

WILLEM F. H. ADELAAR, PILAR VALENZUELA BISMARCK
Y ROBERTO ZARIQUIEY BIONDI

Editores

ESTUDIOS SOBRE LENGUAS ANDINAS Y AMAZÓNICAS

Homenaje a Rodolfo Cerrón-Palomino



Capítulo 7



FONDO
EDITORIAL

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Estudios sobre lenguas andinas y amazónicas
Homenaje a Rodolfo Cerrón-Palomino

Willem F. H. Adelaar, Pilar Valenzuela Bismarck
y Roberto Zariquiey Biondi

© Willem F. H. Adelaar, Pilar Valenzuela Bismarck
y Roberto Zariquiey Biondi, editores

De esta edición:

© Fondo Editorial de la Pontificia Universidad Católica del Perú, 2011

Av. Universitaria 1801, Lima 32 - Perú

Teléfono: (51 1) 626-2650

Fax: (51 1) 626-2913

feditor@pucp.edu.pe

www.pucp.edu.pe/publicaciones

Cuidado de la edición, diseño de cubierta y diagramación de interiores:

Fondo Editorial PUCP

Ilustración de cubierta: Josué Sánchez Cerrón

Foto de Rodolfo Cerrón-Palomino: Roberto Zariquiey

Primera edición, setiembre de 2011

Tiraje: 500 ejemplares

Prohibida la reproducción de este libro por cualquier medio, total o parcialmente,
sin permiso expreso de los editores

ISBN: 978-9972-42-972-9

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2011-11916

Registro de Proyecto Editorial: 31501361101722

Impreso en Tarea Asociación Gráfica Educativa

Pasaje María Auxiliadora 156, Lima 5, Perú

ENTERRANDO EL ESQUELETO QUECHUMARA¹

Paul Heggarty

*Departamento de Lingüística, Instituto Max Planck de Antropología
Evolutiva, Leipzig, Alemania*

1. Quechumara: historia de un concepto controvertido

A lo largo de la prolongada y destacada carrera del doctor Rodolfo Cerrón-Palomino —y no por coincidencia— la lingüística andina ha dado grandes pasos adelante en lo que él mismo ha denominado «lingüística propiamente dicha» (Cerrón-Palomino, 2000: 51). Esto se aplica sobre todo a nuestra comprensión de las dos principales familias lingüísticas de la región, tal como se las estudió en las obras de consulta fundacionales de cada una de ellas, a saber, *Lingüística quechua* ([1987] 2003) y *Lingüística aimara* (2000), de autoría del mencionado investigador. Para empezar, en lo tocante al dificultoso asunto de los nombres mismos *quechua* y *aimara*, suscribo sin vacilaciones el dictamen de Cerrón-Palomino (1993), considerándolo el argumento más convincente sobre cómo hemos de emplear estos términos de la manera más adecuada: esto es, para denotar familias lingüísticas completas, evitando neologismos innecesarios y de poca utilidad, tales como *aru* y *jaqi*, o el añadir una *-n* final innecesaria en el inglés *Quechuan* o *Aymaran*.

Sin embargo, detrás de todo el avance en los estudios de cada familia, en el transcurso de estas décadas formativas siempre ha asomado el esqueleto en el clóset de la lingüística andina: la hipótesis del «quechumara». Esto es, ha permanecido sin respuesta la pregunta más esencial: ¿cuál es, exactamente, la naturaleza de la relación entre las dos grandes familias lingüísticas que sobreviven en los Andes?

¹ El autor expresa su agradecimiento a Marco Ferrell, Pilar Valenzuela y a Roberto Zariquiey por su valiosa ayuda en la traducción y corrección técnica y estilística de la versión en castellano de este artículo.

Resúmenes invalorable de la historia de este debate son los estudios correspondientes de Cerrón-Palomino (2003: cap. XI) y Cerrón-Palomino (2000: cap. VIII). Además, el autor nos ofrece un reconocido trabajo de consulta sobre el núcleo común de notorios paralelismos entre las dos familias, especialmente entre las variedades sureñas de ambas, en Cerrón-Palomino (1995), volumen recientemente reeditado y ampliado como Cerrón-Palomino (2008). En esta obra, el autor emplea el término «quechumara» simplemente para remitir al conjunto de paralelismos estructurales entre ambas familias. Es precisamente gracias a esta interpretación del término que el autor nos puede proporcionar una invalorable indagación de los datos de tales paralelismos, libre de las presuposiciones inherentes a cualquier interpretación particular de cómo surgieron. Vale decir que, no obstante las aseveraciones de Torero (2002: 159-160), el uso por parte de Cerrón-Palomino del término «quechumara» ni supone ni implica una afirmación de que las dos familias necesariamente compartieran un ancestro común. La insinuación de que sí lo tuvieron es a lo que en este ensayo me refiero con la frase específica «hipótesis del quechumara».

La opinión lingüística sobre esta hipótesis ha sido durante décadas una montaña rusa. Efectivamente, el asunto se ha revelado tan espinoso que el mismo Cerrón-Palomino se ha abstenido de aventurar cualquier pronunciamiento especulativo o definitivo sobre el tema. Es verdad que Torero (2002: 154), por su parte, se muestra taxativo al descartar que cualquier concepto «quechumara» pueda tener algún sentido, por lo menos en fechas suficientemente recientes para ser pertinente al entendimiento de la prehistoria andina. No obstante, han permanecido suficientes dudas para que varios destacados lingüistas de la región sigan mostrándose reticentes a cerrar las puertas a la hipótesis de manera contundente. Es la línea de Cerrón-Palomino, entonces, la que tiene más adhesiones: Adelaar con Muysken (2004: 35), por ejemplo, observan que la hipótesis aún se admite «no obstante la falta de pruebas». Por cierto, Campbell (1995: 195), desmintiendo su inmerecida reputación de archiescéptico de cualquier tipo de relaciones distantes que se haya propuesto entre las familias lingüísticas, se expresa más bien a favor de la hipótesis antes que en contra de ella, aunque de manera cautelosa: «pistas sugestivas, pero nada concluyente». Mientras, parece que entre algunos distinguidos estudiosos que no son especializados en lenguas andinas, el parentesco quechua-aimara se da aún por sentado: la amplia pincelada de Diamond y Bellwood (2003: Figura 3), por ejemplo, simplemente supone la existencia de una sola entidad denominada «quechumara».

He buscado aportar a este debate con mis propios datos, análisis y argumentos, expuestos en detalle en Heggarty (2005), con sumarios en inglés en McMahon, Heggarty *et al.* (2005) y McMahon y McMahon (2005: 156-157, 166-173).

No hay espacio más adecuado que este para reiterar mis agradecimientos a los cuatro distinguidos autores cuyos comentarios se publicaron junto a Heggarty (2005), entre ellos el del mismo Rodolfo Cerrón-Palomino. Dichos autores pusieron sobre el tapete varios asuntos importantes acerca de mi análisis y tomo la oportunidad que me ofrece este volumen para responder, llevando mi metodología unos significativos pasos más adelante.

El patrón básico detectado en Heggarty (2005) fue una correspondencia significativa entre los lexemas quechuas y aimaras que representan significados *menos* estables, mientras que en el correspondiente conjunto de significados *más* estables la señal resultó ser mucho más débil. Argumenté en el sentido de que mis datos y análisis apuntaban así a que el quechua y el aimara no habían compartido un origen común; es decir, iban en contra de la hipótesis del quechumara. Uno de los autores que aportó su comentario —así como otros que lo han hecho informalmente— ha sugerido que el contraste que expuse no puede excluir un escenario alternativo que sí incorpore un profundo origen común, seguido de una ola de préstamos en épocas más recientes. En este ensayo me centro precisamente en dicha objeción, proponiendo un análisis más refinado para reforzar mi posición. Si logro convencer a mis colegas andinistas, quizá podamos avanzar finalmente hacia el «cierre» de este tema y podamos, así, por fin poner a descansar el «esqueleto» del quechumara.

Si fuera así, el interés y significancia potenciales de este estudio no se limitarían solo a la lingüística andina. En primer lugar, beneficiarían a los que estudian la prehistoria de la cuna andina de la civilización humana; ya que el presente trabajo nos permitiría formular enunciados más firmes acerca de los orígenes de las poblaciones que hablaban las respectivas protolenguas. Tendrían, pues, orígenes distintos, aun cuando pasaron por largos períodos de contacto muy intenso (véase por ejemplo Heggarty y Beresford-Jones, 2010). Y en segundo lugar, este trabajo beneficiaría a los lingüistas históricos en general, ya que el ejemplo del quechumara podría ilustrar cómo el método aquí propuesto podría aplicarse a casos análogos en los que los métodos existentes no han logrado todavía diagnosticar de manera concluyente si determinadas lenguas están o no (distantemente) emparentadas. El enigma del quechumara constituye, ciertamente, un caso ilustrativo ideal de este tipo de hueso tan duro de roer para la lingüística histórica. Los lectores interesados en explorar estas repercusiones metodológicas más generales podrían con provecho leer este trabajo juntamente con Heggarty (2010) en el volumen de las Actas de la conferencia por el centenario del nacimiento de Morris Swadesh.

2. Medición de la diferencia en semántica léxica

A primera vista podría parecer que la metodología de Heggarty (2005) se puede comparar con la que originalmente desarrolló Swadesh, esto es, la *lexicoestadística*. No obstante ello, a poco de iniciar la lectura se notará cuánto se aleja mi técnica de la metodología lexicoestadística tradicional. Además de desarrollar mediciones más refinadas del grado de diferencia entre las lenguas en su semántica léxica, también esclarezco —y abandono— varios de los principios y suposiciones de la lexicoestadística tradicional. Tanto mejor, pues ese método por largo tiempo ha adolecido de múltiples ambigüedades y confusiones fundamentales. Como estas son tratadas por separado en Heggarty (2010), me limito aquí a lo más pertinente para los fines de este ensayo.

2.1 *Medición del grado de correspondencia, no diagnóstico del parentesco*

Una de estas confusiones es la suposición —a menudo tácita— por parte de varios de los primeros practicantes de la lexicoestadística, de que los ítems del vocabulario «nuclear» de ninguna manera pueden ser prestados entre las lenguas. En realidad, incluso la misma versión referencial de la lista de Swadesh, en inglés, presenta múltiples préstamos escandinavos y franconormandos. Dicha suposición, pues, resulta demostrablemente errónea, sin que por ello hayan faltado varias afirmaciones en tal sentido. Incluso en el caso del «quechumara», ciertos autores han seguido la supuesta lógica de que si se hallaban algunas correspondencias entre una lista quechua y otra aimara, entonces, puesto que «no podían» deberse a préstamos, solo se las podía explicar como herencia de un remoto origen común; y así el método estaba en condiciones de «demostrar» esa relación. Fue precisamente esta falacia la que Büttner (1983) buscó aplicar al tema del quechumara. Invocó el número de «correspondencias» entre sus listas de Swadesh para el quechua y el aimara como si eso constituyera una prueba de que las dos lenguas estaban distantemente emparentadas al nivel de «tronco». Su estudio sirve, más bien, de advertencia del fracaso que espera a los trabajos que parten de un supuesto fundamentalmente equivocado. Los lingüistas de la rama andinística, entre ellos Cerrón-Palomino (2003: 371-373) y Torero (2002: 146-150), rechazan rotundamente el estudio de Büttner.

