

HERRAMIENTAS CUANTITATIVAS EN EL PROYECTO ALTO
(ATLAS LINGÜÍSTICO DEL TSELTAL OCCIDENTAL)
Y EVALUACIÓN DE LA ZONA DIALECTAL SUR

Gilles Polian - CIESAS

Jean-Léo Léonard - IUF/Univ. París 3

1. INTRODUCCIÓN

En este trabajo, describimos el método que desarrollamos en el marco del proyecto ALTO (*Atlas Lingüístico del Tselal Occidental*) para medir cuantitativamente la proximidad respectiva entre diecisiete localidades de habla tselal en términos de tres ámbitos lingüísticos: léxico, fonología y morfosintaxis. Este método, que se centra en el cálculo de un *Índice de Similitud* dialectal, permite modelizar la complejidad de los hechos dialectales de manera útil a pesar de que sólo disponemos de un conjunto limitado de datos, y se basa crucialmente en la selección, la abstracción y la ponderación de los datos siguiendo una perspectiva que proponemos llamar “diasistémica cuantitativa”. Se distingue de los procedimientos utilizados en la corriente conocida como “dialectometría”, articulada originalmente alrededor del lingüista austriaco Hans Goebel y de sus colaboradores (véase Goebel 1981, 2002, 2010, entre otros), que ha podido aprovechar la existencia de varios atlas dialectológicos de tamaño considerable (miles de variables) en el dominio europeo. Vamos a mostrar que es posible desarrollar formas alternativas de computación de los índices de similitud entre variantes de un idioma o dentro de una red dialectal. Además, veremos cómo y por qué resulta muy conveniente desarrollar herramientas cuantitativas del tipo que proponemos aquí, aun cuando no se dispone de bases de datos muy amplias, como es el caso en dialectología amerindia en general.

Organizamos la presentación de la manera siguiente: en 2 introducimos el campo de estudio de los dialectos del tselal y sintetizamos a grandes rasgos las actividades del proyecto ALTO; en 3 describimos los pormenores del cálculo del Índice de Similitud dialectal; en 4 ilustramos la aplicación de este cálculo, tomando por caso de estudio la cuantificación de la similitud de las tres localidades sureñas con respecto a las demás;

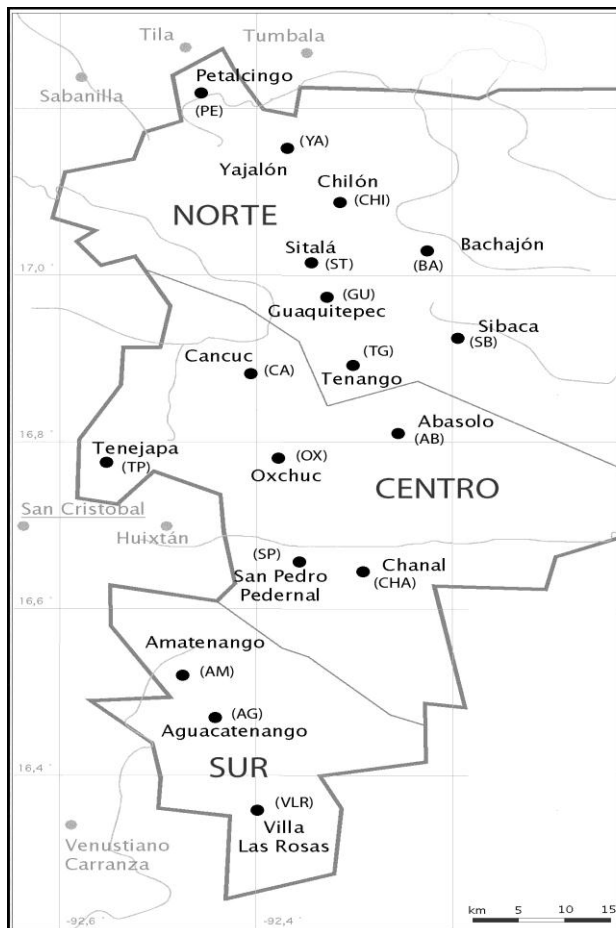
cerramos en 5 con una discusión sobre los alcances, los límites y las perspectivas de este trabajo. Por último, en los anexos se incluye la integralidad de los datos sobre los que se basa este estudio.

2. LA DIALECTOLOGÍA TSELTAL Y EL PROYECTO ALTO

El tselal es un idioma maya, de los idiomas indígenas más hablados en México (y en las Américas en general). Cerca de medio millón de personas¹ lo hablan en las regiones Altos y Selva de Chiapas (centro y este del estado), y su vitalidad es variable según las regiones: se mantiene bien y hasta todavía crece demográficamente en muchas zonas rurales, pero está en declive en los grandes centros urbanos, o incluso está francamente moribundo en ciertas zonas como en Villa Las Rosas, municipio más sureño de los Altos donde se sigue hablando.²

En la segunda mitad del siglo XX, varios autores han estudiado la variación dialectal del tselal, centrándose únicamente en la zona de los Altos, es decir en la zona de ocupación antigua, e ignorando la zona Selva que es de expansión reciente (mismo siglo XX): Robles Uribe (1966), Hopkins (1970), Campbell (1987 y 1988), y sobre todo Kaufman (1972). Todos estos autores concuerdan *grosso modo* en que el tselal actual de los Altos debe dividirse en tres grandes zonas dialectales: norte, centro y sur. El

Mapa 1: Zonas dialectales del tselal y puntos de encuesta de ALTO



¹ 474,298 hablantes de más de tres años según el censo 2010 del Inegi.

² En otra zona más se dejó de hablar recientemente: el tselal suroriental, ubicado al sureste de la ciudad de Comitán hasta la frontera con Guatemala. Los últimos hablantes de este dialecto todavía vivían en los años 1980 (Campbell 1988).

mapa 1 representa esta clasificación tripartita, ubicando únicamente los diecisiete puntos de encuesta tomados en cuenta en ALTO. Entre paréntesis aparecen las abreviaturas usadas para cada localidad en este trabajo.

Las clasificaciones anteriores a las que nos referimos se basan en un pequeño número de variables léxicas y/o fonológicas. En su estudio pionero por ejemplo, Hopkins (1970) sólo toma en cuenta diez sustantivos en veinte localidades.³ Kaufman (1972), a su vez, combina fonología y léxico para caracterizar cada variante. Aunque la perspectiva que adopta es la de la reconstrucción de la protolengua en lugar de la clasificación dialectal, este autor plantea esta misma división tripartita de los dialectos tseltales.

Algunas subdivisiones adicionales dentro de estas tres áreas han sido propuestas por estos autores. Todos reconocen en particular la unidad de OX, CHA y AB (véase el Mapa 1 en cuanto a las abreviaciones utilizadas aquí) dentro de la zona centro: Kaufman (1972) trata incluso estas tres localidades como una sola variante, mientras que Campbell (1987) hace de ellas un “centro oriental”, opuesto a un “centro occidental” compuesto por CA y TP. Por otro lado, Hopkins (1970) establece dentro del norte un grupo entre PE, YA, CHI, BA y ST, que mostrarían más cohesión entre sí que con las demás localidades norteñas (GU, TG, SB y dos localidades más que no se tomaron en cuenta en este estudio, por la heterogeneidad estructural que las caracteriza: Ocosingo y Altamirano).

El proyecto ALTO se planteó profundizar nuestro conocimiento de la variación dialectal tseltal, trayendo al estudio una mayor cantidad de datos léxico y fonológicos, y sobre todo incluyendo datos de índole morfológica y gramatical.⁴ El trabajo de campo inició en 2008 y fue llevado a cabo por los coordinadores y un equipo de siete encuestadores de lengua materna tseltal con 34 hablantes nativos (17 hombres y 17 mujeres de entre 35 y 75 años, bilingües tseltal-español). Los estímulos se dieron en español, pero

³ En este trabajo, usamos de manera indistinta “localidad”, “dialecto” y “variante” como expresiones sinónimas.

