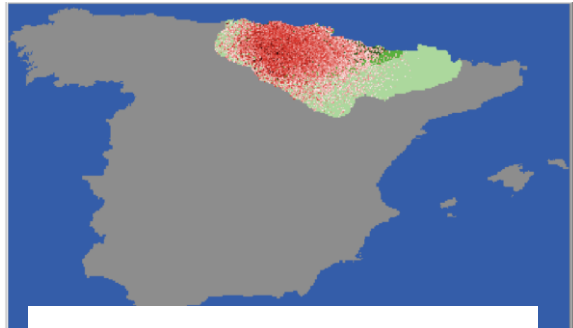


Ilustración 12: Cogolla, 20 ticks, leapfrog

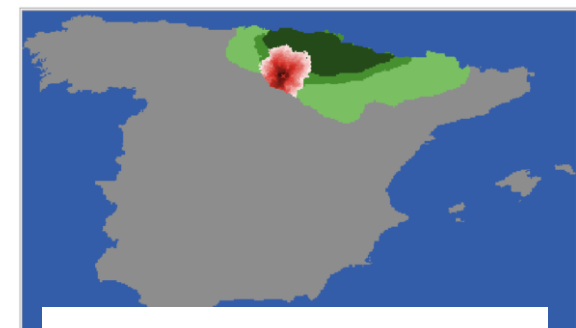


Ilustración 15: Cogolla, 20 ticks, ola

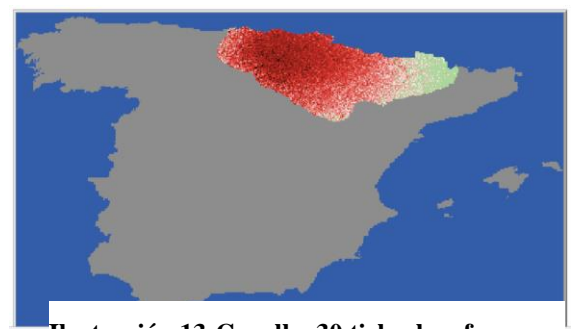


Ilustración 13: Cogolla, 30 ticks, leapfrog

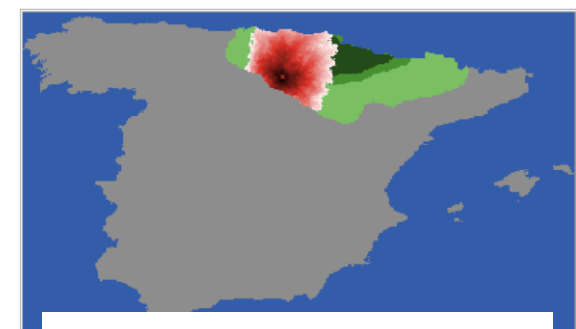


Ilustración 16: Cogolla, 30 ticks, ola

Probablemente la expansión de la pérdida de la F- sería una combinación de ambos movimientos, quizá en un principio sólo a las áreas vecinas, y más adelante a más áreas que presentasen características similares (en este caso ser hablantes de vasco). Esto explicaría que al principio la expansión de la pérdida estuviera más localizada y fuera más densa en un área concreta y que más adelante se fueran extendiendo las desapariciones de F- a áreas más lejanas. Pero esto se queda solo en un nivel hipotético, para poder llegar a conclusiones más certeras haría falta una documentación textual mucho mayor, no abaricable en este trabajo.

De momento, estas imágenes sirven para demostrar la validez del modelo a la hora de representar sistemáticamente y de manera discreta la dispersión diacrónica.

A continuación, se realizan algunas variaciones en los parámetros proporcionados por el programa para intentar ajustar los resultados de la dispersión y para comprobar las posibilidades del modelo.