2.2 *No «cognados versus préstamos», sino más bien «correlatos»*

Pasamos ahora a otro asunto, aunque relacionado con el precedente, en el cual se observan inconsistencias y confusiones en los intentos de parte de diferentes estudios por aplicar la lexicoestadística: ¿qué hacer cuando estamos ante posibles

préstamos? ¿Cómo debería uno lidiar con las aparentes «correspondencias» en el conjunto de datos que no son verdaderos cognados sino palabras que —se sabe o por lo menos se sospecha— son préstamos?

Para seguir correctamente la lexicoestadística tradicional, el primer paso debe consistir en depurar el conjunto de datos de todo préstamo conocido, tal como lo intentan Dyen, Kruskal y Black (1992), por ejemplo, en su famoso conjunto de datos sobre el indoeuropeo. Sin embargo, incluso para la más investigada de todas las familias lingüísticas, no les fue posible lograr esta «depuración» de manera completamente exitosa (véase Heggarty, 2010). Con mayor razón, en el caso del quechua y del aimara, apenas si podemos considerar tal tarea viable.

Incluso en el vocabulario nuclear, el quechua y el aimara comparten algunas palabras idénticas, como la que significa ‘fuego’, que se presenta como /nina/ en ambas familias. Otras palabras muestran por su parte correspondencias inconfundibles, como el aimara sureño /quta/ ‘lago’ y /tunka/ ‘diez’, que han de cotejarse con las formas del quechua central /qut̪sa/ y /t̪sunka/. Que estas y muchas otras formas lingüísticas demuestran alguna forma de «correlación» es algo en lo que ambos bandos del debate sobre el quechumara están plenamente de acuerdo. En lo que sí difieren es en cómo interpretan y explican muchas de las correlaciones: ya sea como cognados supervivientes de un origen quechumara común, o como préstamos entre dos familias lingüísticas no emparentadas. En una u otra explicación, las justificaciones que se aducen a menudo acusan una lógica circular, fundada en una presuposición por parte del autor acerca de si las familias están emparentadas o no.

Habida cuenta de las disputas al interpretar lo que puede o no puede ser un préstamo, la misma tarea de «depurar» los datos para excluir tales préstamos presupone —quiérase o no— la existencia de una lengua ancestral común, el hipotético quechumara. Vale decir, se prejuzga precisamente acerca de aquel tema de investigación que tratamos de enfrentar, sobre el cual el método más bien debe evitar cualquier presuposición *a priori* en uno u otro sentido. Por lo tanto, para que ese asunto se pueda investigar, no hay más remedio que abandonar el principio de la lexicoestadística tradicional que requiere que los préstamos sean identificados y excluidos. (Recordemos, además, que en cualquier caso esta filtración no se requería sino para una tarea muy diferente: no para establecer el hecho del parentesco, sino para cuantificar el *grado* de un parentesco ya conocido.)

El método que expongo aquí asume, entonces, una orientación del todo diferente, que evita completamente los recargados juicios caso por caso que se requieren si se intenta detectar y excluir los préstamos, pero no los verdaderos cognados. Es decir, se logra evitar el problema porque simplemente desistimos de cualquier intento de discriminarlos. A primera vista, podría parecer que tal

decisión tiene consecuencias serias, pero, como veremos, es imprescindible si queremos lograr una perspectiva más amplia que pueda proveernos de un nuevo instrumento diagnosticador, libre de juicios *a priori*, para establecer si determinadas lenguas están emparentados o no. Ciertamente, en casos tales como el que nos presentan el quechua y el aimara, el enfoque que se acaba de proponer es el único viable. Nos permite hacer uso provechoso de lo que sí se conoce con seguridad acerca de muchas formas lingüísticas —que claramente tienen alguna correlación entre las dos familias—, incluso cuando los datos y las investigaciones existentes no son concluyentes o son insuficientes para confirmar si esta correlación se explica mejor por el contacto o por el origen común. Así, en la base de datos de la que doy cuenta aquí, a las formas lingüísticas quechuas y aimaras que muestran claros paralelismos de sonido-significado se las califica simplemente de correlatos, empleando este término de una forma muy amplia, para aludir a cualquier tipo de correspondencia, sea por origen común o por préstamo. En otras palabras, en lo que toca al quechua y al aimara, mi base de datos no diferencia los conceptos de *cognado* y *préstamo*: ambos no son sino correlatos. En el mismo sentido, hablaremos igualmente del grado de correlación entre dos lenguas en su vocabulario básico.

Un tema aparte es el de evaluar qué palabras se han de considerar correlatos o no. En la gran mayoría de los casos hay poca discusión sobre la identificación de cuáles formas son correlatos entre el quechua y el aimara —antes de explicar *por qué* lo son— visto cuanto se parecen en el nivel fonético (por ejemplo, /nina/ o /quta/ versus /quṭsa/). Los desacuerdos se limitan a un pequeño número de palabras en las que las correspondencias de sonido son mucho menos directas, tales como las que significan ‘ojo’: quechua *ñawi* y aimara *ayra* (véase Cerrón-Palomino, 2000: 311-2). Algunos investigadores han considerado correlatos a unos cuantos de estos pares, mientras otros estudiosos descartan dicho análisis, viendo en ellos correspondencias simplemente casuales en uno o dos sonidos. Para tratar de reducir el inevitable subjetivismo en este punto, en Heggarty (2005) recurrí a los conocimientos existentes de la fonotáctica del quechua y del aimara y a la historia de sus cambios fonéticos, con el fin de al menos ordenar las alegadas correlaciones en una escala de «niveles de plausibilidad», y efectué análisis separados en cada nivel. En cualquier caso, ya que se encontró poca diferencia en los resultados globales entre los varios niveles de plausibilidad, surgió cierta confianza en cuanto a los riesgos de la subjetividad —que eran menos serios de lo que se podría temer—, lo que —se presume— refleja cuán pocos son los casos en disputa en comparación con los casos sobre los que hay consenso. Como en Heggarty (2005), en lo que sigue uso nuevamente los datos calculados para un nivel de plausibilidad intermedio.

2.3 *¿Emparentados o no? Cuando el vocabulario nuclear por sí solo no es suficiente para determinar el parentesco*

Varios trabajos recientes se han basado en datos de tipo lexicoestadístico con miras a diagnosticar el parentesco lingüístico, aunque centrándose casi exclusivamente en refinar la lista de referencia hasta llegar a un presunto «santo grial» de un núcleo mínimo de los significados más estables. Sin embargo, como se señala en Heggarty (2010), esto no necesariamente resuelve la cuestión, ya que cualquiera sea el número de los significados de la lista mínima, persiste el problema fundamental. ¿Exactamente cuántas coincidencias en este núcleo (o qué proporción de él) constituyen un diagnóstico contundente de parentesco y divergencia, antes que de convergencia?

En el caso del quechua-aimara, ciertas variedades de cada familia muestran no menos de trece «correspondencias» aparentes entre los cuarenta significados de la lista nuclear de Holman *et al.* (2008), por ejemplo. ¿Pero cómo hemos de interpretar tal proporción? ¿Es suficiente por sí sola para establecer el parentesco entre estas familias lingüísticas, a pesar de las muchas evidencias en el sentido contrario y de las opiniones de varios especialistas en estas familias? Desafortunadamente, el solo mirar listas de cuarenta significados e incluso el hacer cálculos de su grado de correlación dejan sin resolver la cuestión quechumara. No nos acercamos en nada al esclarecimiento del enigma.

La proporción exacta de correlatos no es síntoma confiable de que exista o no parentesco, pues en ambos escenarios —esto es, sean dos lenguas cualquiera emparentadas o no— la proporción exacta es el resultado más bien de otros factores. Entre dos lenguas que sí están emparentadas, la proporción de correlatos resulta en primer lugar del lapso transcurrido desde que estas se separaron. Entre dos lenguas no emparentadas, la proporción se da en función de cuán intensos y duraderos han sido los contactos entre ellas, y en qué etapas de su evolución ocurrieron. Son estos factores del «mundo real» —que o separan a las poblaciones que hablan las dos lenguas en cuestión o las ponen en contacto— los que determinan cuántos cognados originales sobreviven o cuántos préstamos se intercambian, incluso al interior de la lista nuclear. Pero estamos aquí frente a las incógnitas de la prehistoria andina, y un simple conteo de correlatos en cualquier lista nuclear no nos acerca en nada a la determinación de cuál de los dos escenarios es el que mejor explica el resultado lingüístico que hoy en día observamos.

Que el controvertido concepto del quechumara continúe sin resolverse se debe a que representa precisamente un caso del tipo más difícil: estamos enfrentados a cierto número de correlatos entre las listas del léxico nuclear de ambas lenguas, pero no podemos resolver las disputas acerca de cuál de los dos escenarios diametralmente

opuestos da mejor cuenta de ellos. Desde hace mucho tiempo tenemos disponibles listas «Swadesh» para el quechua y el aimara, y hasta para varias de las lenguas o dialectos de cada familia. Pero las proporciones de correlatos (20% a 30%) son tales que tanto los que abogan por la hipótesis quechumara como los que la rechazan han encontrado en ellas «evidencias» a favor de sus posiciones. ¿Serán estos correlatos los últimos *cognados* obstinados que sobreviven de un remoto origen común? ¿Serán testimonio del proceso opuesto de convergencia, es decir, préstamos que surgen de contactos tan intensos o de tan larga duración que llegan a entrar incluso en el vocabulario nuclear? ¿O serán más bien una mezcla de ambos: divergencia temprana y luego una etapa posterior de convergencia?

3. Cómo extraer de una lista de vocabulario básico más datos sobre el parentesco

Dado que un simple conteo de cuántas formas son correlatos en una lista de léxico nuclear no alcanza a zanjar el debate, necesitamos más datos que nos puedan proporcionar un cuadro más claro. Afortunadamente, incluso basándonos en las mismas listas del léxico nuclear ya existentes, efectivamente *hay* maneras de extraer de ellas una señal diagnosticadora mucho más precisa de lo que es parentesco frente a lo que es contacto. Hay dos niveles en los que esto puede hacerse, expuestos en §3.1 y §3.2, líneas abajo.

3.1 No «sí o no», sino mediciones de gradación

En primer lugar, podemos aumentar la sensibilidad de las mediciones que se han empleado. Una posibilidad sería recurrir a listas de más significados, aunque intentos como el de Embleton (1986: 92-93) sugieren que más allá de aproximadamente doscientos significados hay en realidad poca ganancia en este sentido. Un enfoque alternativo sería no buscar más significados sino extraer de cada uno de ellos una medición más detallada y refinada. Ciertamente, es este aspecto en donde la lexicoestadística ha logrado en menor medida hacer frente a la tormenta de críticas. El método tradicional impone una caracterización de «todo o nada», cuando en realidad las relaciones entre los léxicos de dos lenguas a menudo no son tan absolutas, sino un asunto de *gradación*. Para cualquier significado, las «medidas» de la lexicoestadística se limitan a valores de apenas 1 (cognados) o 0 (no cognados). Se trata de una camisa de fuerza poco adecuada para muchos ejemplos reales, según las consabidas objeciones en cada uno de tres niveles: sinónimos múltiples (por ej. *small/little* en inglés); supervivencias de cognados en significados ligeramente diferentes pero todavía estrechamente relacionados

(alemán *hund* frente a inglés *dog/hound*); y una correlación entre algunos pero no todos los morfemas en palabras morfológicamente complejas (francés *cœur* pero español *cor-azón*). Entonces, para aumentar la sensibilidad de las valoraciones de correlatos, mi método se aparta significativamente de la lexicoestadística al acomodar y calcular medidas de traslapamiento o divergencia en el léxico que pueden ser más sensitivas que solo 0 o 1, en cada uno de los tres niveles. En Heggarty (2005) se expone en detalle cómo se calculan estos valores intermedios, con ejemplos específicos tomados de las lenguas andinas aquí tocadas, en tanto que en <<http://www.quechua.org.uk/cpv>> se puede consultar los datos completos.