⁴ Los dos autores de este trabajo son corresponsables del proyecto ALTO. El financiamiento ha provenido de una variedad de financiamientos: 1) The *Hans Rausing Endangered Languages Project*, Major Documentation Project núm. 164 (véase <http://www.hrelp.org/grants/projects/index.php?projid=138>); 2) el *Institut Universitaire de France*, proyecto “Mesoamerican (Morpho)Phonology”; 3) CONACYT e INALI, a través del proyecto “Documentación lingüística, estudio de la variación dialectal y elaboración de un vocabulario comparativo en tseltal”. El proyecto ALTO también se inscribe dentro de un proyecto más amplio de documentación general del tseltal conducido desde el CIESAS-Sureste por G. Polian, gracias al cual se ha constituido hasta ahora un corpus amplio de unas 400 horas de grabaciones diversas (entrevistas, habla ritual, conversaciones espontáneas, etcétera), cubriendo una diversidad de dialectos. Este es el corpus al que nos referimos en este estudio.

las interacciones entre encuestado y encuestador, dicho de otra manera la “negociación” de las formas para elicitación, se llevaron a cabo enteramente en tselta. El corpus está compuesto por un cuestionario léxico (175 preguntas), fonológico (196 preguntas), morfológico/gramatical (92 preguntas) y un cuestionario sociolingüístico de 63 preguntas que permiten identificar el perfil idiolectal de los informantes, su historia de vida sociolingüística y la extensión de su repertorio lingüístico. Las encuestas se registraron mediante grabadoras digitales de audio y siempre que se pudo se usó también cámaras de video. Las grabaciones fueron transcritas con el programa ELAN (en acceso libre en la página www.lat-mpi.eu/tools/elan/). La transcripción fonética del cuestionario fonológico se realizó gracias al programa PRAAT (en acceso libre en la página www.fon.hum.uva.nl/praat/) con el fin de verificar las impresiones auditivas del transcriptor a partir de la señal.

Los datos recolectados están siendo analizados para identificar las variables léxicas, fonológicas y morfosintácticas⁵ y serán publicados integralmente a mediano plazo en un atlas dialectológico, pero se ha empezado a dar a conocer algunos resultados parciales mediante estudios de temas particulares (véase Polian y Léonard 2009, Gendrot, Léonard y Polian 2010 y Léonard, Gendrot y Polian 2011). El presente trabajo se inscribe en esta última línea: en esta etapa del proyecto, contamos para el estudio cuantitativo con un total de 70 variables representativas de lo que podemos llamar el núcleo diasistémico (véase Tabla 1). La noción de *diasistema* se remonta a Uriel Weinreich (1954) y a la dialectología estructural de mediados del siglo pasado: el diasistema es una descripción coordinada por fracciones y módulos descriptivos de las variables estructurales de la variación dialectal averiguadas en el espacio dialectal de un dominio lingüístico. De manera más abstracta, pero tal vez más fácil de entender, podemos decir que el diasistema es la gema multifacética en que se reflejan todas las características de las muchas variantes que componen la trama de la red dialectal: con el diasistema, ya no importan las isoglosas como tales – es decir, como fronteras o límites de áreas, o como líneas que delimitan secciones de un espacio geolingüístico. Lo que importa, son las relaciones proporcionales entre las variables estructurales, o las funciones diferenciales entre los elementos variables en la

⁵ El tercer componente puede llamarse tanto morfosintáctico como gramatical. Conciernen principalmente la conjugación de los verbos (marcas y construcciones de persona, aspecto y voz), pero también marcadores gramaticales diversos, como direccionales y partícula interrogativa.

dimensión variacional de una lengua o de un dominio lingüístico. Con la diasistémica, la dialectología se emancipa del territorio de la lengua para abarcar la gramática o el léxico de un idioma (o de varios idiomas en situación de continuum estructural) como sistema a geometría variable, ya no como conjunto de áreas contrastivas en un espacio geográfico. Las variables y las matrices de variables que analizaremos aquí (presentadas integralmente en los anexos al final de este trabajo) aluden a este enfoque diasistémico.

Tabla 1: Número de variables seleccionadas por ámbito:

Léxico: 36	Fonología: 13	Morfosintaxis: 21
TOTAL: 70		

Estas 70 variables son el fruto de un proceso de selección y de abstracción en base a las transcripciones de las encuestas, proceso que no está concluido por lo que el número de variables irá creciendo en el futuro. La tarea consiste en extraer los elementos variables más frecuentes e identificar las categorías o niveles donde conviene ubicar la variación. En otras palabras, definir variables, como lo estamos haciendo aquí en el marco de una investigación de diasistémica cuantitativa, equivale a hacer un trabajo de tipología lingüística, mas allá de la tipología dialectológica del dominio. Para tomar un solo ejemplo, la construcción perifrástica del aspecto progresivo, ámbito donde muchos dialectos han operado reconfiguraciones e innovaciones estructurales muy diversas en tseltal,⁶ fue escindida en tres variables diferentes (variables morfosintácticas 11, 12 y 13, véase anexo 3), a saber: 1) uso de la construcción progresiva con objeto de 3ª persona, 2) con objeto de 1ª o 2ª persona y 3) auxiliar empleado.⁷ En este caso, el análisis lleva a juzgar más conveniente establecer tres variables con tres o cuatro valores cada una, en lugar de una sola variable que cubriría una multiplicidad de fenómenos disímiles y tendría que especificar un número grande y poco estructurado de valores. Las demás variables fueron elaboradas siguiendo un proceso similar.

Este número de 70 variables puede parecer pequeño para emprender intentos de cuantificación, sobre todo con respecto a la masa considerable de datos manejados por los

⁶ Véase al respecto Polian (en prensa - a).

⁷ Véase el ejemplo 5 más abajo para una ilustración de la construcción progresiva.

trabajos de dialectometría clásica – que llamaremos también aquí de “dialectometría estándar”. Sin embargo, vamos a argumentar la validez de nuestra propuesta por dos principales hechos: 1) estas variables han sido el objeto de una cuidadosa selección, basada en quince años de describir la gramática del idioma por parte del primer autor,⁸ lo cual da cierto crédito al carácter representativo de las mismas y 2) se compensa el número limitado de variables por una ponderación, que asigna un peso a cada variable según su frecuencia de uso y su relevancia sistémica (véase más adelante sobre estos últimos puntos), lo cual remedia en cierta medida el carácter fortuito de esta colección de elementos lingüísticos.

3. EL CÁLCULO DEL ÍNDICE DE SIMILITUD

La complejidad del estudio de la variación lingüística en todas sus modalidades radica en el carácter continuo, heterogéneo y multifactorial de la variación. Frente a tal elasticidad del dato, el recurso de herramientas cuantitativas resulta de mucho valor, tal como lo mostró Labov en sus clásicos estudios sobre las variaciones sociolingüísticas del inglés norteamericano (Labov 1972). En dialectología, la dialectometría estándar ha desarrollado herramientas que permiten medir cuantitativamente la distancia o proximidad entre dialectos, usando para este fin bases de datos con miles de variables.⁹ Nuestro propósito es más modesto y se ubica en margen de la dialectometría estándar, ya que no contamos, ni contaremos antes de mucho tiempo, con bases de datos tan extensas: sólo pretendemos inspirarnos de este tipo de procedimiento para obtener mediante fórmulas matemáticas triviales mediciones preliminares de similitud respectiva entre puntos de encuesta usando el pequeño número de variables disponibles. A continuación, describimos nuestro método.¹⁰

El cálculo que empleamos se desarrolla en dos etapas. En un primer tiempo, los valores de cada variable son comparados para cada par de dialectos, según los requisitos del

⁸ Véase en particular Polian (2006) y Polian (en prensa - b).

⁹ Acerca del método dialectométrico, véase Goebel (1981 y 2010), donde se presenta lo más esencial de los fundamentos matemáticos del método y, para una aplicación más reciente al ámbito galo-romano véase Goebel (2003).

¹⁰ Nuestra propuesta es parte de un conjunto de alternativas a la dialectometría estándar, como los varios ensayos presentados en Niit (1981) o el estudio cuantitativo sobre variación dialectal del finés publicado por Kalevi Wiik a inicios del siglo (Wiik 2004). Si, por un lado, no hay duda de que la dialectometría estándar tiene mucho poder descriptivo, tanto a nivel geolingüístico como taxonómico (es decir, la clasificación de las variantes), por otro lado, queda todavía mucho por hacer para vincular la dialectología cuantitativa a la tipología lingüística y para entender mejor la causalidad sociolingüística que determina las redes y subredes dialectales observables gracias a los mapas dialectométricos (mapas coropletos, con gradación cromática en la semiología gráfica).

cálculo del Índice de Similitud. Un índice de 0 a 1 es establecido para cada variable, tomando 1 para una coincidencia completa y 0 cuando la correspondencia es nula. Cuando la coincidencia es sólo parcial, se establece un valor intermedio entre 0 y 1, según el grado de similitud (véase más adelante). En un segundo tiempo se multiplica este índice de 0 a 1 por un número establecido por nosotros que representa la importancia relativa de cada variable para una clasificación dialectal desde nuestro análisis como lingüistas. A esta segunda etapa del cálculo se le llama *ponderación*, y se fundamenta en dos criterios funcionales: el *rendimiento* y la *transparencia* estructural, como veremos adelante. La suma de los valores obtenidos para cada variable da el Índice de Similitud para los dos dialectos considerados, cuyo valor está diseñado para fluctuar entre 0 y 100. Vamos a presentar cada uno de estos dos pasos en los párrafos siguientes.