Primero se muestra una comparación de los datos según el tipo de autoestop

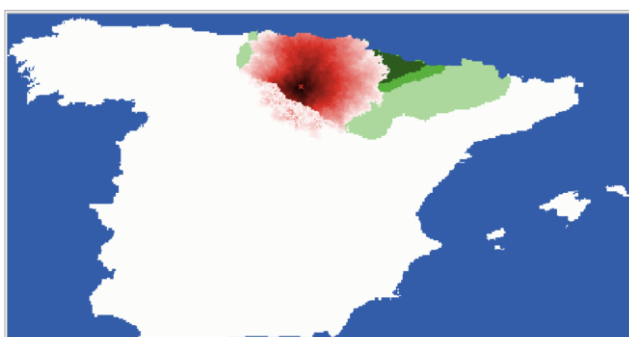


Ilustración 7: Cogolla, ola, 40 ticks, biased

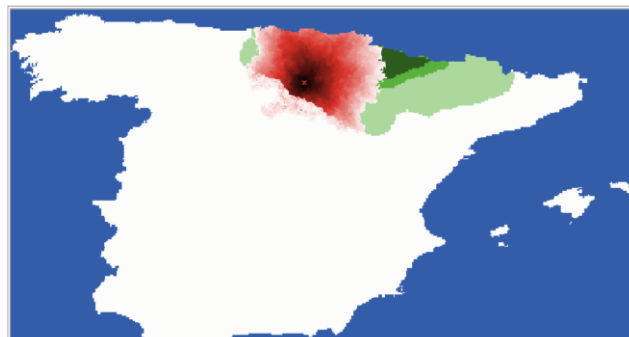


Ilustración 18: Cogolla, ola, 40 ticks, unbiased

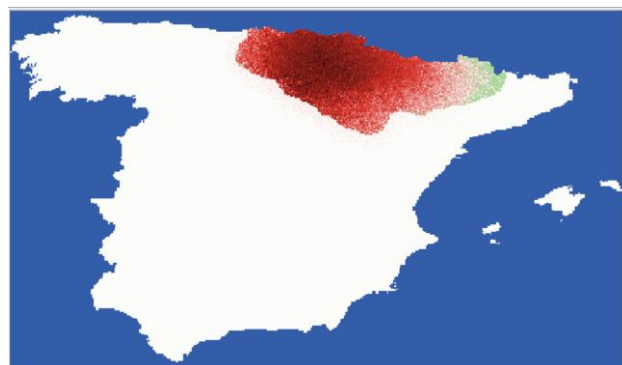


Ilustración 19: Cogolla, leapfrog, 40 ticks, biased

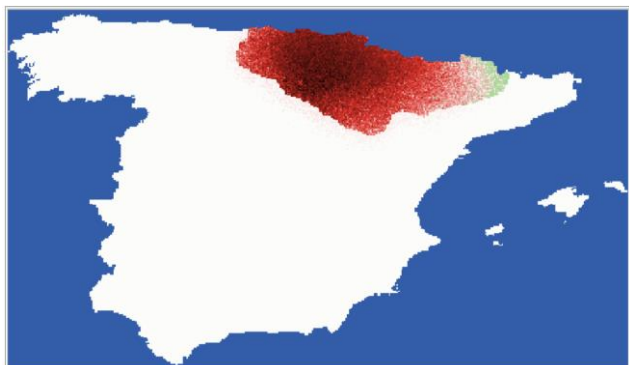


Ilustración 20: Cogolla, leapfrog, 40 ticks, unbiased

seleccionado:

En estas imágenes se observa muy poca variación. De aquí podemos concluir que el carácter *biased* o *unbiased* del autoestop tiene más relevancia en el movimiento de ola de expansión que en el de leapfrog.

Finalmente, presentamos unas imágenes en las que se ha ido cambiando la ratio de innovación. Este parámetro se considera útil para observar la velocidad y grado de innovación de los distintos fenómenos lingüísticos.

Las siguientes ilustraciones muestran simulaciones realizadas con ambos tipos de movimiento: leapfrog y ola, con autoestop biased, durante 20 ticks y lo que se ha variado para realizar la comparación ha sido la ratio de innovación. Esta se ha situado en los límites de los sugeridos como normales para la innovación cultural por Pardo-Gordó (2015), en 4 y 6.