3.2 *No conteos generales de correlatos, sino patrones de correlación*

Ciertamente, si podemos lograr cifras de mayor precisión por significado, estas serán, en principio, valiosas. Sin embargo, para nuestros fines mucho del beneficio se pierde si con estos nuevos cálculos, ahora más detallados para cada significado individual, no hacemos sino volver a confundirlos en un solo conteo general de grado de correlación. Esto se debe a que, como hemos visto, una misma proporción global puede resultar o bien de un parentesco lejano, o bien de un proceso de convergencia. Entonces, el simplemente calcular con más finura ese mismo conteo general no añade ninguna precisión que sea directamente informativa sobre ese asunto. Para ello, necesitamos dirigir la mirada también a un segundo nivel en el cual se pueda extraer de nuestro conjunto de datos más señales «diagnosticadoras» de parentesco o de convergencia.

Afortunadamente, tales señales están disponibles desde ya en el *detalle* de los mismos datos individuales que se sumaron para calcular el conteo general de correlación. Porque los correlatos no necesariamente se distribuyen de manera uniforme a través de la lista; y es justamente en el detalle de su distribución que se puede encontrar un mejor diagnóstico. Así, antes que fusionar todos estos datos en una suma total del número de correlatos a través de una sola y monolítica lista «nuclear», sirve más bien retener estos datos más precisos e incluirlos individualmente en nuestro análisis.

Sin embargo, para que puedan surgir de entre tales datos patrones de distribución diferenciados el uno del otro, necesitamos primero identificar algún criterio lingüísticamente significativo que nos permita distinguir entre los varios datos individuales que conforman nuestra base de datos. Además, debe ser un criterio que pueda darnos una señal clara de una u otra de las posibles explicaciones para la existencia de correlatos: o divergencia a partir de un parentesco original, o convergencia por contacto. En el caso de las listas léxicas, un criterio clave es

el de la «estabilidad» relativa de los diferentes significados, el mismo concepto empleado por Heggarty (2005), pero llevado aquí a otro nivel.

3.3 Significados más estables versus significados menos estables

Muchos de los trabajos recientes que buscan aprovecharse de los datos de tipo lexicoestadístico con el propósito de establecer el parentesco se basan en un hecho evidente y ahora ampliamente reconocido: que los varios significados que conforman las listas «Swadesh» no son, de ninguna manera, igualmente estables a lo largo del tiempo. En el mismo conjunto de lenguas y el mismo lapso, los lexemas que se emplean para expresar significados como *sucio* o *golpear* son reemplazados con una frecuencia mucho mayor que los lexemas que expresan los significados *dos* o *morir*. Dyen, Kruskal y Black (1992), por ejemplo, calculan cuántos reemplazos se observan a través de las 95 lenguas de su base de datos del indoeuropeo, en cada uno de los doscientos significados de la lista de Swadesh. Estas «tasas de reemplazo», tal como aparecen listadas en su archivo IE-RATE1.TXT, expresan el número promedio de reemplazos en este conjunto de datos en una unidad de tiempo que los autores estiman en aproximadamente 2.400 años. Las tasas varían enormemente: de apenas 0,01 reemplazos en significados como el número *dos* (*two*), hasta 3,61 en significados como el de *carretera* (*road*), el significado más inestable en la lista de Swadesh de cien significados, y 6,23 en *sucio* (*dirty*), el más inestable en su lista de doscientos significados.

Estos conteos de Dyen, Kruskal y Black (1992) representan apenas uno de entre varios enfoques alternativos que buscan proporcionar alguna cuantificación de la estabilidad inherente de cada significado. McMahon y McMahon (2005: 105-110) siguen una idea de Lohr (1999), quien tomó datos de cuatro familias de lenguas independientes —el indoeuropeo, el afroasiático, el austronesio y el sinotibetano— para clasificar cada uno de quinientos significados básicos según sus propios criterios de estabilidad: «propensión a ser retenido» y «reconstructibilidad». Holman et al. (2008) se basaron en criterios diferentes para proporcionar un nuevo ordenamiento alternativo, del cual también nos aprovecharemos aquí, aplicándolo a datos sobre las lenguas andinas.

4. Buscando patrones en cálculos de correlación

(A): Sublistas de significados más estables y menos estables

Cualquiera sea el ordenamiento que se emplee, lo importante es que nos permita identificar, en cualquier lista de significados, los subconjuntos de significados más estables y menos estables que la conforman. Esta diferencia nos permite ir

un paso más allá de la incertidumbre de algún «número mágico» de correlatos, arbitrario y subjetivo, al cual se le considera un «umbral» diagnosticador de parentesco. Más informativo que cualquier cifra sería un contraste entre el número de correlatos hallados por un lado entre los significados más estables, y por otro lado, entre los menos estables. Aunque bastante simple, este puede constituir un primer tipo de patrón discernible al interior del conjunto global de datos: es decir, un patrón de *contraste*.

4.1 Base de datos y subconjuntos

En Heggarty (2005) apliqué este enfoque a una nueva base de datos sobre lenguas andinas (recolectada en mis propios trabajos de campo), la misma que ahora vuelvo a emplear. La diferencia es que en el presente artículo expongo un nuevo método más refinado de analizar esos mismos datos, a fin de extraer de ellos una señal que sea aún más clara y «diagnosticadora» de si determinadas familias lingüísticas están emparentadas o no.

Puesto que ya en Heggarty (2005) y en <<http://www.quechua.org.uk/cpv>>, se dan los detalles completos de esta base de datos y de las tasas de correlación de ella derivadas, me limito aquí a recapitular brevemente lo esencial. Mi base de datos cubre quince variedades regionales de quechua y cinco de aimara. Además, gracias a que el mismo Rodolfo Cerrón-Palomino generosamente me permitió el acceso a sus abundantes materiales no publicados sobre esa lengua, pude incluir también el chipaya. La lista de referencia comprende 150 significados, que se basan principalmente en la lista de Swadesh de 200 significados, aunque con considerables modificaciones para adaptarla al contexto andino y para asegurar un contraste más nítido entre significados más estables y menos estables.

En la lexicoestadística tradicional, los resultados para todos los significados individuales simplemente se combinan en una cifra promedio para la lista entera (es decir, de 150 significados, en el caso de mi base de datos).

En Heggarty (2005), por el contrario, eché mano de las evaluaciones de estabilidad de significados de Lohr (1999) para aislar, entre esos 150 significados, los subconjuntos de los 42 más estables y de los 32 menos estables —los que encajan en los criterios de Lohr de estabilidad máxima y mínima—.

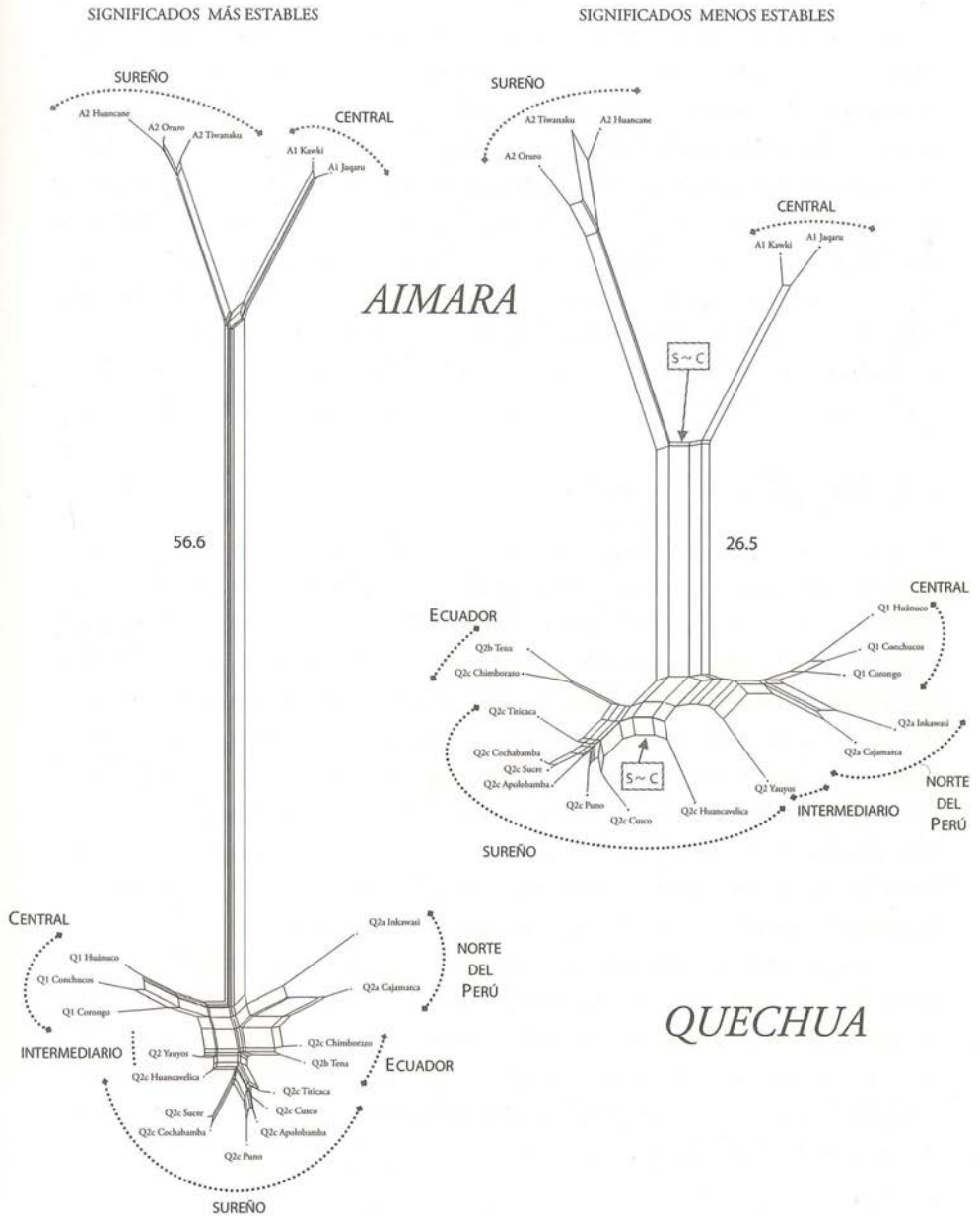
Para cada sublista se calculó separadamente el grado promedio de correlación entre dos lenguas, expresado mediante una cifra entre 1 (correlatos completos en todos los significados) y 0 (ningún correlato, ni siquiera parcial; véase §3.1 líneas arriba). Este proceso se repitió para todas las pares de lenguas, hasta obtener una matriz de tales cuantificaciones del grado de correlación para cada uno de los 191 pares que conforman las 20 lenguas estudiadas (la primera lengua con las

19 demás, más la segunda con las 18 restantes, y así en adelante). Poder evaluar tales datos resulta un tanto difícil con tan solo una mirada de conjunto a esta masa de cifras, pero se trata del formato adecuado para cierto tipo de algoritmo de análisis filogenético, que precisamente requiere de esta matriz de cálculos de distancia (véase Heggarty et al., 2010). (Aquí «distancia» no es otra cosa que el valor inverso, 1 menos x , de un dado grado de correlación x .) Estos algoritmos son capaces de sintetizar tales resultados de una manera objetiva y equilibrada, y de visualizarlos en formas más fáciles de aprehender: es decir, como árboles genealógicos o redes de conexiones. La figura 1 muestra el gráfico de salida de uno de tales algoritmos, a saber el *NeighborNet*, desarrollado por Bryant y Moulton (2004) y ahora integrado en el paquete SplitsTree 4 (Huson y Bryant, 2006). (*NeighborNet* va ganando popularidad en la Lingüística histórica, aunque recomendamos ver Heggarty et al. (2010) para conocer algunas de sus limitaciones.) El *NeighborNet* que está a la izquierda de la figura 1 se ha calculado a partir de la sublista de significados más estables; mientras que el que está a la derecha, a partir de la sublista de significados menos estables.