El cálculo del grado de similitud parte de las premisas siguientes: para una variable dada, se considera que a cada valor corresponde una proporción. Por ejemplo, si determinada variable tiene un solo valor {a}, este valor único corresponde a una proporción de 1. Si hay dos valores {a,b}, cada uno de estos valores corresponde a una proporción de 0.5. Si hay tres valores {a,b,c}, la proporción de cada valor es de 0.33 (1/3), etcétera. También se toman en cuenta valores secundarios, cuando al lado de una forma primaria más común y frecuente coexiste una forma equivalente secundaria. Por ejemplo la variable fonológica 6 (evolución de los grupos *VhV) tiene dos valores: a) conservación de la aspiración /h/ entre las vocales, b) pérdida de la /h/; en GU, el valor (a) es dominante, pero también se observan casos de (b), por ejemplo ‘humo’ se dice generalmente *ch’ajil*,¹¹ pero a veces también *ch’ail*. En este caso, se nota el valor secundario entre paréntesis: {a,(b)}, y se atribuye a *a* una proporción de 0.75 y a *b* un valor de 0.25 (dicho de otro modo, un valor secundario vale la tercera parte de un valor primario).¹² Siguiendo la misma lógica, si para una variable observamos dos valores primarios y uno secundario {a,b,(c)}, las proporciones son las siguientes: 0.43 (3/7) para *a*, 0.43 (3/7) para *b* y 0.14 (1/7) para *c*, siguiendo el mismo principio de que un valor primario vale por tres valores secundarios.

¹¹Transcribimos el tseltal siguiendo la ortografía usual del idioma, donde *j* corresponde a una fricativa velar [x] o laríngea [h], en las variantes que, como GU aquí, han neutralizado los fonemas /x/ y /h/ de la protolengua. Esta distinción sólo se conserva ahora en PE y BA.

¹² Desde luego, esta proporción de 3 a 1 es arbitraria, pero permite tomar en cuenta de manera intuitivamente tolerable casos intermedios entre presencia y ausencia de un valor.

Para comparar los valores de una misma variable de dos dialectos diferentes, identificamos primero la coincidencia entre los dos dialectos para esta variable, es decir las formas compartidas. Por ejemplo, si un dialecto presenta los valores {a,b} para determinada variable, y otro presenta {a,c} para la misma variable, la forma compartida es sólo *a*. Si los valores son {a,(b)} de un lado y {a,b,c} del otro lado, las formas compartidas son *a* y *b*. Establecemos entonces la intersección entre ambos grupos de valores, es decir, la proporción mínima compartida. Por ejemplo, si tenemos {a,b,c} por un lado y {a,d} por otro lado, la forma compartida es *a*. En el primer caso, *a* presenta una proporción de 0.33, y en el segundo caso de 0.5. La intersección es por lo tanto de 0.33.

La segunda parte del cálculo del Índice de Similitud consiste en sacar el promedio entre las proporciones respectivas de las formas compartidas. En el ejemplo precedente, las proporciones son de 0.33 y 0.5, respectivamente, lo que nos da un promedio de 0.42.

Finalmente, el índice de 0 a 1 se obtiene sacando el promedio entre la intersección y el promedio de las proporciones, es decir entre los dos valores obtenidos mediante las operaciones anteriores. En nuestro caso, el promedio entre 0.33 y 0.42 da 0.38. Es fácil averiguar que si no hay ninguna forma compartida, el resultado es de 0, y que si todas las formas son compartidas el resultado será de 1.

La mejor forma de hacer inteligible este cálculo es probablemente ilustrándolo con ejemplos concretos. Vamos a tomar el caso de un dialecto norteño, PE, comparado con un dialecto sureño AM, en cuanto a las 13 variables fonológicas seleccionadas en este estudio: estos datos se presentan en la Tabla 2 (véase más detalles sobre las variables fonológicas en el Anexo 2). Las variables 1-10 y 12 no plantean ningún problema: en todos estos casos, los valores son ya sea estrictamente idénticos o totalmente divergentes, por lo que el índice de similitud resultante es de 1 (identidad) o 0 (diferencia). El caso es más interesante para las variables 11 y 13. Para la variable 11, PE tiene un solo valor {a}, mientras que AM tiene dos valores {a,b}. La forma compartida es (a), que ocupa una proporción de 1 para PE (único valor) y de 0.5 en AM. El cálculo se hace entonces de la manera siguiente: intersección (Inters.) = 0.5; promedio entre las proporciones compartidas (Prom.) = 0.75; índice final de similitud (promedio entre los dos resultados anteriores) = 0.625. En el caso de la variable 13, tenemos una configuración similar, con la diferencia de que el segundo valor presente en AM es secundario {a,(b)}. El valor (a) compartido tiene por lo tanto una

proporción compartida de 0.75, conforme a lo estipulado arriba. Será fácil comprobar que el cálculo da entonces un resultado de $(0.75 + ((1 + 0.75) / 2) / 2 = 0.8125$.

Tabla 2: Similitud fonológica entre PE y AM

Variabes Fonológicas	PE	AM	Inters.	Prom.	Similitud
1. p' - b	a	a	1	1	1
2. x - h	a	b	0	0	0
3. h + Res.	a	b	0	0	0
4. h + Eyect.	a	a	1	1	1
5. CVhCV	a	a	1	1	1
6. VhV	a	b	0	0	0
7. V'V	a	a	1	1	1
8. V'C	a	a	1	1	1
9. h + Oc/Af	a	a	1	1	1
10. s > [ʃ]	a	b	0	0	0
11. VV > V	a	a,b	0.5	0.75	0.625
12. VV > VjV	b	a	0	0	0
13. CVh-C...	a	a,(b)	0.75	0.875	0.8125
TOTAL:					7.4

Como resultado preliminar, observamos que los dialectos PE y AM tienen una similitud fonológica redondeada de 7.4, sobre una escala de 0 (diferencia total) a 13 (identidad perfecta). Sin embargo, este resultado es todavía insatisfactorio, ya que todas las variables no tienen la misma importancia: la variable 1 corresponde a la conservación o pérdida de una distinción fonémica (entre /p'/ y /b/, distinciones que ciertas variantes centrales han perdido a favor del segundo miembro del par), mientras que por ejemplo la variable 5 concierne la evolución de la aspiración /h/ en un contexto muy restringido: la posición de coda preconsonántica en la primera sílaba de palabras polisilábicas (por lo tanto, de estructura CVh.CV...). Este último contexto no es tan raro en tsel'tal pero mucho menos frecuente que la aparición del fonema /p'/. Por lo tanto, es legítimo considerar que la variable 1 es más decisiva para caracterizar la similitud o la diferencia entre dialectos que la variable 5. En consecuencia, la variable 1 debe tener un mayor peso en la computación ponderada del índice que la variable 5.

Esto se resuelve asignando un peso diferente a cada variable. Concretamente, establecemos una ponderación de manera que el peso respectivo de cada variable sea proporcional a su importancia en el sistema considerado y a su frecuencia de aparición (estimación aproximativa en base al corpus), y de manera que el total de las ponderaciones sea de 100. Esto se presenta en la Tabla 3: se agrega a la tabla anterior una columna con la ponderación por variable, que en el caso de la fonología varía de 5 a 14. La última columna a la derecha muestra entonces la similitud ponderada por variable que resulta de multiplicar la similitud no ponderada por la ponderación establecida. El total de estos valores corresponde al Índice de Similitud fonológica final para estos dos dialectos, PE y AM, que es de 54.5, sobre una escala de 0 a 100.

Tabla 3: Similitud fonológica entre PE y AM con ponderación

Variables Fonológicas	PE	AM	Similitud no ponderada	Ponderación por variable	Similitud ponderada
1. p' - b	a	a	1	14	14
2. x - h	a	b	0	12	0
3. h + Res.	a	b	0	7	0
4. h + Eyect.	a	a	1	5	5
5. CVhCV	a	a	1	5	5
6. VhV	a	b	0	9	0
7. V'V	a	a	1	6	6
8. V'C	a	a	1	6	6
9. h + Oc/Af	a	a	1	7	7
10. s > [ʃ]	a	b	0	8	0
11. VV > V	a	a,b	0.625	8	5
12. VV > VjV	b	a	0	5	0
13. CVh-C...	a	a,(b)	0.8125	8	6.5
TOTAL:			7.4	100	54.5

La justificación de la ponderación por variable radica en las propiedades estructurales involucradas en cada variable: por ejemplo en cuanto a la primera variable (1: p' - b), cambios radicales en la estructura interna de una unidad funcional (fonema) tan importante como una eyectiva labial /p'/, que corresponde a una sencilla oclusiva labial sonora /b/, implica el máximo rango de ponderación (14). En el caso de la variable

siguiente (2: x – h), también se trata de un fenómeno que tiene que ver con una correlación básica en el inventario fonémico del idioma: la oposición entre la fricativa velar /x/ y la fricativa laríngea /h/. Pero su peso en la ponderación no alcanza el valor de 14, sino de 12, por ser menor el contraste de estructura interna entre los dos términos de la correlación. En el caso de la correlación p' – b, se dan dos cambios en la estructura interna: pérdida del rasgo glotal (eyectividad) por un lado, y sonorización por otro lado. Se da un factor acumulativo dentro del cambio fonológico. En cambio, con la correlación laringovelar x – h, se da un contraste entre lugar de realización velar (x) contra simple fonación (h).¹³ Además, este contraste resulta sutil en términos perceptivos, y hasta puede pasar desapercibido en el habla. Por supuesto, el hecho de que le demos 14 o 12 o 9 u otro valor a este fenómeno de neutralización es discutible, y podríamos establecer configuraciones distintas de valores, según las decisiones que tomemos. Pero tal como construimos la escala de valores de ponderación para nuestros conjuntos de variables apareados, se alcanza un equilibrio o una consistencia estructura interna a la serie. Por ejemplo, la discrepancia entre el valor 5 atribuido a la variable 5 (CVhCV), versus el valor 9 otorgado a la variable 6 (VhV) sigue también esta lógica de equilibrio entre criterios de relevancia sistémica y de transparencia de los contrastes: la caída de la fricativa velar en 5 (CVhCV) acontece menos, en términos de frecuencia en el léxico o de combinatoria silábica, que en los núcleos complejos de tipo (VhV) de la variable 6. El criterio del rendimiento estructural baja el valor o el peso de la variable CVhCV.