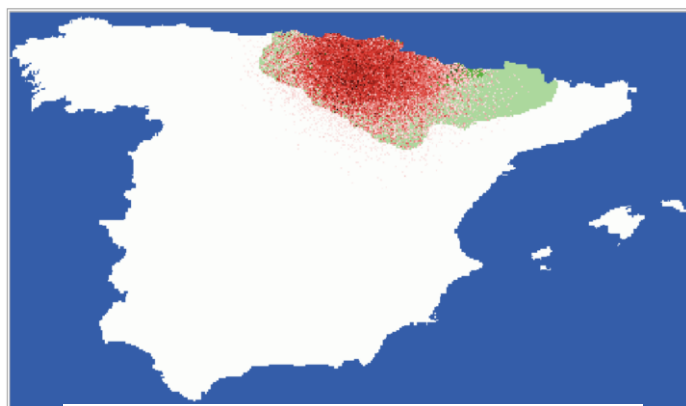


Ilustración 21: ratio innovación máximo, leapfrog

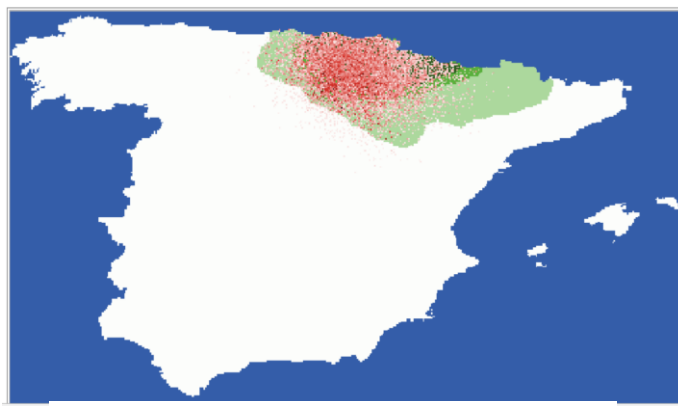


Ilustración 22: ratio innovación mínimo, leapfrog

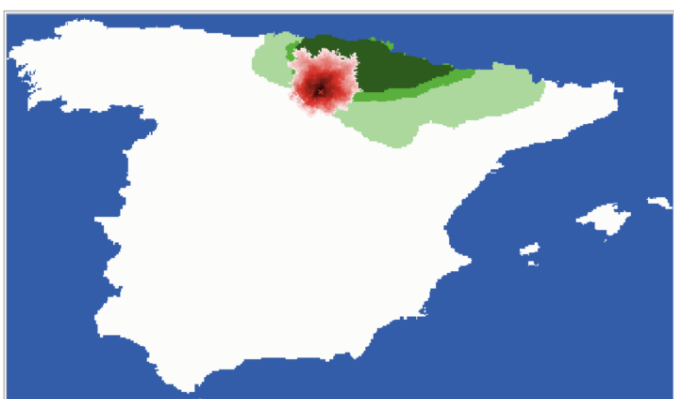


Ilustración 23: ratio innovación máximo, ola

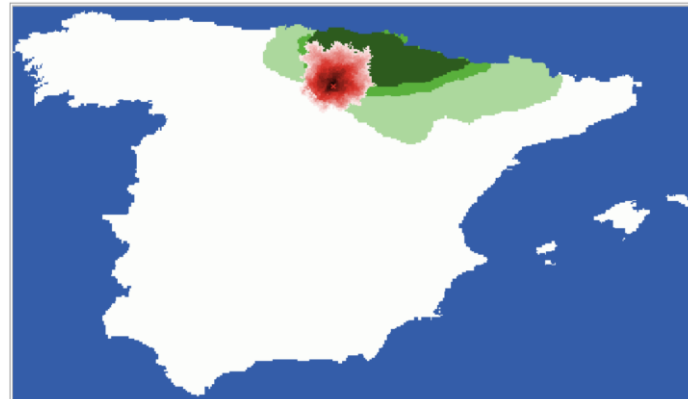


Ilustración 24: ratio expansión mínimo, ola

La diferencia en la ratio de innovación se ve sobre todo en el movimiento leapfrog; en la imagen 21 se ve cómo la simulación registra una pérdida de F- más densa y extendida en el mapa. En el movimiento de ola de expansión, la diferencia es muy leve.

La modulación de este parámetro en estas simulaciones no es muy representativa, debido a la falta de información con la que comparar los datos obtenidos, sin embargo, se presenta aquí como muestra de las posibilidades del modelo de cara a futuras aplicaciones. La aplicación de este parámetro podría usarse para comprobar la velocidad de expansión de diferentes fenómenos lingüísticos.

4. Conclusiones

Con este trabajo se ha mostrado solo una aplicación muy concreta y aproximada de un modelo basado en la teoría de los sistemas complejos. Sin embargo, y a pesar de la reducción y simplificación teórica del programa propuesto, este ya deja ver su validez, al demostrar una gran capacidad de síntesis de datos y fácil visualización de estos, ambas características muy valiosas para el estudio de la lingüística diacrónica y de contacto.

El modelo aquí planteado pretende ser un punto de partida, aún falta completarlo y afinarlo hasta poder resolver cuestiones como la validez de la teoría sustratística. Por ello, y vistos los datos obtenidos y los parámetros analizables cabe proponer las siguientes mejoras:

- indagar y profundizar en el alcance y ajuste de los parámetros estudiados a través de la resolución de preguntas como: ¿cuántos ticks equivalen a un siglo? o ¿de qué manera se produce la expansión de la pérdida de la F- (condicionada, no condicionada, con movimiento de tipo leapfrog, de ola de expansión, con qué ratio de innovación...)?
- trabajar desde el mismo modelo las distintas teorías de expansión de la pérdida de la F- (cambio interno de la lengua, reajustes en el sistema consonántico, otros sustratos...) y compararlas para ver cuál se ajusta más a la realidad,
- añadir información relevante para la expansión lingüística en el mapa como rutas comerciales o ponderación de núcleos culturales o políticos que podrían haber fomentado la dispersión de una lengua o fenómeno lingüístico.