4.2 Inferencias por defecto: el quechua y el aimara no están emparentados

La característica más elocuente del contraste que emerge entre los dos subconjuntos de estabilidad es que resulta repetidamente consistente en diversos aspectos. Primero, vale la pena fijarse en cómo el quechua y el aimara se comportan individualmente, como familias lingüísticas. Conforme uno pasa de los significados menos estables a los más estables, cada familia se «contrae» sobre sí misma, ya que las distancias entre las lenguas que son miembros de la misma familia se reducen. Vale decir, los resultados corresponden a lo que encontramos en lenguas que sabemos pertenecen a una misma familia, exactamente tal como se esperaría. Pero si ahora preguntamos cómo cambian las dos familias, la una respecto de la otra, descubrimos que se comportan de manera exactamente contraria, alejándose mucho más. Esto por sí solo ya sugiere que el quechua y el aimara no actúan como lenguas distantemente emparentadas, sino lo contrario. Los detalles de la figura 1 ofrecen otros indicios más, pues en los resultados de la sublista menos estable surge una reticulación (la gran «caja» rectangular). Ciertas variedades de cada familia se encuentran «jaladas» en cierto grado las unas hacia las otras, alejadas así de los demás miembros de la familia que integran. Esta reticulación se ha identificado en la figura 1 como S~C, esto es, sureño~central, puesto que coloca a las variedades *sureñas* del aimara al lado de las del quechua *sureño* (ambas a la izquierda), apartadas así de las correspondientes variedades *centrales* de cada familia (ambas a la derecha):

Gráfico 1. NeighborNets de cálculos de correlatos entre veinte variedades de quechua y de aimara, en subconjuntos de significados más estables y menos estables



Ciertamente, el parentesco lingüístico (por un lado, entre todas las variedades del quechua y, por otro lado, entre las del aimara) permanece como la señal más fuerte —el rectángulo es más alto que ancho—, pero su ancho bien podría verse como el «segundo componente principal» del patrón general. Y este también, tal como el primero, está lejos de ser producto del azar; solo que esta vez la señal no coincide con el parentesco sino con otra realidad —la proximidad geográfica. El candidato obvio como proceso (pre)histórico que pueda explicar este *reacercamiento* entre las variedades contiguas de cada familia es aquella otra gran fuerza determinante de la lingüística histórica: la convergencia por medio del contacto. Además, esta señal muy clara en los significados menos estables simplemente desaparece en los más estables, es decir, los menos sujetos a préstamo. (Para un análisis de la naturaleza del vínculo entre la (in)estabilidad de determinados significados y su propensión a ser prestados, véase Tadmor et al., 2010.) Esto vale como apoyo adicional a la conclusión de que las dos familias no están emparentadas, y las correspondencias entre ellas se explican mejor mediante el contacto.

4.3 ¿Una explicación alternativa?

Lo previamente expuesto parece ser la interpretación más directa de los múltiples patrones que se disciernen en el gráfico 1. Sin embargo, como lo destaca Adelaar en su comentario a Heggarty (2005), es posible concebir por lo menos una explicación alternativa. ¿No podrían ser las proporciones de correlatos en los diferentes subconjuntos señales de etapas separadas de las historias de estas lenguas? ¿No podrían los correlatos que se encuentran en el subconjunto estable representar justamente los últimos cognados sobrevivientes de un remoto origen común? Mientras tanto, los correlatos del conjunto menos estable serían préstamos intercambiados en un episodio de contacto mucho más reciente, tan intenso que ha traído a este subconjunto tantos préstamos que ahora son más numerosos que los pocos cognados que subsisten en el subconjunto más estable.

Aunque no es la explicación más económica, es cierto que el solo contraste entre dos sublistas —la más estable frente a la menos estable— no la puede excluir. ¿De qué otra manera, entonces, podría uno discernir una señal más fina que pueda ayudar a diferenciar estos dos escenarios posibles? En lo que sigue, propongo una manera más finamente discriminante de analizar estos mismos datos, con el fin de descubrir tal señal.

5. Buscando patrones en cálculos de correlación (B): Significado por significado

5.1 No contrastes binarios de estabilidad sino gradientes progresivas

Ya hemos adelantado algo de lo que se quiere presentar en esta sección: en vez de prestarle atención solamente a un conteo global —y de baja «resolución»— del número de correlatos en una sola lista general, hemos llegado a un diagnóstico más detallado de parentesco que radica en el *contraste* entre dos subconjuntos, es decir, los extremos más y menos estables de esa lista. Pero podemos refinar más aún nuestra perspectiva sobre estos datos si continuamos por la misma ruta. Antes que solo un contraste binario entre dos subconjuntos opuestos al máximo, más sensitivo y detallado sería una comparación continua en una secuencia de muchos subconjuntos de significados, ordenados de manera más precisa por su grado de estabilidad, siguiendo lo que (por supuesto) es una *gradiente* de lo más a lo menos estable. Como veremos líneas abajo, por medio de una comparación continua como esta, proyectada a lo largo de lo que llamaré una «gradiente de estabilidad», se puede lograr una representación gráfica muy útil para diagnosticar si la explicación que la subyace es con más probabilidad el parentesco o la convergencia.

5.2 Primeras impresiones: patrones en los números del 1 al 6

Empiezo con un ejemplo tan pequeño en escala, que solo es de valor ilustrativo; pero por lo tanto más claro aún, pues los datos consisten nada más que en el conjunto de números del 1 al 6. La figura 1 muestra transcripciones fonéticas de dos lenguas muestra de la familia quechua, y dos de la familia aimara. (Estas se toman de mi base de datos de registros de unas veinte lenguas andinas, que se pueden escuchar en: <<http://www.quechua.org.uk/sounds>>.) Aunque son instantáneamente reconocibles para quien está familiarizado con lenguas andinas, para los fines que nos proponemos es útil *no* identificar todavía las variedades, para centrarnos en los patrones discernibles en los datos.

Figura 1. Los números del 1 al 6 en dos lenguas muestra de la familia quechua, y dos de la familia aimara

NÚMERO →	1	2	3	4	5	6
Lengua A	hʊk	ɪʃke	kime	tʃusku	pitsqe	hoqte
Lengua B	ux	jskæj	kĩnsɛ	tawa	p ^h ʒzɛra	soqta
Lengua C	mæjɛ	paxɛ	kimsɛ	pufɛ	pitʃqe	sufte
Lengua D	maje	paje	kimsa	pusi	p ^h ʒsqɑ	sɔxtɑ

No se necesita gran conocimiento lingüístico, ni menos en las lenguas andinas, para poder extraer de estas transcripciones la siguiente tabla (figura 2) que resume cuáles de estas formas son correlatos (sean préstamos o cognados) entre determinadas combinaciones de lenguas.

Figura 2. Patrones de correlatos entre los números del 1 al 6 en quechua y aimara

Para este número →	1	2	3	4	5	6
¿Son correlatos las formas en las lenguas A y B?	✓	✓	✓	✗	✓	✓
¿Son correlatos las formas en las lenguas C y D?	✓	✓	✓	✓	✓	✓
¿Son correlatos las formas en los pares de lenguas AB y CD?	✗	✗	✓	✗	✓	✓

A menudo se invoca a los números, igual que a los términos de parentesco (*padre, madre, etcétera*) como evidencia privilegiada de que determinadas lenguas derivan de un origen común, tal como podría suponerse a partir de un análisis superficial del indoeuropeo. Es bien sabido, sin embargo, que de igual manera los números son especialmente susceptibles a complicaciones. En el mismo indoeuropeo, un análisis más detallado del estatus de los números lo demuestra, ya que descubrimos posibles relaciones de contacto con otras familias lingüísticas (Gamkrelidze y Ivanov, 1995). Más aún en otras partes del mundo —y no menos en los Andes y en otros lugares de América del Sur— es clarísimo que se necesita mucha cautela. Numerosas son las lenguas que han tomado prestado de otras su sistema numérico, especialmente lenguas indígenas frente a lenguas europeas sociolingüísticamente dominantes.

Difícilmente se puede afirmar, entonces, que incluso las fuertes correlaciones dentro de cada uno de los pares de lenguas AB y CD sean prueba de parentesco. Con todo y eso, resulta que estas vinculaciones se establecen de forma indubitable mediante una abundancia de otros datos lingüísticos. La lengua A es una forma de quechua central (de una zona cercana a Chavín de Huántar, en el Callejón de Conchucos, Áncash), la lengua B es una forma del quechua sureño (de la región de Cochabamba, Bolivia); mientras que la lengua C es una forma del aimara central (el «cauqui» de Cachuy, en la provincia de Yauyos, departamento de Lima), en tanto que la forma D es del aimara sureño (de la ribera meridional del lago Titicaca, en Bolivia).

Lo que esa abundancia de otros datos lingüísticos aún *no* ha podido resolver de manera completamente satisfactoria, sin embargo, es si AB (quechua) comparte

o no un remoto origen común con CD (aimara). Y aquí incluso el minúsculo conjunto de datos de los números del 1 al 6 está revestido de interés particular. Primero, ponemos a un lado los casos en que o *todo* un sistema de números es heredado o prestado, o ninguna parte de él lo es. Porque los casos de interés mayor son justamente aquellos en los cuales surge algún *patrón* distintivo, en el que una lengua tiene correlatos con solo *parte* del sistema numérico de otra. El quechua y el aimara constituyen un caso en el que empieza a surgir un patrón que es precisamente de este tipo: un contraste entre los números menores y mayores.

El patrón radica en cómo los correlatos no aparecen al azar, sino bastante coherentemente en conformidad con el valor matemático de cada número. Ya que tal distribución solo muy inverosímilmente se podría achacar al azar, reclama una explicación. Una primera línea de pensamiento que surge inmediatamente es que la explicación podría radicar en las conocidas diferencias de estabilidad —o propensión a ser prestado— de los diferentes significados léxicos. Es revelador, entonces, que de entre los números básicos Swadesh eligió incluir todos aquellos del 1 al 5 (más no mayores) en su lista de 200 significados, mientras que en su lista de los 100 «más estables» dejó solamente el 1 y el 2. Gracias a que los números básicos forman un subsistema del vocabulario tan cerrado y altamente estructurado, a partir de tan solo los seis significados de la figura 2, uno ya puede empezar a discernir una primera señal, sorprendentemente clara y consistente, que puede contribuir a nuestro juicio acerca de si determinadas lenguas están emparentadas o no.