A esta propiedad, podemos agregar también consideraciones categoriales: el patrón VhV no es una sola cadena de vocales con inserción de una fricativa laríngea, sino que es una vocal rearticulada – mejor dicho, en este caso específico, una vocal reaspirada –, es decir, un núcleo complejo (véase Brown y Wichmann 2004). La dinámica variacional de la categoría VhV resulta muy compleja, y entra en correlación con otros patrones como el

¹³Este análisis se fundamenta en el modelo de la geometría de los rasgos (Clements y Hume 1996): desde este punto de vista, la estructura de los segmentos fonológicos se reparte entre rasgos categoriales (o raíz de la representación), como la clase silábica (vocoide *versus* contoide *versus* aproximante), en el nivel más alto de la arborescencia, y módulos de especificación segmental, como el nodo de fonación (nodo laríngeo) y el nodo de clase mayor, de manera, que incluye el *modo de articulación* (manera) y el *punto articulatorio* (labial, coronal y dorsal). Esta jerarquía permite argumentar que la variable p' – b involucra activamente dos rasgos de diferentes ámbitos de estructura interna (de nodo laríngeo por la efectividad, y de sonorización), mientras que la variable h – x sólo involucra activamente uno (de *manera*, para /x/, con rasgo dorsal, mientras que la realización [h] puede definirse como una debucalización, que neutraliza la especificación dorsal, operando defectivamente a nivel del nodo laríngeo).

núcleo rearticulado V'V, con un alto grado de complejidad en cuanto a las realizaciones, como se demostró en Léonard, Gendrot y Polian (2011). Pero el continuo sutil de realizaciones (VhV, VV, V:, V), la debilidad acústica y, por ende, perceptiva, de la fase de fricción laríngea en el núcleo, son factores que bajan la transparencia del fenómeno en la comunicación real. Por su importancia en cuanto paradigma en el idioma (su *paradigmaticidad*), la vocal reaspirada merece un valor de ponderación fuerte, pero que no alcance 14, por su complejidad variacional y su bajo valor funcional a nivel perceptivo y comunicativo. La solución consistió en darle un peso superior a la media de sólo dos puntos: el coeficiente 9. El mismo criterio comunicativo lleva a bajar el valor de la otra variable de tipo núcleo complejo, es decir la variable 7 (V'V) al coeficiente 6, en lugar de 9: como demostramos en el artículo previamente mencionado (Léonard *et al.* 2011), la fenomenología variacional del núcleo rearticulado V'V (V'V, V~V, V~:, etcétera) es aún más delicada que la del núcleo reaspirado VhV. La transparencia de las realizaciones, a nivel comunicativo, es bastante baja, lo que empuja a bajar de 9 a 6 su peso.

Las variables 2 (x-h), 6 (VhV) y 7 (V'V) han sido el tema de dos estudios previos muy pormenorizados (Gendrot *et al.* 2010, Léonard *et al.* 2011), que nos permiten sugerir estos valores de ponderación, fundamentados principalmente en dos criterios funcionales básicos: por un lado, el carácter categórico de una relación estructural entre dos términos de una correlación, y por otro lado, la transparencia o la univocidad de las realizaciones, que apuntan a una discreta sociolingüística de la variable. Para resumir de manera más amena estos dos criterios, podríamos decir que en el primer caso, el sonido se escucha a menudo, o se escucha mucho, en el segundo caso, se escucha bien.

La ponderación de las variables léxicas se realizó principalmente mediante estimaciones de frecuencia de uso de los lexemas correspondientes en base al corpus, aunque no vamos a profundizar aquí la justificación para esta área. En cuanto a los criterios de ponderación de las variables de morfosintaxis, ya dimos las explicaciones en Polian y Léonard, 2009. En este dominio de la estructura del diasistema, también prevalecen, para aumentar o bajar el peso de cada variable, los dos criterios funcionales del *rendimiento* y de la *transparencia* estructural. El primer criterio tiene que ver con la relevancia estructural de la categoría variable, mientras que el segundo criterio tiene que ver con su integridad

perceptiva: cuanto más *frecuente* y *clara* es una forma, más peso tendrá en la ponderación de la similitud entre variantes dialectales del diasistema.

Después del ejemplo pormenorizado del cálculo del Índice de Similitud fonológica aplicado al par PE-AM presentado en la Tabla 3, extendemos la misma metodología a cada uno de los 272 pares de dialectos (17x16, comparando cada una de las 17 localidades estudiadas con las 16 otras) en los tres ámbitos estudiados; obtenemos así los Índices de Similitud léxica, fonológica y morfosintáctica (redondeados al entero más próximo). Por ejemplo, ilustramos en la Tabla 4 los índices de PE con los otros 16 dialectos, incluyendo AM. Vemos aquí en un vistazo que este índice va de un mínimo de 18 (con OX y CHA, dialectos máximamente diferentes de PE en su fonología) a 95 (con BA, dialecto fonológicamente casi idéntico a PE), por lo que AM se encuentra a una distancia mediana a PE con respecto a su fonología dentro del espectro de toda la variación registrada.

Tabla 4: Índices de Similitud fonológica de PE con los demás dialectos

	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
PE	65	78	95	63	57	73	59	58	63	18	51	48	18	55	56	60

Las matrices integrales de resultados se presentan en los anexos.¹⁴ En base a estas matrices, el trabajo de análisis propiamente dicho puede empezar, con una fineza de detalle que no permite la simple evaluación cualitativa de los datos, por el hecho destacado arriba de que estos datos pocas veces son consistentes entre sí. Este trabajo no pretende ahondar mucho en el análisis: nos conformaremos con tomar un caso entre varios para ilustrar el aprovechamiento que pensamos hacer de esta herramienta cuantitativa: el problema de la definición de una zona dialectal sur dentro de la red de variantes del tseltal, tema que desarrollamos en la siguiente sección.

A modo de conclusión de esta parte, insistimos sobre el hecho de que nuestro método de cálculo del Índice de Similitud no se basa en la aplicación de principios estadísticos establecidos, sino en el carácter satisfactorio de los índices obtenidos desde nuestro punto de vista como lingüistas. Es decir, se trata fundamentalmente de un método

¹⁴ El cálculo fue automatizado, para aligerar la tarea y sobre todo evitar errores, mediante un programa en Java elaborado por Christian Métairie y Nicole Polian, a quienes agradecemos cálidamente su ayuda.

heurístico que abre nuevas ventanas de representación y de exploración de la realidad dialectal.

4. APLICACIÓN DEL CÁLCULO: CUANTIFICACIÓN DE LA ZONA SUR

En esta sección, desarrollamos un estudio de caso que permite ilustrar el aprovechamiento del estudio cuantitativo diferencial para léxico, fonología y morfosintaxis mediante los Índices de Similitud, cuyo cálculo se presentó en la sección anterior. Concretamente, vamos a evaluar la solidez de la zona dialectal sur a la luz de estos nuevos datos.

La llamada zona sur se articula alrededor de tres principales localidades: Amatenango (AM), Aguacatenango (AG) y Villa Las Rosas (VLR) (antes conocida como Pinola). Fue destacada por Hopkins (1970) en base a pocas isoglosas léxicas, como una forma de inicio *tu* (*tu, tut, tu'ti, tut'*) para el interrogativo inanimado 'qué', cuando las demás regiones tienen formas de inicio *b...* (*bi, bin, beluk*, entre otras), o como *xulbal* 'su cuerno', frente a la forma *xulub* registrada en otros lugares. Sin embargo, vamos a mostrar que esta cohesión sureña se desvanece en parte al considerar una gama más amplia de datos, en particular fonológicos y morfosintácticos.