Este trabajo también ha servido para observar algunas de las dificultades que puede plantear la aplicación de este tipo de herramientas a la lingüística, siendo las principales, que al ser un campo muy nuevo la información creada y los programas utilizados aún necesitan desarrollarse mucho y que la información debe ir creándose y aprendiéndose prácticamente sin guía y desde cero. Debido a la amplitud y diversidad de los conocimientos necesarios para realizar este tipo de modelos se recomienda la creación de grupos interdisciplinarios para su aplicación, en el caso concreto de este

trabajo se recomienda la colaboración de lingüistas, geógrafos y programadores con conocimientos de modelos basados en agentes y de sistemas complejos dinámicos.

5. Bibliografía:

- Alarcos Llorach, E. (1992). De F- latina en castellano. *Aetas del XVIIe. Colloque International de Linguistique Fonctionnelle*. León: Universidad de León, Secretariado de Publicaciones, 31-39.
- Albesa Pedrola, E. (2017). *La lengua en el Bajo Aragón a través de la documentación notarial (1450-1453). Transcripción y estudio lingüístico*. (tesis doctoral). Zaragoza: Universidad de Zaragoza.
- Alvar, M. (1947). Sobre pérdida de F- en el aragonés del siglo XI. *Archivo de Filología Aragonesa*, II, 151-162.
- Arnal Purroy, M.L. y Enguita Utrilla, J.M. (1993). Aragonés y castellano en el ocaso de la Edad Media. *Aragón en la Edad Media*, 10-11, 51-84.
- Bosque, I. (2004). *Redes. Diccionario combinatorio del español contemporáneo*. Madrid: SM.
- Calvo del Olmo F. J. (2012). Una frontera lingüística en las lenguas románicas: la pérdida de f-latina en castellano. *Abehache*, 2, 127-142.
- Cid Abasolo, C. (2002). Las fronteras de la lengua vasca a lo largo de la historia. *Revista de Filología Románica*, 19, 15-36.
- GITHE (Grupo de Investigación Textos para la Historia del Español): *CODEA+ 2015 (Corpus de documentos españoles anteriores a 1800)* [en línea] Sitio web [consultado en 3 agosto 2020]
- Igartua, I. (2011). Historia abreviada de la aspiración en las lenguas circumpirenaicas. En: *Euskaltzaindiaren XVI*. Biltzarra.
- Five graces group. (2009). Language is a complex adaptive system. *Language Learning*, 59, 1-26.
- Lapesa, R. (2008). *Historia de la lengua española*. Madrid: Gredos.
- Larsen-Freeman, D., y Cameron, L. (2008). *Complex Systems and Applied Linguistics*, New York: Oxford University Press.
- Larsen-Freeman, D. [Oxford University Press ELT]. (10 de mayo 2016). Complex, not complicated: Diane Larsen Freeman on Complexity Theory in Applied Linguistics (Part 1). Recuperado de: https://www.youtube.com/watch?v=SX_ItiZfUOQ.
- López Rivera, J, J. (2013). Aplicación del marco de los sistemas complejos adaptativos a un modelo de variación lingüística. *Moenia*, 19, 5-24.

- Massip Bonet, A. (2013). Language as a Complex Adaptative System: Towards an Intergrative Linguistics. En Massip Bonet, A., y Bastardas Boada, A., (Eds.), *Complexity Perspectives on Language, Communication and Society*, Berlin: Heidelberg, Springer-Verlag
- Menéndez Pidal, R. (1964). *Orígenes del español*. Madrid: Espasa-Calpeoríori.
- Menéndez Pidal, R. (1985). *Manual de gramática histórica española*. Madrid: Espasa-Calpe.
- Michelena, L. (1977). *Fonética Histórica Vasca*. San Sebastián: Imprenta de la diputación de Guipuzcoa.
- Mitchell, M. (2009). *Complexity, A Guided Tour*. New York: Oxford University Press.
- Moreno Fernández, F. (2009). *La lengua española en su geografía*. Madrid: Arco Libros.
- Pardo-Gordó, S. (2015). *La diversidad cultural del primer neolítico en el mediterráneo occidental: un análisis desde los sistemas complejos y la simulación basada en agentes* (tesis doctoral). Valencia: Universidad de València.
- Pardo-Gordó, S. (2020, April 23). "Cultural Spread" (Version 1.2.0). *CoMSES Computational Model Library*. Accesible en: <https://www.comses.net/codebases/4592/releases/1.2.0/>
- Pla Colomer, F. P. (2013). *Reconstrucción de la pronunciación castellana medieval: La voz de los poetas* (Tesis doctoral). Valencia: Universitat de València.
- Princeton University. (2010). About WordNet. *WordNet*. Princeton University.
- Quilis, M. (1996). La F- inicial latina en los orígenes de la lengua española (I). *ASJU*, XXX-2, 385-453.
- Terrado Pablo, J. (1991). *La lengua de Teruel a fines de la Edad Media*, Teruel, Instituto de Estudios Turolenses, Excma. Diputación Provincial de Teruel.