Primero, en los casos en los cuales los números que son correlatos entre dos lenguas son *solo* los mayores, sería imprudente ver en ellos alguna evidencia de un origen común —cuando más bien serían préstamos—, pues de ser así, tanto más notoria sería la ausencia de correlatos en los números menores. A la inversa, la falta de correlatos en los números mayores no necesariamente debilita una presunción por defecto de parentesco si los números más bajos —ceranos a cero— sí resultan ser claramente correlatos. Las lenguas indígenas de Incahuasi (provincia de Ferreñafe, departamento de Lambayeque) y Chetilla (provincia y departamento de Cajamarca), en las serranías del norte del Perú, ilustran efectivamente lo que exponemos. Sus números menores —hasta 3 y 5, respectivamente— son correlatos con el quechua, mientras que los mayores lo son con el español. Estos son *patrones* claros y completamente coherentes con el hecho —en este caso, consabido— de que se trata de lenguas de la familia quechua influidas por el español, pero poco coherentes con la idea de que estamos ante lenguas romances influidas por el quechua. Pocos lingüistas andinos verían algo remotamente inusual en tales patrones, por supuesto; pero, por la misma razón, nos incumbe de sobremanera reflexionar sobre las presunciones por defecto que

entrañan los datos correspondientes, por simples que sean, también respecto de la cuestión del quechumara.

Vale la pena subrayar dónde radica precisamente, en este caso ilustrativo, la señal que podría ayudar a diagnosticar lo «emparentado o no». Ciertamente, no es en el comportamiento de ningún número en particular, cosa que está lejos de ser un dato a prueba de fallas. Véase cómo, incluso en el quechua, el número *cuatro* representa un caso insólito, el único número para el cual existen dos raíces diferentes y no explicadas en diferentes regiones: */tawa/ versus */t̥suku/.

El diagnóstico tampoco radica en el número absoluto de correlatos presentes en una lista mínima «nuclear». Es revelador que en casos en que *ningún* número es correlato —o incluso cuando lo son *todos*— es precisamente la falta de una señal diferencial lo que nos deja con más dudas, y reacios a formular siquiera un pronunciamiento por defecto. Para cualquier par dado de lenguas x e y, los totales simples no pueden excluir que la lengua x haya adoptado su sistema numérico en bloque, tomándolo en préstamo: o bien de y, de modo tal que todos los números son correlatos incluso si x e y *no* son emparentadas; o bien de una tercera lengua z, caso en el cual los números no son correlatos aun si x e y sí son emparentadas. Hay innumerables ejemplos, y no en poca medida en América del Sur, donde el español y el portugués son las lenguas donantes preferidas.

Más bien, es precisamente en aquellos casos en los cuales surge un *patrón* de *contraste* entre significados más estables y menos estables que dicho patrón resulta más intrigante y potencialmente valioso. La verdadera señal diagnosticadora de «parentesco o no» radica no tanto en alguna proporción particular de correlatos dentro de un determinado conjunto de significados, sino en un *patrón* diferencial que surge cuando se clasifican estos correlatos en un ordenamiento *progresivo* según la estabilidad de sus significados. Aun en este ejemplo puramente ilustrativo, se logra entrever un patrón tal incluso con apenas seis significados. Compárese cómo en el método comparativo, herramienta de base de la lingüística histórica, la señal más confiable de parentesco no radica en sonidos *idénticos* que se repiten, sino igualmente en *patrones* regulares y repetidos de *diferencias*, indicativos de que ha operado alguna «ley fonética».

6. Diagnóstico del parentesco mediante gradientes de estabilidad

¿Pero cuáles son, para ser más precisos, los diversos patrones que esperaríamos ver surgir a lo largo de la «gradiente de estabilidad» en diferentes escenarios, según si las lenguas que se comparan están o no emparentadas? ¿Y cómo se podrían visualizar tales patrones de la mejor manera?

En cualquier par de lenguas, si se retiene en el análisis los datos detallados, significado por significado, sobre los correlatos que comparten, estos permiten

trazar en un gráfico la proporción de correlatos (en el eje x) en cada una de una serie de subconjuntos de significados de estabilidad decreciente (a lo largo del eje y). En conjunto, esta serie de datos conformaría una *línea de tendencia* continua (tomando por ahora este término en un sentido general, no técnico) del grado de correlación entre estas lenguas, según la estabilidad decreciente de los subconjuntos de significados. En todos los gráficos restantes de este trabajo aparecen ejemplos de tales líneas de tendencia. En tales conjuntos de datos existen diversos patrones de contraste posibles: las diferencias entre ellas se manifiestan en la forma particular que adopta la línea de tendencia en cada caso. Y es en estas diferentes líneas de tendencia que radican las señales que pueden ayudarnos a responder a la vez la pregunta de «sí o no» sobre el parentesco de las dos lenguas comparadas, y la del *grado* de convergencia o divergencia entre ellas, según veremos en §6.2.

Este enfoque tiene ventajas significativas. En primer lugar, mantiene, en lugar de descartar, los datos que no encajan ni en la sublista más estable ni en la menos estable, sino entre ellas; esto es, los datos en los significados de estabilidad intermedia. Aprovecha plena e igualmente todos los significados, para que todos contribuyan a darle un mayor detalle al *patrón*, en forma de línea de tendencia, que nos dará el diagnóstico más claro de lo que está emparentado o no. En segundo lugar, nos permite encapsular esa señal diagnósticadora en un solo diagrama, en vez de dos representaciones diferentes que han de verse una al lado de la otra, como en el gráfico 1, presentado anteriormente.

6.1 Parentesco versus convergencia: cuatro posibilidades

Antes de echar una mirada a las líneas de tendencia que se presentan en el caso andino, esta sección resumirá las formas generales de líneas de tendencia que se preve surjan de cuatro escenarios hipotéticos posibles en las historias de las lenguas. Naturalmente, cada línea de tendencia corresponde a un *par* de lenguas comparadas la una con la otra. Los escenarios históricos se definen como la matriz de dos variables: en primer lugar, las dos lenguas o tienen un origen común, o no; en segundo lugar, el contacto e intercambio de préstamos entre ellas ha sido o débil (y hasta nulo), o intenso. Estas posibilidades forman las previsiones idealizadas frente a las cuales avanzo en la sección §7 a los datos concretos del quechua y del aimara, para evaluar si responden a alguno de estos escenarios hipotéticos; y si es así, a cuál.

Cada uno de los cuatro escenarios básicos concebidos tiene su correspondiente línea de tendencia y las distintas líneas de tendencia son contrastadas en la figura 3, más abajo. Estas expectativas hipotéticas se basan en:

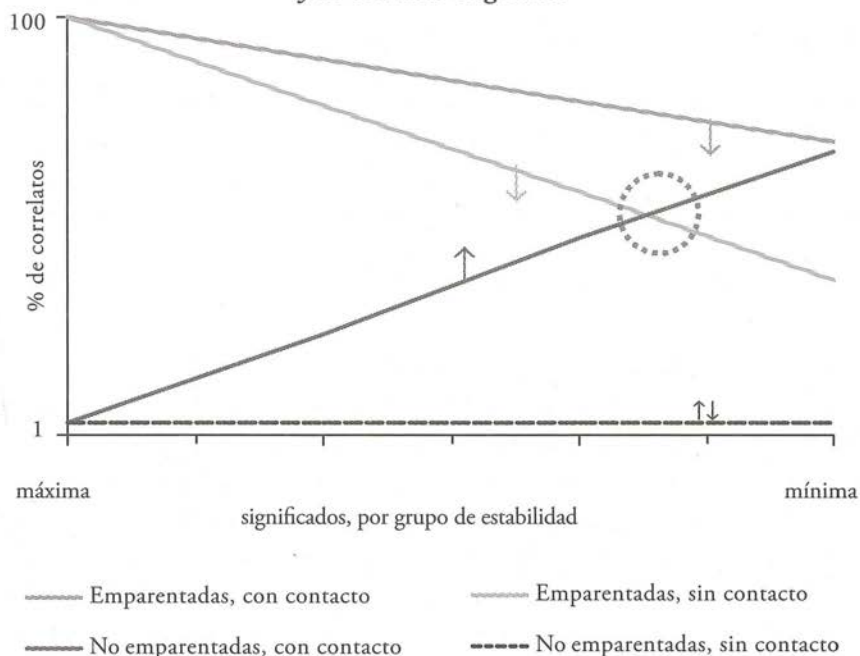
- El concepto mismo de la estabilidad de un significado, el cual por definición quiere decir *a lo largo del tiempo*.
- El hecho de que la estabilidad refleja en gran parte cuán probable es que un préstamo reemplace a la palabra nativa en un significado dado (ver Tadmor et al., 2010).
- Las direcciones opuestas en que el grado de correlación entre las lenguas puede cambiar a lo largo del tiempo, bien divergiendo de un origen común, bien convergiendo por contacto.

Cada línea de tendencia va acompañada por una flecha hacia arriba \uparrow o hacia abajo \downarrow , para indicar la dirección en que se preve que la línea se mueva a lo largo del tiempo. En lenguas emparentadas que divergen desde un ancestro común, el punto de partida es una línea horizontal que roza el límite superior del gráfico, en el valor más alto del eje y de 100% de correlatos (en este caso, cognados originales). Una línea descendente (de izquierda a derecha) corresponde a la pérdida gradual de esos correlatos, más entre los significados menos estables (a la derecha) que entre los más estables (a la izquierda). A la inversa, en el caso de lenguas enteramente no emparentadas que nunca entraron en contacto, en una primera etapa la línea es plana, rayando el eje x, en el nivel de 0% de correlatos —o con ocasionales subidas de muy bajo perfil en el caso de que alguna semejanza fortuita se tome como correlato—. Una flecha hacia arriba representa una adquisición gradual de correlatos (en este caso, préstamos), según son intercambiados por las lenguas que constituyen el par (sea cualquiera de ellas la lengua donante).

Inherente a la definición de estabilidad (en el tiempo) es que las pérdidas y las ganancias ocurran más rápidamente entre los significados menos estables que entre los más estables. Por eso, el movimiento en cualquier dirección será más pronunciado en el extremo menos estable a la derecha del eje y, que en el extremo más estable a la izquierda. Así, de bajada o de subida, las gradientes de las líneas de tendencia *se hacen más empinadas* con el tiempo. Tanto es así que una vez transcurrido el tiempo necesario para que los efectos de convergencia o divergencia se acumulen suficientemente, la línea de tendencia de un par de lenguas emparentadas puede acabar cruzando la de un par de lenguas no emparentadas, hacia el extremo menos estable a la derecha, como se destaca en la figura 3 mediante el círculo punteado. Esto es, en los significados menos estables, las lenguas emparentadas pueden terminar compartiendo menos correlatos (cognados originales) que las lenguas no emparentadas que adquieren correlatos por convergencia. El español, por ejemplo, sin duda cuenta ahora con más correlatos con el quechua que con el bengalí. Esto es así porque los pocos correlatos (tipo cognados) con el bengalí, que sobreviven desde su distante ancestro común en el

proto-indoeuropeo, son ahora superados en número por los muchos correlatos (tipo préstamo) que el español ahora comparte con el quechua, a medida que este converge hacia aquél.

Figura 3. Líneas de tendencia de porcentaje de correlatos por estabilidad de significado hipotetizadas para cuatro escenarios de parentesco y de contacto lingüístico



6.2 ¿Inferir la historia lingüística a partir de líneas de tendencia?

Cuando se acude a las líneas de tendencia como herramienta para ayudar a diagnosticar la historia lingüística, son pertinentes dos aspectos independientes de ellas, cada uno de los cuales ofrece una señal histórica de forma diferente.