En un primer momento, el léxico parece confirmar lo observado por Hopkins (1970). En particular, las tres localidades del sur comparten rasgos léxicos únicos, que en varios casos denotan innovaciones compartidas, o al revés, una impermeabilidad a una innovación difundida tanto en el norte como en el centro. Así, donde todas las demás variantes tienen para 'ahora, hoy' *yo'tik* (o *ya'tik*, TP), el sur emplea una forma reducida del español *ahora* más un deíctico proximal: AM *or to*, AG *or ili ~ or le*, VLR *or le > o'le*. De forma similar, para 'pueblo, ciudad' el sur dice *tejk-lum*¹⁵ cuando los demás lugares dicen simplemente *lum* (que significa también 'tierra, suelo'); los sureños designan la 'mano derecha' como *bats'il k'ab* (*bats'il* 'verdadero', *k'ab* 'mano'), cuando todos los demás dicen *wa'el k'ab* (*wa'el* 'derecha'); por último, el sustantivo *tsots*, que significa originalmente 'pelo, cabello' en todos los demás lugares ha pasado a significar 'cobija', salvo en el sur. Estos datos se sintetizan en la Tabla 5. Ejemplos así no son raros, que

¹⁵*Tejk* es un elemento semánticamente opaco, que se encuentra también en tsotsil en un compuesto cognado: *tek-lum* 'pueblo'.

vienen a reforzar la idea de que estas tres localidades conforman efectivamente un área cohesionada, diferenciada de las demás zonas dialectales.

Tabla 5: ejemplos de cohesión léxica del sur

	Norte	Centro	Sur (AM, AG, VLR)
‘qué’	<i>bin, binti</i>	<i>bi, bin, binti, beluk, ...</i>	<i>tu, tut, ...</i>
‘su cuerno’	<i>xulub</i>	<i>xulub</i>	<i>xulbal</i>
‘ahora, hoy’	<i>yo'tik</i>	<i>yo'tik (TP: ya'tik)</i>	<i>or to, or ili, or le, o'le</i>
‘pueblo’	<i>lum</i>	<i>lum</i>	<i>tejk-lum</i>
‘mano derecha’	<i>wa'el k'ab</i>	<i>wa'el k'ab</i>	<i>bats'il k'ab</i>
<i>tsots</i>	‘cobija’	‘cobija’	‘pelo, cabello’

Esto no excluye la existencia de contraejemplos, que vienen a matizar la afirmación anterior. En particular, no es raro encontrar lexemas o evoluciones particulares de ciertos lexemas donde las tres localidades sureñas han divergido, como se muestra en la Tabla 6.

Tabla 6: ejemplos de divergencias (fono)léxicas entre localidades sureñas

	AM	AG	VLR
‘sol’	<i>k'ajk'al</i>	<i>k'ajal</i>	<i>k'a'al</i>
‘día’	<i>k'aal</i>	<i>k'ajal</i>	<i>k'a'al</i>
‘largo, alto’	<i>najt'</i>	<i>najt'</i>	<i>najt</i>
‘mojarse’	<i>ajch'el</i>	<i>k'ojsel</i>	<i>ach'el</i>
‘cielo’	<i>ch'ulxan</i>	<i>ch'ulchan</i>	<i>ch'ulchan</i>

Primero, para ‘sol’ y ‘día’, AM es excepcional en tseltal por tener formas diferentes, cuando todos los demás dialectos presentan una forma única polisémica. Efectivamente, AM tiene una forma conservadora *k'ajk'al* para ‘sol’, similar al tseltal norteño (donde es transparente todavía la derivación a partir de *k'ajk'* ‘fuego’), mientras que para ‘día’ presenta la forma reducida *k'aal*, idéntica a la forma más común en tseltal central. A su vez, AG presenta una variación *k'ajal* de la forma reducida, y VLR tiene *k'a'al* que, siguiendo a Kaufman (1972), analizamos como una reducción de *k'ak'al*, evolución predecible de

**k'ahk'al* en ese dialecto (omisión de /*h/ delante de eyectivas). Así, ninguno de los tres lugares coincide para esta variable. En cuanto a las tres variables (fono)léxicas siguientes, las tres agrupaciones posibles se presentan: primero, AM y AG coinciden en una forma con eyectiva final *najt'* para 'largo, alto' frente a la forma sin eyectiva *najt* de VLR;¹⁶ luego AM covaría con VLR para 'mojarse' en la forma *ajch'el* (AM) - *ach'el* (VLR, la pérdida de la aspiración es esperable en ese contexto) contra *k'ojtse* para AG; finalmente, AG y VLR son iguales para 'cielo' con la forma conservadora *ch'ulchan*, cuando AM presenta la evolución *ch'ulxan* con desafricación de la africada palatal.¹⁷

Estos ejemplos no necesariamente ponen en cuestión la existencia de una zona dialectal sureña, mientras se queden como hechos minoritarios con respecto a los datos del tipo de la Tabla 5. Aquí es donde la posibilidad de una cuantificación nos salva de tener que recurrir a la pura intuición para resolver esta cuestión: en base a las 36 variables léxicas tomadas en cuenta en este estudio, llegamos al panorama presentado en la Tabla 7, lo cual viene a reforzar claramente la idea de que existe una zona dialectal sur, al menos en lo que respecta al léxico (véase matriz completa en el Anexo 1, Tabla 11).

Tabla 7: Matriz cuantitativa del Índice de Similitud léxica para el *tseltal* sureño

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
AM	47	46	45	46	48	49	46	44	44	39	37	52	44	38		81	80
AG	31	34	30	31	34	36	36	39	41	41	39	50	42	40	81		77
VLR	38	35	37	38	38	35	40	39	37	38	33	44	40	34	80	77	

Efectivamente, vemos que la cohesión léxica entre las tres localidades sureñas se traduce en Índices de Similitud mucho más elevados que con el resto de los dialectos: la localidad no sureña más similar a cualquiera de las tres sureñas, TP (centro occidental), no rebasa 52 de similitud (con AM), mientras que la similitud respectiva dentro del sur está alrededor de 80. En conclusión: hasta donde la muestra tomada en cuenta aquí es significativa, existe un sur léxico, estas tres localidades tienen rasgos léxicos que les dan una identidad como zona dialectal hasta cierto grado diferenciada de las demás, aunque también difieran la una de la otra en menor grado.

¹⁶ Se trata de una variable léxica y no fonológica, ya que no corresponde a ninguna evolución regular. Para este tipo de variable, nos referimos a *divergencias (fono)léxicas*, es decir paradigmas donde no se da un cambio de lexema, sino una modificación irregular de la forma fonológica del lexema.

¹⁷ La misma forma aparece en otras dos variantes norteñas: YA y ST.

Ahora, cabe preguntarse si la fonología y la morfosintaxis siguen necesariamente esta misma tendencia. Vamos a ver que no es tan claro. Primero, vemos que las tres variantes sureñas muestran una cohesión fonológica innegable, pero que el sur en su conjunto no se distingue como una zona diferenciada de un conjunto más amplio de dialectos. Básicamente, podemos decir que se ubican en el grupo de las variantes tseltales fonológicamente ni muy conservadoras (a diferencia de PE y BA sobre todo) ni muy innovadoras (caso especialmente de OX y CHA), es decir un tipo de promedio poco diferenciado. Esto se refleja en los valores muy altos para los Índices de Similitud fonológica que se observan en la Tabla 8 a continuación (véase matriz completa en el Anexo 2, Tabla 13).

Tabla 8: Matriz cuantitativa del Índice de Similitud fonológica para el tseltal sureño

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VR
AM	55	74	78	60	82	92	84	96	97	94	60	96	90	60		99	85
AG	56	76	80	61	84	90	86	97	96	95	56	97	86	56	99		86
VLR	60	71	78	65	79	89	81	83	82	87	42	83	72	42	85	86	

El Índice de Similitud de 99 entre AM y AG muestra que estas dos localidades son virtualmente idénticas en cuanto a su fonología. VLR en cambio se distingue en cierta medida, sobre todo por dos variables: la simplificación de /*h/ ante eyectivas ya referida arriba (Variable 4) por una parte, y la conservación de la aspiración intervocálica (*VhV), donde AM y AG (al igual que todos los dialectos centrales) la pierden.

En conclusión, en el ámbito fonológico la zona sur se desdibuja y se funde en un continuo más amplio de variantes neutras, que en la red dialectal se encuentran sobre todo en el centro geográfico y al sur de la zona norte (véase en particular las variantes que corresponden a los Índices de Similitud en negritas en la Tabla 8). Sobre todo, no hay ninguna innovación fonológica o retención que sea específica del sur, todos son rasgos compartidos con otras localidades.