- La pregunta de tipo «sí o no» acerca de si la **tendencia** de la línea es de bajada o de subida —esto es, de más correlatos a menos correlatos, o lo contrario—, conforme uno vaya de izquierda a derecha a lo largo de la gradiente de lo más estable a lo menos estable. Esto sirve como indicador potencial de la historia correspondiente, igualmente de tipo «sí o no»: o de una divergencia a partir de un origen común, si la tendencia es de caída; o de una convergencia de lenguas originalmente no emparentadas, si la tendencia es de subida.

- Lo **empinado** de la línea, o sea si es abrupta o más leve representa un dato continuo que sirve como indicador potencial de dos asuntos progresivos: en pares de lenguas emparentadas, refleja el grado de divergencia y así —aunque de un modo más bien simplista— el periodo de tiempo *relativo* transcurrido desde la separación; o en casos de convergencia, el grado de intensidad del contacto entre las dos lenguas en cuestión.

6.3 Aspectos prácticos y análisis adicionales

Hasta aquí hemos considerado las líneas de tendencia en el nivel de principios teóricos; pero ¿corresponden los datos lingüísticos reales del quechua y del aymara a alguna de ellas en la práctica? Y de hecho, ¿cómo debería proceder uno al analizar los datos lingüísticos reales para convertirlos en tales líneas de tendencia?

Teóricamente, la gradiente más detallada posible ha de obtenerse dividiendo el *continuum* de estabilidad de los significados en el mayor número de conjuntos discretos, de manera tal que cada uno de los conjuntos abarcará el menor número posible de significados. En la práctica, sin embargo, reducir el tamaño del conjunto hasta llegar a un solo significado por conjunto no es una solución viable. Según el dicho de que «cada palabra tiene su propia historia», los significados individuales bien pueden comportarse atípicamente respecto de su «grupo de (casi) iguales» en estabilidad: veamos el aberrante caso de *cuatro*, el único número que muestra formas no cognadas en quechua. El problema se agudiza por el hecho de que para cualquier significado individual, el grado de correlación será forzosamente un valor extremo de 1 o 0, sea correlato o no, en los enfoques lexicostatísticos tradicionales. Las valoraciones intermedias permitidas gracias a los refinamientos metodológicos mencionados anteriormente en §3.1 logran en cierta medida atenuar estos extremos, pero aun así quedan muchos casos o bien de 1, o bien de 0. Cualquier gráfico que se traza significado por significado fluctuaría así muy erráticamente de un extremo del eje y al otro.

Así que necesitamos todavía alguna forma de valoración del grado promedio de correlación por «grupos de iguales» en estabilidad. No hay, sin embargo, pautas que no sean arbitrarias para indicarnos en qué puntos precisos deberíamos escindir la gradiente de estabilidad en subconjuntos discretos. La mejor solución de compromiso entre estas dos consideraciones es trazar una *media móvil* de grados de correlación a lo largo de la gradiente de estabilidad. Esto es, el grado de correlación graficado para cualquier significado individual corresponde en realidad al promedio del valor de correlación en ese mismo significado, más los valores en un número x de significados que están a cada uno de sus lados en el ordenamiento de los significados por estabilidad. Así, donde x tiene un valor de 10, el

subconjunto móvil abarca un total de 21 significados (10+1+10). Nuevamente, no hay principio que pueda guiarnos en identificar algún valor lingüísticamente «correcto» de x —aunque el número total de significados de la lista pone un límite obvio—. Así, tras experimentar con varios valores asignados a x , en los gráficos aquí presentados he seguido lo que parece ser la mejor solución de compromiso: suficiente «atenuación» para que la tendencia sea clara, pero reteniendo al mismo tiempo cierta discriminación, para que todavía sea visible cualquier anomalía significativa. Existe igualmente una gama de transformaciones matemáticas alternativas que pueden ser de utilidad al proveernos con líneas de tendencia más atenuadas y más claras, pero estas se dejan más bien para trabajos ulteriores.

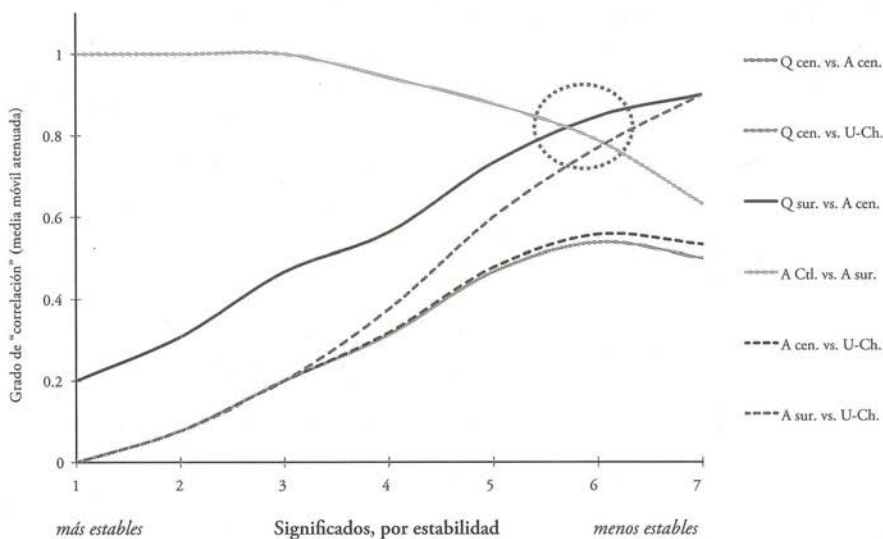
7. Gradientes de estabilidad entre el quechua y el aimara

7.1 *Relaciones lingüísticas ya establecidas, como caso de prueba*

Siempre para propósitos ilustrativos, empecemos de nuevo con apenas el minúsculo conjunto de datos que constituyen los números más bajos (extendido aquí hasta 7 para proporcionar perfiles atenuados algo más fiables e informativos). Se presupone —de manera algo simplista— que las valoraciones de estabilidad caen progresivamente, conforme uno cuenta los números hacia arriba. Y a fin de proveernos de la señal más clara posible, he seleccionado aquí pares de lenguas que representan los extremos divergentes dentro de las familias quechua y aimara: esto es, una variedad central y una sureña de cada una de ellas. Además, he añadido comparaciones que contrastan ambas familias con los datos de Cerrón-Palomino de la lengua chipaya del altiplano boliviano, la única superviviente de la familia uro, que no está emparentada ni con el quechua ni con el aimara —pese a un considerable impacto por contacto con esta última lengua—. La figura 5 muestra las líneas de tendencia de diversos pares de estas lenguas.

Para aquellos lectores enterados de las formas que tienen los números en las respectivas lenguas, el impacto de los cálculos de una media móvil —en este caso usando un valor de $x = 3$ — se nota inmediatamente, y no menos en la línea de tendencia del quechua frente al aimara (central). Pues aun cuando estas lenguas no tienen correlatos en el 1 o en el 2, la línea de tendencia empieza no exactamente en el origen 0, sino un poco por encima, porque sí comparten correlatos en el 3, el cual contribuye a aumentar las medias móviles de los significados inmediatamente precedentes 1 y 2 (y de igual manera, las de los inmediatamente siguientes 4 y 5). El deseado efecto de atenuación es visible en cómo, a pesar de que de un significado individual al siguiente el valor de correlación pueda fluctuar de modo errático de 0 a 1, la línea de tendencia no lo hace otro tanto, sino más bien captura la trayectoria general ascendente. Sin embargo, uno puede todavía detectar

Figura 4. Líneas de tendencia de correlación en lenguas andinas de las mismas o de diferentes familias, en los números del 1 al 7



el detalle de cuáles significados particulares ciertamente muestran correlatos y cuáles no, en la forma ondulante de la línea entre «crestas» (en 3 y 5, por ejemplo, donde hay correlatos) y «valles» (en 4, donde no los hay), respectivamente.

Si ahora recordamos los casos hipotéticos explorados en el subcapítulo §6.1 y comparamos con ellos estas líneas de tendencia que ahora surgen de datos tomados de lenguas reales, las comparaciones del chipaya con cualquier otra lengua resultan en encarnaciones prototípicas del escenario de lenguas no emparentadas que luego entran en contacto (intenso). La línea parte del origen, y sube parejamente con la estabilidad decreciente de los significados. En cambio, comparar las variedades sureñas y centrales del aimara una con otra produce a su vez un caso muy prototípico de la trayectoria inversa, precisamente tal como se predice a partir de los supuestos teóricos en el caso de dos variedades lingüísticas que con seguridad se conocen como emparentadas: empezando del 1 y yendo gradualmente hacia abajo. La mayoría de comparaciones quechua~quechua (es decir, entre dos variedades del quechua) simplemente permanecen altas de una manera consistente, pues las variedades son lo bastante cercanas para no haber perdido correlato alguno —salvo por el caso aislado del significado ‘cuatro’ en algunos pares de dialectos, véase §5.2 arriba—. Sin embargo, donde uno de los dos dialectos quechuas comparados ha perdido algunos números ante el español —en la práctica, siempre los más altos o mayores— la línea de tendencia efectivamente cae de manera bastante abrupta, precisamente como se preve en el caso de lenguas emparentadas cuya divergencia en el léxico se acelera por el contacto con lenguas externas. (Como es el caso en los

dialectos quechuas muy aislados de Incahuasi y de Chetilla —véase §5.1— más impactados por el castellano en este aspecto que las variedades mayores tal como el «Cuzco-Colla»). Además, como se destacó en el óvalo punteado, salta a la vista que efectivamente las líneas de tendencia sí pueden cruzarse en la práctica: el aimara central tiene más correlatos con el aimara sureño en los números bajos, pero en los números altos tiene más correlatos con el quechua.

Finalmente, si buscamos algún indicador sobre la cuestión más fundamental y todavía pendiente de la prehistoria lingüística andina —¿están el quechua y el aimara emparentados?— solo necesitamos mirar las dos líneas de tendencia que representan tales pares de lenguas. ¿Remedan a las líneas de tendencia descendentes que corresponden a lenguas que se sabe son emparentadas, o a las líneas de tendencia ascendentes de lenguas que se sabe no están emparentadas? En una señal diagnóstica mucho más elocuente que un simple conteo de cuántos de los números son correlatos o no, las líneas de tendencia de la figura 4 ponen al aimara y al quechua clara e inconfundiblemente en el campo de lo no emparentado.

No es, por supuesto, que estemos prestos a declarar muerta la hipótesis del quechumara sobre la base de un conjunto de datos de apenas siete palabras. Los números sirven aquí solo por la claridad y elocuencia inhabitual que tiene este particular subconjunto cerrado de significados para efectos de nuestros propósitos ilustrativos iniciales. Una demostración decisiva solamente radicaría en una lista de significados mucho más grande y variada.

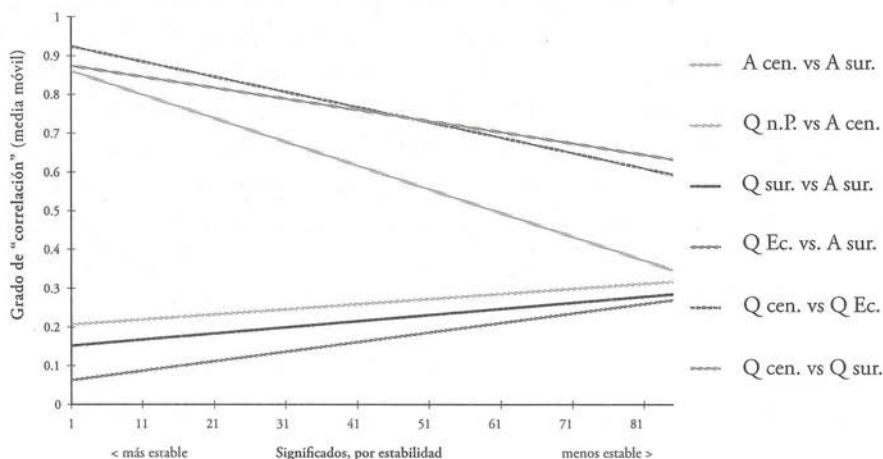
7.2 *Un conjunto completo de datos: ¿a fin de indagar relaciones de parentesco remoto?*

La figura 5 que sigue muestra las líneas de tendencia que resultan de una base de datos de 85 significados en total. Esto representa el traslapamiento, es decir, los significados que se tienen en común, entre dos listas de significados: los 100 para los cuales Holman *et al.* (2008) proporcionan el ordenamiento por estabilidad que empleo aquí; y las 150 de mi conjunto de datos de lenguas andinas. Además, para proporcionar más datos sobre el asunto clave de si el aimara y el quechua están o no distantemente emparentados, las comparaciones de la figura 5 muestran una gama más amplia de variedades representativas del quechua, una por cada uno de sus cuatro principales grupos geográficos y clasificatorios: central («QI» en la terminología tradicional), sureño («QIIc»), ecuatoriano («QIIb») y norteño peruano (parte de la muy discutida rama «QIIa»).

Evidentemente, las proporciones globales de correlatos entre determinados pares de lenguas sí difieren entre este conjunto de datos más grande y el minúsculo compuesto de solo los números básicos. No obstante, lo que no cambia de

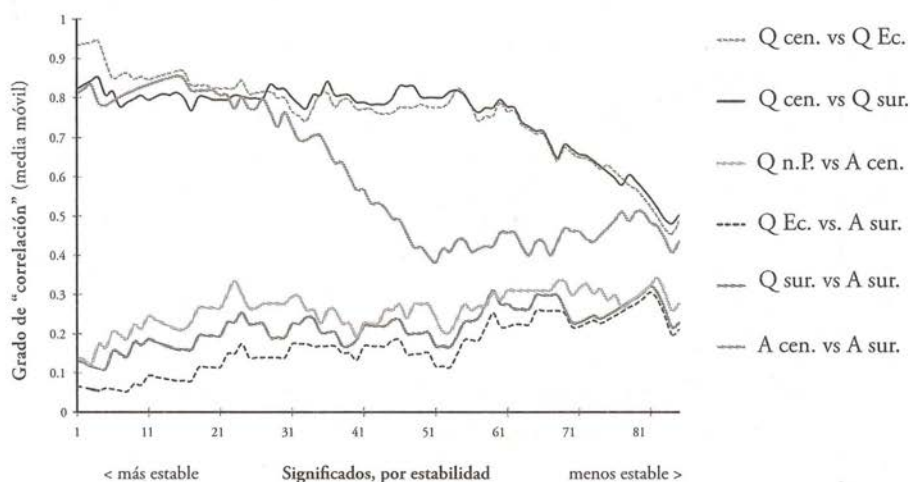
manera significativa es el patrón general, la tendencia, que para ambos conjuntos de datos arroja una señal consistente que apunta a las mismas conclusiones. Esto se ve aún más claramente si uno hace abstracción de las fluctuaciones en los promedios móviles de significado a significado, computando a partir de los mismos datos un análisis de regresión lineal simple —una verdadera *línea de tendencia* en el sentido matemático—. La figura 6 muestra las regresiones lineales de los mismos datos comparativos que están en la figura 5 —según cálculos por la función correspondiente integrada en Microsoft Excel. En cualquiera de las dos visualizaciones, todas las comparaciones quechua~quechua empiezan muy altas y siguen una tendencia de caída, tal como se predice para lenguas emparentadas; semejante es lo que se ve en la comparación de los dos representantes extremos dentro de la familia aimara. Hay, sin embargo, una diferencia muy clara entre las familias, porque la línea de tendencia aimara declina mucho más abruptamente que la del quechua. Retornando a §6.2, las presunciones por defecto indicarían que esto refleja un contacto particularmente intenso que afectó al menos a una de las variedades del aimara, y/o una mayor antigüedad de divergencia de la familia aimara respecto al quechua. La presunción anterior en particular encaja con la conocida e intensa convergencia léxica del aimara central hacia el «mar» quechua que por largo tiempo rodeó sus últimos enclaves pequeños en la provincia de Yauyos. En verdad, pese a los muy altos niveles de correlación en el extremo más estable del eje y, el grado de correlación del aimara central con su variedad hermana del aimara sureño está por debajo del 50% en la mayor parte de la mitad menos estable de la lista de 85 significados.

Figura 5. Líneas de tendencia de correlación en pares de muestra de lenguas andinas en 85 significados



Las líneas de tendencia quechua~aimara, en cambio, nos cuentan una historia muy diferente; en verdad, nos transmiten lo opuesto. Nuevamente, como con los números, empiezan bajo y van subiendo progresivamente. Esto es, una vez más encajan en nuestra línea de tendencia teórica en el escenario de convergencia de lenguas *no* emparentadas, y contradicen la hipótesis quechumara de la divergencia de lenguas lejanamente emparentadas. Y aunque estas líneas de tendencia de «no parentesco» por poco no alcanzan a entrecruzarse con las de los pares de lenguas que se saben emparentadas, como ocurre en el caso de los números, sí se acercan muy marcadamente por el extremo menos estable de la lista. Esto da cuenta de la consabida y tan impresionante cantidad de correlatos quechua~aimara, incluso en este «vocabulario nuclear». Lo que importa para nuestro análisis, sin embargo, es que esta proporción global, pese a su valor relativamente alto, no puede ocultar la *tendencia* claramente ascendente que indica que no hay parentesco.

Figura 6: Análisis de regresión lineal de los pares de muestra de lenguas andinas de la figura 5

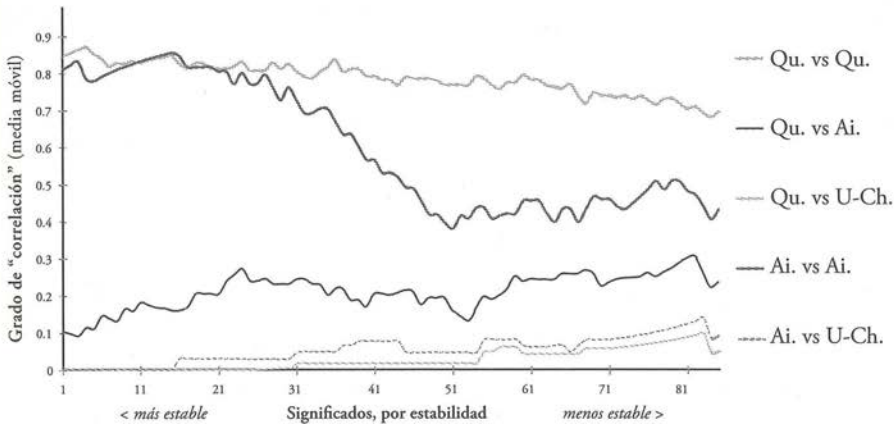


7.3 Promedios de familia

Hay un paso más que se puede dar para consolidar estos resultados. Las líneas de tendencia hasta el momento mostradas se refieren a pares individuales ilustrativos de entre las veintiún variedades de lenguas de mi base de datos andina. Pero para un cuadro más representativo, por cada significado individual también es posible combinar los conteos de correlación de múltiples pares de lenguas en un solo conteo promedio. Esto es, los representantes de los cuatro principales grupos

de quechua se pueden comparar uno con otro, y los conteos de correlación de los seis pares de lenguas que conforman se pueden promediar, para obtener un conteo promedio compuesto de todas las comparaciones que se hacen dentro del quechua, lo que denomino «promedio intrafamiliar». De modo semejante, uno puede promediar todas las comparaciones entre cualquier variedad de quechua y cualquier variedad de aimara, para obtener un «promedio interfamiliar». Estos conteos promedios de correlación, calculados separadamente para cada significado de la lista, a su vez pueden representarse gráficamente como líneas de tendencia, según se muestra en la figura 7. Una vez más estas líneas de tendencia de promedio continúan apuntando en exactamente la misma dirección que todas las de los pares de lenguas individuales que ya hemos visto. La comparación quechua-aimara se comporta, nuevamente, más como un par de lenguas que se conocen como no emparentadas, que como un par conocido por estar emparentado.

Figura 7. Líneas de tendencia de correlación intrafamiliares e interfamiliares en las lenguas andinas



7.4 Escenarios más complejos: ¿remoto origen común seguido por contactos más tardíos?

Los altibajos y vuelcos de la historia a menudo hacen pasar a las poblaciones que hablan determinadas lenguas por trayectorias bastante tumultuosas que podría esperarse lleven a otras formas de línea de tendencia más complejas que los escenarios idealizados y sencillos modelados en el subcapítulo §6.1. Por ejemplo, lenguas originalmente no emparentadas pueden en una etapa anterior haber

experimentado una fase de contacto intenso e intercambio de préstamos, antes de separarse nuevamente, sin que luego haya habido más contacto. O posteriormente podrían entrar en contacto con otra lengua que «sobrescribe» parte del impacto del episodio de contacto precedente. En verdad, escenarios de ambos tipos parecen haberse desarrollado entre las lenguas de los Andes. Para dilucidar tales escenarios más complejos, hay ciertamente más todavía que se puede extraer del nuevo enfoque aquí propuesto, llevándolo más adelante en varios aspectos. Una posibilidad es calcular las líneas de tendencia de correlación no por valoraciones de estabilidad sino por valoraciones de *propensión al préstamo* (tal como las calculan Tadmor *et al.*, 2010), con un potencial para informarnos no (solo) en cuanto a parentesco, sino más específicamente sobre las historias de contacto entre las lenguas. Sin embargo, esta y otras vertientes de indagación han de dejarse más bien para trabajos ulteriores.

Aquí ofrezco nada más que una última demostración del poder diagnóstico potencial de las líneas de tendencia de correlación, pues de los escenarios más complejos que hemos contemplado, hay uno que ciertamente merece que lo consideremos en detalle: la interpretación alternativa «quechumara» de los *NeighborNets* contrastados en la figura 1, ya expuesta en el apartado §4.3. Como ya hemos reconocido, el simple contraste binario entre subconjuntos monolíticos de significados más estables *versus* significados menos estables no pudo negar la posibilidad de una prolongada divergencia a partir de un remoto origen común, seguido de un episodio de contacto más reciente. Así, retornamos ahora a este asunto, para evaluar qué nueva luz podría arrojar a la cuestión mediante la señal más detallada que pueden proporcionar las líneas de tendencia.

¿Qué forma de línea de tendencia se prevé que surgiría de este escenario alternativo? Si solo supervivieran unos pocos cognados originales de un remoto origen común, entonces naturalmente deberían ser los de la más alta estabilidad, así producirían una línea de tendencia con un punto de partida relativamente alto en el mismísimo extremo izquierdo del eje *y*. Esta línea entonces seguiría la típica tendencia generalmente declinante de los casos de parentesco; y dado el gran tiempo transcurrido desde su separación, la pendiente del declive sería particularmente empinada (ver §6.2). Por otra parte, según y conforme uno se mueve más hacia la derecha, llegando a significados cada vez menos estables y cada vez más posibles de ser prestados, sería cada vez más probable que estos registren correlatos, esta vez de tipo préstamo, debido al reciente episodio de contacto. La línea de tendencia, entonces, volvería a subir a valores altos hacia la derecha. En suma, la forma general de la línea de tendencia que prevemos surgiría de tal escenario sería una que se «comba» o se «hunde».