Por último, vamos a estudiar los resultados del estudio aplicado a las 21 variables morfológicas/gramaticales. Aquí, la tarea se complejiza. Hay efectivamente algunos rasgos compartidos únicos entre las tres localidades del sur, pero son pocos, y se confrontan a una cantidad de otros rasgos que apuntan en otras diversas direcciones, como lo vamos a ver. Los rasgos gramaticales exclusivos del sur son los siguientes:

a) una construcción prohibitiva diferente para los verbos intransitivos (Variable 8): donde los demás dialectos usan la negación más el verbo en aspecto incompletivo como en 1a, el sur usa la negación más la forma imperativa como en 1b (véase abreviaturas al final del trabajo).

- | | | |
|-----|---|---|
| (1) | a. ma x-ʔokʔ-at
NEG INC-llorar-B2
‘no llores’ | b. ma okʔ-an
NEG llorar-IMP
‘no llores’ |
|-----|---|---|

b) Un sufijo distributivo de forma *-talan*, diferente de la forma observada en otros dialectos, donde varía entre *-tiklan*, *-tilay*, etcétera (Variable 10).

c) Todos los demás dialectos presentan una “construcción agentiva”, alternativa a la conjugación verbal canónica para los verbos intransitivos agentivos (como ‘hablar’, ‘trabajar’, ‘correr’, pero no ‘caerse’, ‘morir’, etcétera, que son “no agentivos”), y que enfatiza el sentido agentivo mediante el uso del verbo ligero *aʔiy* más una forma no finita del verbo, como en 2 (Variable 19). En cambio, ningún dialecto sureño usa esta construcción.

- | | |
|-----|--|
| (2) | ya k-aʔiy nux-el
INC A1-VL nadar-NF
‘nado’ |
|-----|--|

d) Las tres variantes del sur usan una partícula interrogativa *me* de posición inicial para marcar las preguntas polares, mientras que todas las demás variantes emplean un clítico *bal* de segunda posición (Variable 21).

e) Por último, mencionamos un rasgo adicional, aunque sólo concierne AM y VLR, pero no AG: AM y VLR innovaron una construcción pasiva única, por reanálisis de las formas no finitas transitivas con sufijo *-el*, como en 3a, donde se ve que aparece directamente conjugada (Variable 20). En otros dialectos (incluyendo AG), las formas no finitas con *-el* sólo pueden aparecer como objeto del auxiliar de pasivo *ichʔ*, como en 3b, pero nunca directamente conjugadas.

- (3) a. la s-maj-el-on b. la k-ich' maj-el
 COM A3-pegar-PAS-B1 COM A1-PAS pegar-NF
 'fui pegado' 'fui pegado'

Estos rasgos gramaticales únicos en la red dialectal y compartidos entre las tres localidades sureñas parecen apoyar fuertemente la clasificación que los une en una zona dialectal distinta. Sin embargo, más allá de los rasgos compartidos saltan a la vista varias idiosincrasias morfológicas y sintácticas de cada una de estas localidades. En particular, observamos en VLR rasgos únicos o marcados, como una forma *-atex* del sufijo absoluto de segunda persona plural (siendo *-at* la forma singular y *-ex* la forma plural original, Variable 3), o una forma *-tikon* del sufijo de primera persona plural exclusivo (Variable 4), que sólo aparece en otro dialecto en el norte (SB). En cuanto a esta última variable, AM y AG esta vez comparten la característica típica de la zona central de haber perdido la oposición inclusivo-exclusivo. Luego, VLR tiene otro rasgo excepcional: es el único dialecto que conserva la antigua voz inversa con sufijo *-on*, como en 4a, que fue descrita en tsotsil zinacanteco por Aissen (1999) (Variable 18). Todos los demás dialectos han sustituido a la forma inversa una forma pasiva con sufijo *-ot*, como en 4b.

- (4) a. ¿mach'a maj-on? b. ¿mach'a maj-ot y-u'un?
 quién pegar-INV quién pegar-PAS A3-SR
 '¿quién le pegó?' '¿quién le pegó?' ('¿por quién fue pegado?')

En cuanto a AM, tiene una forma única de auxiliar de aspecto progresivo *yipal* (Variable 13), y una construcción del mismo aspecto muy marcada donde el auxiliar no flexionado precede una forma finita del verbo (Variable 11 y 12), como en 5a; esta construcción, fuera de AM, sólo se encuentra en SP (centro). Ilustramos para contraste, la manera más común de construir el progresivo en VLR, con el auxiliar *yak* flexionado con sufijo absoluto, seguido de una forma no finita (sufijo *-el*) del verbo, donde aparece el paciente marcado por un prefijo de Juego A.¹⁸ El contraste en términos de distancia estructural es patente.

¹⁸ En lingüística maya, se conoce como “Juego A” un conjunto de prefijos que tienen dos funciones básicas: 1) marcar el sujeto de un verbo transitivo (ergativo); 2) marcar el poseedor de un sustantivo. Se opone a un

BA, ST y GU. Estamos hablando aquí de una zona con mucho mayor cohesión morfosintáctica, como se observa con los índices superiores a 90.

Tabla 9: Matriz cuantitativa del Índice de Similitud morfosintáctica para el tseltal más norteño

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
PE		83	80	80	77	80	76	77	60	51	59	46	51	56	29	51	26
YA	83		89	89	94	97	85	85	74	60	68	63	60	65	33	56	30
CHI	80	89		97	91	92	89	88	70	60	68	55	64	65	37	56	34
BA	80	89	97		87	89	85	84	66	56	64	51	60	61	35	57	31
ST	77	94	91	87		99	88	89	76	62	70	65	66	67	39	54	36
GU	80	97	92	89	99		89	88	75	60	68	63	64	65	37	56	34

La asimetría de cohesión morfosintáctica por zona se puede visibilizar mediante una gráfica como la que se presenta en la Figura 1, donde se representó la similitud morfosintáctica de manera icónica, con líneas más gruesas conforme asciende el índice (los umbrales de 87.5, 75 y 59 se seleccionaron por conveniencia). Se observa claramente aquí la formación de al menos dos paquetes densos de dialectos: uno en el norte, alrededor de los núcleos YA-ST-GU por una parte y CHI-BA por otra parte, otro en el centro, sobre todo alrededor del triángulo AB-OX-CHA.

Leyenda:
 Línea plena gruesa: $87.5 < ISM < 100$
 L. hueca: $75 < ISM < 87.5$
 L. punteada: $59 < ISM < 75$
 $ISM < 59$: no representado
 (ISM = Índice de Similitud Morfosintáctica)

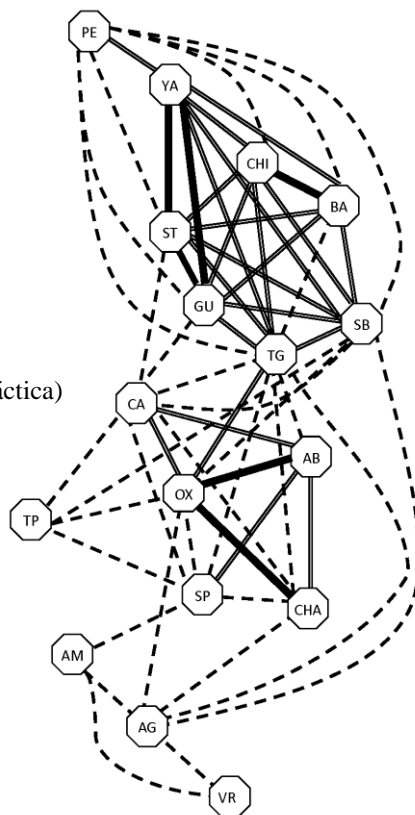


Figura 1: Representación gráfica de la similitud morfosintáctica

En cambio, en el sur no emerge ningún polo de densidad a nivel de similitud morfosintáctica. Más bien, vemos que los tres dialectos sureños se acercan tanto o más a localidades centrales y/o norteñas que entre sí (por ejemplo, AG resulta más cercano al dialecto central OX, con un índice de 63, que con su vecino inmediato VLR, con un índice de 59, véase Tabla 8). Esto indica que el nivel de cohesión morfosintáctica del sur es mucho inferior al que se observa en otras áreas de la red dialectal, y que de hecho no sobresale particularmente ninguna zona sureña: el sur en este respecto no tiene existencia como entidad dialectal discreta, sino que se diluye en una red más amplia y compleja de vínculos múltiples. En otras palabras, no hay ningún sur morfosintáctico claro, mientras que sí hay un norte y un centro morfosintáctico. Más allá de ciertos rasgos gramaticales propios del sur, la modelización cuantitativa que desarrollamos aquí le da más relieve a la heterogeneidad de estas tres variantes, lo cual invita a renunciar a constituir las como una zona dialectal diferenciada.

Esta es la conclusión más sorprendente de este estudio: el divorcio relativo de las características morfológicas/gramaticales y fonológicas comparadas con las características léxicas. Gracias a la perspectiva cuantitativa adoptada aquí, logramos evaluar en qué medida ciertas variantes, como las tres sureñas, pueden formar una zona cohesionada en cuanto a su léxico, al mismo tiempo que se dispersan en cuanto estudiamos sus propiedades morfosintácticas.¹⁹ Esta dispersión radica en parte en las idiosincrasias propias de estas localidades, pero también se debe a los múltiples vínculos estructurales que se revelan no sólo con lugares del centro – el área vecina – sino también con variantes norteñas lejanas (por ejemplo los vínculos de AG con TG y SB visibles en la Figura 1). En pocas palabras, a primera vista, desde el punto de vista cualitativo (enfoque isoglótico), el tseltal sureño tiene un perfil que recuerda las áreas transicionales en dialectología general, cuando su ubicación no es de ningún modo transicional.²⁰ Este resultado interesante habría sido imposible de formular con precisión sin un método cuantitativo del tipo que hemos propuesto aquí.

5. DISCUSIÓN Y PERSPECTIVAS

En síntesis, en secciones anteriores describimos detalladamente el método que escogimos en el marco del proyecto ALTO para cuantificar los datos dialectales del tseltal, e ilustramos su utilidad para evaluar de manera fina el grado de los vínculos interdialectales sobre tres componentes – léxico, fonología y morfosintaxis –, lo cual resulta provechoso a la hora de tomar decisiones analíticas en cuanto a la clasificación dialectal.

De entrada, aclaramos que no pretendemos reemplazar o descartar metodologías cuantitativas ya existentes, como la dialectometría estándar. La razón esencial que nos llevó a desarrollar herramientas propias es el hecho ineludible de que nuestro corpus de datos es apenas incipiente, sin que eso implique renunciar a una perspectiva cuantitativa. Ahora, pensamos que las precauciones metodológicas que tomamos pueden compensar en parte esta limitación del corpus, tal como lo vamos a remarcar aquí. Concretamente, las especificidades metodológicas que deben subrayarse son tres: condensación sistémica,

¹⁹ El caso de la relación entre AG y OX (centro) mencionado arriba es significativo a este respecto. AG y OX son muy diferentes en cuanto a su léxico (índice de similitud: 39, comparado con 81 entre AG y AM, véase Tabla 7), y también en su fonología (índice: 56, comparado con 99 entre AG y AM, véase Tabla 8). En cambio, en cuanto a la morfosintaxis OX es el dialecto más cercano a AG en igualdad con AM, con un índice de 63.

²⁰ Se plantea entonces la cuestión del origen histórico de esta situación mixta o ambigua, pero esto rebasa el alcance del presente estudio.

monitoreo de los datos y ponderación de las variables. Cada uno de estos tres pasos se complementa con el otro, en una cadena lógica.

La condensación sistémica se refiere al hecho de delimitar las variables según criterios diasistémicos, como vemos en el carácter estructurado y/o abstracto de muchas de las variables seleccionadas (véase anexos). Desde este punto de vista, la diasistémica cuantitativa recuerda más el covariacionismo laboviano que la dialectometría: el campo empírico está determinado por el conocimiento previo del núcleo estructural variable del idioma que se analiza. En cuanto al enfoque mesoamericano, es importante subrayar que esta metodología discrepa por completo del método que consiste en utilizar listas de palabras genéricas, como la lista de Swadesh, para comparar cualquier idioma de la zona. En el caso de la diasistémica cuantitativa – como en el covariacionismo –, la lista de palabras que constan en el cuestionario está elaborada según criterios específicos al dominio dialectal analizado, conforme a las variables y a su complejidad interna (por ejemplo, en el caso de la variable fonológica 2: x – h, los contextos inicial, preconsonántico y final de la fricativa, véase Gendrot *et al.* 2010).

El monitoreo de los datos es consecuencia de esta condensación sistémica que acabamos de mencionar: es decir, por el hecho de trabajar con un conjunto parsimonioso de variables, siempre hay manera de volver a los datos para averiguar su incidencia en los patrones de similitud entre las variantes y los grupos de variantes. Las variables tienen su jerarquía y complejidad interna, por ejemplo, en cuanto a su estratificación contextual y a su potencial variacional, a nivel tipológico (véase Léonard *et al.* 2011; Polian y Léonard 2009; Gendrot *et al.* 2010, que presentan etapas del monitoreo de datos). Esta dimensión de análisis de la complejidad interna de las variables es uno de los requisitos del tercer componente del dispositivo analítico de la diasistémica cuantitativa, es decir, de la ponderación.

La ponderación del índice de similitud, como vimos, es la dimensión más falsificable, en términos epistemológicos, de esta metodología: se fundamenta en estos dos criterios funcionales elementales que son el *rendimiento* y la *transparencia* estructural. Las decisiones que toma el lingüista son de su responsabilidad, según su juicio, y se pueden discutir. Este espacio de discusión sobre el valor estructural de las variables y de sus objetos no se abre en la dialectometría estándar, que se sustenta por el efecto de la ley de

los grandes números. Por muy valiosa que sea esta ley, resulta en una paradoja: el lingüista contempla los mapas que representan la gran diversidad de las áreas fonológicas, morfológicas o léxicas de un dominio dialectal, pero estos objetos cartográficos se substituyen a los datos lingüísticos y a la complejidad interna de las series correlacionadas de fenómenos y de variables que componen el diasistema. Al usar un corpus limitado como en nuestro caso, se corre claramente el riesgo de obviar una cantidad de datos importantes que podrían trastocar el panorama resultante, pero la ventaja es el control cercano que se ejerce sobre la exactitud y la pertinencia estructural de los mismos datos. En el mismo tenor, la diasistémica cuantitativa tiene el interés de ser un ejercicio de lingüista y de tipólogo, no sólo de dialectólogo. La elaboración de las variables se hace a través de un cuestionamiento tipológico de las estructuras del idioma, mostrando, como suele hacerlo la dialectología, que los tipos lingüísticos están hechos de muchos subtipos: el grano fino de la tipología, a través de la observación y la caracterización cualitativa de las variables de un idioma dentro de una red dialectal, abre un horizonte nuevo para la tipología lingüística. Más allá de formas de modelización mixta de la tipología lingüística con la dialectología (Kortmann *et al.* 2003), pero focalizadas en un solo componente, como el componente morfosintáctico, la diasistémica cuantitativa abre perspectivas para una tipología dialectológica de los idiomas, en los tres componentes (léxico, fonología y morfosintaxis), manteniendo el control sobre los datos, del principio al final de la cadena argumentativa.

ABREVIATURAS

A: marca personal de “Juego A” (ergativo o posesivo); B: marca personal de “Juego B” (sujeto intransitivo u objeto); COM: aspecto completivo; IMP: imperativo; INC: aspecto incompletivo; INV: inverso; NEG: negación; NF: no finito; PAS: pasivo; PREP: preposición; PROG: auxiliar de progresivo; SR: sustantivo relacional; VL: verbo ligero

REFERENCIAS BIBLIOGRÁFICAS

- AISSSEN JUDITH 1999. "Agent Focus and Inverse in Tzotzil", *Language*, 75, pp. 451-485.
- BROWN CECIL y SØREN WICHMANN 2004. "Proto-Mayan Syllable Nuclei", *International Journal of American Linguistics*, 70 : 128-186.

- CAMPBELL, LYLE 1987. "Tzeltal dialects: new and old", *Anthropological Linguistics*, 29, pp. 549-70.
- CAMPBELL, LYLE 1988. *The linguistics of Southeast Chiapas, Mexico*. New world Archaeological Foundation 50. Provo: Brigham Young University.
- CLEMENTS, NICK y ELISABETH HUME 1996. "The internal organization of speech sounds", en *The Handbook of Phonological Theory*. Ed. John Goldsmith (2ª edición). Oxford/Cambridge: Blackwell, pp. 245-306.
- GENDROT, CEDRIC, JEAN LEO LEONARD y GILLES POLIAN 2010. "Correlación laringovelar y variación dialectal del tzeltal (Maya occidental, Chiapas, México). Enfoques del proyecto ALTO", *Estudis Romànics*, 32, pp. 311-329.
- GOEBL, HANS 1981. "Eléments d'analyse dialectométrique (avec application à l'ALS)", *Revue de Linguistique romane*, 45, pp. 349-420.
- GOEBL, HANS 2002. "Analyse dialectométrique des structures de profondeur de l'ALF », *Revue de Linguistique Romane*, 66, pp. 5-63.
- GOEBL, HANS 2003. "Regards dialectométriques sur les données de l'atlas linguistique de la France (ALF): relations quantitatives et structures de profondeur", *Estudis Romànics*, 25, pp. 59-96.
- GOEBL, HANS 2010. "Dialectometry: theoretical prerequisites, practical problems, and concrete applications (mainly with examples drawn from the 'Atlas Linguistique de la France', 1902-1910)", *Dialectologia*, Special Issue I, pp. 63-77.
- HOPKINS, NICHOLAS 1970, reed. 1990. "Estudio preliminar de los dialectos del tzeltal y del tzotzil", en *Ensayos de antropología en la zona central de Chiapas*. Ed. N. Mcquown y J. Pitt-Rivers. México: INI., pp. 185-235.
- KAUFMAN, TERRENCE 1972. *El proto-tzeltal-tzotzil. Fonología comparada y diccionario reconstruido*. México: UNAM.
- KORTMANN, BERNDT, WALTER BISANG, HANS HENRICH HOCK y WERNER WINTER (eds.) 2003. *Dialectology Meets Typology. Dialect Grammar from a Cross-Linguistic Perspective*. Berlin: Mouton de Gruyter.
- LABOV, WILLIAM 1972. *Sociolinguistic Patterns*. Philadelphia: University of Pennsylvania Press.