En ninguno de los pares de lengua quechua~aimara analizados en este artículo, en ninguno de los diversos análisis, aparece tal forma. Más bien, cada caso retoma la línea de tendencia baja y gradualmente ascendente, típicamente diagnosticadora de lenguas no emparentadas cuyos correlatos léxicos derivan de la convergencia a través del contacto.

Todos los indicadores que puede ofrecer el nuevo enfoque metodológico aquí propuesto apuntan a la misma conclusión. Aquellos mismos correlatos quechua~aimara que tan a menudo se invocan (sea vagamente) como «sugeren-tes» de un origen común remoto terminan en realidad, mirados más de cerca, señalando completamente lo contrario.

8. Conclusión

Las propuestas metodológicas recientes que buscan aprovechar datos de la semántica léxica en la investigación sobre el parentesco lingüístico han privilegiado la búsqueda de una lista mínima, lo más breve posible, de significados máximamente estables (véase Heggarty, 2010: 316-318). Pero descartar todos los significados menos estables no es menos que echar a un lado la riqueza de datos que existe precisamente en el contraste —y aun más en la gradiente de estabilidad— de lo más a lo *menos* estable. Un enfoque más fino y continuo hacia tales contrastes, en la forma de líneas de tendencia de correlación por estabilidad, puede producir un diagnóstico más claro de la historia y del parentesco de las lenguas que el que puede ofrecer el núcleo más estable por sí solo. Además, es un oráculo capaz de responder con firmeza o «sí» o «no» — en vez de un simple «sí»... o un silencio impenetrable.

Deseo cerrar este ensayo con un llamado a retornar al principio más fundamental de la lingüística comparativa. Dos ramas de nuestra disciplina entran en esta descripción de «comparativa»: por una parte, la lingüística histórica; y por otra, el estudio de «tipología y universales», incluyendo el fenómeno de las *Sprachbünde* o «áreas lingüísticas» (es decir, de convergencia). Para ver más claramente el contraste entre ambas, vale la pena aclarar explícitamente la equivocación más frecuente en las visiones populares de la lingüística, tal como es entendida desde otras disciplinas y entre el público en general.

El trabajo de la lingüística comparativa al buscar «paralelismos» entre las lenguas —para echar mano del término del mismo Cerrón-Palomino (1995, 2008)— en modo alguno sirve para establecer única y necesariamente que ellas provienen de un origen común. Antes bien, a la vista de cualquier conjunto dado de paralelismos entre cualesquiera lenguas —o familias lingüísticas— los lingüistas no las registramos automáticamente como derivadas de un ancestro

común. Examinamos cuidadosamente la naturaleza de los paralelismos que muestran esas lenguas, para preguntarnos sobre todo de qué tipo son, pues solo ciertos paralelismos valen como indicio de un origen común, y estos son del dominio de la lingüística histórica. (Planteando esto de la manera clásica, se busca correspondencias repetidas y regulares entre forma y significado, las cuales revelen que han operado «leyes» de cambio fonético; tales correspondencias son especialmente convincentes a través de amplios paradigmas morfológicos.) Alternativamente, los paralelismos pueden ser de tipo muy diferente, que viene a dar en el dominio de la tipología y los universales: semejanzas estructurales más generales, allí donde están ausentes las correspondencias fonéticas. Los paralelismos de este tipo reflejan, por el contrario, nada más que tendencias universales que actúan en las lenguas humanas, y/o rasgos susceptibles de convergir por efecto del contacto entre las lenguas —es decir, susceptibles de conformar patrones geográficos, sin que interese la filiación dentro de una familia lingüística.

Los investigadores de la lejana prehistoria de las lenguas desafortunadamente enfrentan una «cosa de la vida» en el cambio lingüístico: ocurre tan rápido y es tan abarcador, que nuestro único medio ortodoxo de establecer el parentesco lingüístico, que es el método comparativo, finalmente choca contra el muy discutido «techo» de profundidad en el tiempo, más allá del cual deja de poderse aplicar. Sin embargo, más desafortunado todavía es que la frustración producida por este techo ha tentado a algunos a imaginar que, pese a todo, podemos continuar descubriendo relaciones más profundas y amplias que imaginan están más allá de esa profundidad en el tiempo; y esto con solo dirigirnos al otro campo de la lingüística comparativa, es decir, los paralelismos tipológicos, como si estos más bien fuesen indicadores de un remoto origen común. Los intentos de aplicar tal «lógica», por ejemplo el de Nichols (1992: 3-7), están condenados a la controversia y a ser vistos como heterodoxos, por pisotear el mismísimo principio que distingue en primera instancia estas dos ramas de la lingüística comparativa. ¿Qué mejor recordatorio que el caso del quechua y del aimara de que, de hecho, los paralelismos estructurales no prueban un origen común?

No es que se pueda negar los evidentes paralelismos entre ambas lenguas, por supuesto: y no hay testimonio más elocuente de ellos que en Cerrón-Palomino (1995, 2008). Pero para poder evaluar de manera equilibrada cuán llamativos son en realidad, nos falta todavía un estudio detallado que las ponga en un contexto más amplio aún. Para la más amplia perspectiva, los paralelismos quechua-aimara necesitan verse al lado de otras lenguas «de control», en dos niveles: lenguas similares en estructura aglutinante pero indiscutiblemente no emparentadas con las lenguas andinas, por ejemplo el vasco, húngaro, finés o turco; y lenguas

geográficamente más cercanas al interior de las mismas áreas de convergencia lingüística en América del Sur, pero de otros tipos morfológicos.

En cualquier caso, por muy impresionantes que puedan parecer los paralelismos estructurales entre el quechua y el aimara, eso no importa aquí para el tema de si comparten un ancestro común. Para determinar esto, tenemos que formular la pregunta correcta: no «¿hay algunos paralelismos notorios?», ni siquiera «¿cuán notorios son estos paralelismos?» sino «¿de qué tipo son estos paralelismos?». Y resulta que son más bien del tipo que indica solo semejanza tipológica general o convergencia intensa. Correspondencias del tipo contrario, las que verdaderamente establecen origen común y subsecuente divergencia, brillan por su ausencia.

Casi medio siglo ha transcurrido desde que nuestra especialidad «entró en mayoría de edad», y ya es tiempo de que los especialistas en lingüística andina tomen por las astas el toro metodológico que es la mismísima base de la lingüística comparativa y define la división entre sus dos subdisciplinas. Por muy notorias que sean, las profundas correspondencias estructurales entre el quechua y el aimara son tema de la tipología, los universales y las áreas de convergencia lingüística, no el legado de un profundo parentesco prehistórico. Es tiempo de reenterrar finalmente el «esqueleto en el clóset» de la lingüística andina. Las familias quechua y aimara no son demostrablemente *emparentadas*, en ningún nivel significativo ni recuperable.

9. Referencias bibliográficas

- Adelaar, Willem F. H. con la colaboración de Pieter C. Muysken (2004). *Languages of the Andes*. Cambridge: Cambridge University Press.
- Bryant, David y V. Moulton (2004). NeighborNet: an agglomerative algorithm for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21(2), pp. 255-265. Disponible en: <<http://www-ab.informatik.uni-tuebingen.de/software/jsplits>>.
- Büttner, Thomas Th. (1983). *Las lenguas de los Andes centrales*. Madrid: Ediciones Cultura Hispánica.
- Campbell, Lyle (1995). The Quechumaran Hypothesis and lessons for distant genetic comparison. *Diachronica* XII: 2, pp. 157-199.
- Cerrón-Palomino, Rodolfo (2008). *Quechumara: estructuras paralelas del quechua y del aimara*. Segunda edición revisada y ampliada. La Paz: Plural.
- Cerrón-Palomino, Rodolfo (2003 [1987]). *Lingüística quechua*. Cuzco: Centro Bartolomé de Las Casas.
- Cerrón-Palomino, Rodolfo (2000). *Lingüística aimara*. Cuzco: Centro Bartolomé de Las Casas.

- Cerrón-Palomino, Rodolfo (1995a). *Quechumara: estructuras paralelas de las lenguas quechua y aimara*. La Paz: Centro de Investigación y Promoción del Campesinado.
- Diamond, Jared y Peter Bellwood (2003). Farmers and their languages: the first expansions. *Science* 300, pp. 597-603.
- Dyen, Isidore; Joseph B. Kruskal y Paul Black (1992). An Indoeuropean classification: a lexicostatistical experiment. *Transactions of the American Philosophical Society*, 82 (5). Datos disponibles en: <<http://www.wordgumbo.com/ie/cmp/iedata.txt>>.
- Embleton, Sheila M. (1986). *Statistics in Historical Linguistics*. Bochum: Brockmeyer.
- Gamkrelidze, Thomas V. y Vjačeslav V. Ivanov (1995). *Indo-European and the Indo-Europeans: a reconstruction and historical analysis of a Proto-language and a Proto-culture*. Berlín: Mouton de Gruyter.
- Heggarty, Paul (2010). Borrowability and the notion of basic vocabulary. *Diachronica* 27 (2). Volumen especial sobre *Quantitative Approaches to Linguistic Diversity: Commemorating the centenary of the birth of Morris Swadesh*, editores invitados Wichmann, Søren y Anthony P. Grant.
- Heggarty, Paul (2005). Enigmas en el origen de las lenguas andinas: aplicando nuevas técnicas a las incógnitas por resolver. *Revista Andina* 40, pp. 9-57. Cuzco: Centro Bartolomé de Las Casas.
- Heggarty, Paul y David Beresford-Jones (2010). Agriculture and language dispersals: limitations, refinements, and an Andean exception? *Current Anthropology* 51 (2), pp. 163-191.
- Heggarty, Paul; Warren Maguire y April McMahon (en prensa). Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories. En: Steele, James, Ethan Cochrane y Peter Jordan. *Cultural and Linguistic Diversity: Evolutionary Approaches —Philosophical Transactions of the Royal Society (B)*.
- Holman, Eric W.; Søren Wichmann; Cecil H. Brown; Viveka Velupillai; André Müller y Dik Bakker (2008). Explorations in automated language classification, *Folia Linguistica* 42 (2), pp. 331-354.
- Huson, D.H. y D. Bryant (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23 (2), pp. 254-267. Disponible en: <<http://www-ab.informatik.uni-tuebingen.de/software/jsplits>>.
- Lohr, Marisa (1999). *Methods for the Genetic Classification of Languages*. Tesis doctoral, sin publicar. Universidad de Cambridge.
- McMahon, April y Robert McMahon (2005). *Language Classification by Numbers*. Oxford: Oxford University Press.
- McMahon, April; Paul Heggarty; Robert McMahon y Natalia Slaska (2005). Swadesh sublists and the benefits of borrowing: an Andean case study. En: McMahon, April (editora). *Quantitative Methods in Language Comparison (Transactions of the Philological Society)*, 103 (2), pp. 147-169. Oxford: Blackwell.

- Nichols, J. (1992). *Linguistic Diversity in Space and Time*. Chicago: University of Chicago Press.
- Tadmor, Uri; Martin Haspelmath y Bradley Taylor (2010). Borrowability and the notion of basic vocabulary, *Diachronica* 27(2). Volumen especial sobre: *Quantitative Approaches to Linguistic Diversity: Commemorating the centenary of the birth of Morris Swadesh*, editores invitados Wichmann, Søren y Anthony P. Grant.
- Torero, Alfredo (2002). *Idiomas de los Andes - Lingüística e Historia*. Lima: Editorial Horizonte / Institut Français des Études Andines.