- LEONARD, JEAN LEO, CEDRIC GENDROT y GILLES POLIAN 2011. “Voyelles réarticulées en tseltal (maya occidental): pertinence d’une approche dialectologique et expérimentale pour la typologie phonologique”, *Bulletin de la Société de Linguistique de Paris*, 106.1, pp. 395-449.
- NIIT, ELEN (ed.), 1981. *Symposium: Processing of Dialectological Data*. Tallinn: Academy of Sciences of the Estonian SSR.
- POLIAN GILLES y JEAN LEO LEONARD 2009. “La morphologie dans ALTO (Atlas Linguistique du Tseltal Occidental). Réseau dialectal et systèmes à décideurs multiples”. *Géolinguistique*, 11, pp. 149-201.
- POLIAN, GILLES 2006. *Eléments de grammaire du tseltal*. Paris, L’Harmattan.
- POLIAN, GILLES (en prensa - a). “Infinitivos transitivos: innovaciones del tseltal en la familia maya”. En *Ensayos de sintaxis en lenguas de Mesoamérica*. Ed. Roberto Zavala y Enrique Palancar. Mexico: CIESAS.
- POLIAN, GILLES (en prensa - b). *Gramática del tseltal de Oxchuc*. Mexico: CIESAS.
- ROBLES URIBE, CARLOS 1966. *La dialectología tzeltal y el diccionario compacto*. México: Instituto Nacional de Antropología.
- WEINREICH, URIEL 1954. “Is a structural dialectology possible?”, *Word*, 4, pp. 388-400.
- WIJK, KALEVI 2004. *Suomen murteet. Kvantitatiivinen tutkimus* [Dialectos fineses. Estudio cuantitativo]. Helsinki: SKS.

ANEXOS

ANEXO 1: VARIABLES LÉXICAS

Tabla 10: Variables léxicas [= Ponderación]*

Variables	*	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VR
1. ‘qué?’ a: <i>b...(bin, ...)</i> b: <i>tu...</i>	6	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
2. ‘ahora’ a: <i>yo’tik</i> b: <i>or+DEIC</i>	4	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
3. ‘niños’ a: <i>alaetik</i> , b: <i>untik</i>	4	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
4. ‘pueblo’ a: <i>lum</i> b: <i>teyklum</i>	3	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
5. ‘derecha’ a: <i>wa’el</i> b: <i>bats’il</i>	1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
6. <i>tsots</i> a: ‘cobija’ b: ‘pelo’	1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
7. ‘su cuerno’ a: <i>xulub</i> b: <i>xulbal</i>	1	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b

8. '(mi) tortilla' a: <i>_waj</i> b: <i>_ot</i>	2	a	a	a	a	a	a	a	a	a	a	a	b	b	a,b	b	b	b
9. <i>stak</i> 'poder' a: sí b: no	4	b	b	b	b	b	b	b	a	a	a	a	a	a	a	b	b	b
10. 'DEM.LOC' a: <i>le'</i> b: <i>te(y)</i>	6	b	b	b	b	b	b	b	b	b	a	a	b	a	a	b	b	b
11. 'vamos' a: <i>kunik</i> b: <i>konik</i>	3	b	b	b	b	b	b	b	b	a,b	a	a	a	a	a	b	a	b
12. 'también' a: <i>u(u)k, uuk</i> b: <i>e(u)k</i>	6	b	b	b	b	b	b	b	c	b	a	a	b	a	a	b	b	a
13. 'mojarse' a: <i>ajch'el</i> b: <i>t'uxel</i> c: <i>k'ojsel</i>	1	b	b	b	a,b	b	b	b	b	b	a	a	b	a	a	b	c	b
14. 'primero' a: <i>na(h)il</i> b: <i>neel</i> c: <i>nal</i>	4	b	b	b	b	b	b	b	b	a,b	a	b	b	a	b	c	b	
15. 'algodón' a: <i>tunim</i> b: <i>tumin</i>	1	b	b	b	b	b	b	b	b	a	a	a	a	a	a	b	b	b
16. 'moco' a: <i>sim</i> b: <i>tsa'Vlni'</i>	2	b	b	b	b	a,b	b	b	b	a,b	a,b	a	a	a	a	a	b	b
17. <i>amen</i> 'malo' a: sí b: no	1	b	b	b	b	b	b	b	b	b	a	a	a	a	a	b	b	b
18. 'grillo' a: <i>chilchil</i> b: <i>chil</i> c: <i>chulchul</i>	1	b	b	b	b	b	b	b	a	a	a	a	a	a	a	b	c	b
19. 'huevo' a: <i>to(n)mut</i> b: <i>tumut</i>	3	b	b	b	b	b	b	b	a	a	a	a	a	a	a	b	c	b
20. 'pie' a: <i>ok</i> b: <i>akan</i>	3	b	b	b	b	a,b	a,b	a,b	a,b	a	a,b	a	a	a	a	b	c	b
21. 'caballo' a: <i>kawu</i> b: <i>kawayo</i> c: <i>kaj</i>	3	b	b	b	b	b	b	a	a	a	a	a	a	a	a	b	c	c
22. 'ADMIR' a: <i>mati, meti</i> b: <i>kati(k)</i>	3	b	b	b	b	b	a	a	a	a	a	a	a	a	a	b	a,b	b
23. 'encender' a: <i>tsun</i> b: <i>tsum</i>	1	b	b	a,b	b	a	a	a	a	a	a	a	a	a	b	a	b	b
24. 'tarde' a: <i>tibil</i> b: <i>(x)mal~k'a(jk')al</i>	3	a	a	a	a	a	a	a	a	a,b	b	b	b	b	b	b	b	b
25. 'mucho' a: <i>bayel</i> b: <i>bayal</i>	4	a	a	a	a	a	a	b	b	a	b	b	a	b	b	b	b	b
26. 'día, sol' a: <i>k'a(h)k'al</i> b: <i>k'aal</i> c: <i>k'ajal</i>	4	a	a	a	a	a	a	a	b	b	b	b	b	b	b	a,b	b	a
27. 'no' a: <i>ma'uk</i> b: <i>ju'u(k)</i>	6	a	a	a	a	a	a	a	a,b	a,b	b	b	b	b	b	b	b	b
28. <i>namal</i> 'lejos' a: sí b: no	2	a	a	a	a	a	a	a	a	a	b	b	b	b	b	b	b	b
29. <i>a'yey</i> 'hablar' a: sí b: no	2	a	a	a	a	a,b	a,b	a	b	b	b	b	b	b	b	b	b	b
30. 'acompañar' a: <i>jokin</i> b: <i>jo(y)in</i>	2	b	a	a	a	a,b	a,b	a,b	b	b	b	b	b	b	b	b	b	b
31. 'jugar' a: <i>ixta'</i> b: <i>taj(i)mal</i>	3	a	a	a	a	a,b	a,b	a,b	a,b	b	b	b	b	b	b	b	b	b
32. <i>mohlol</i> 'compañero' a: sí b: no	2	a	a	a	a	a	a	b	b	b	b	b	b	b	b	b	b	b
33. <i>xi'lel</i> 'hermano de mujer' a: sí b: no	1	a	a	a	a	a	b	a	b	b	a	b	b	b	b	b	b	b
34. <i>mi</i> 'mucho' a: sí b: no	2	a	a	a,b	b	a	a,b	a	b	b	b	b	b	b	b	b	b	b
35. 'cielo' a: <i>ch'ulchan</i> b: <i>ch'ulxan</i>	2	a	b	a	a	b	a	a	a	a	a	a	a	a	a	a	b	b
36. 'alto, largo' a: <i>na(h/j)t</i> b: <i>najt'</i>	3	a	b	a	a	a,b	b	a	b	a	b	b	a	b	b	b	b	a

Tabla 11: Matriz cuantitativa del Índice de Similitud léxica

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
PE		93	97	96	91	88	84	57	63	33	26	53	29	25	47	31	38
YA	93		94	93	93	89	79	56	56	32	25	46	28	24	46	34	35
CHI	97	94		99	91	89	84	56	63	33	26	53	28	25	45	30	37
BA	96	93	99		89	87	82	56	62	33	27	53	30	26	46	31	38
ST	91	93	91	89		91	83	59	66	38	32	56	33	32	48	34	38
GU	88	89	89	87	91		86	69	70	43	40	58	41	39	49	36	35
SB	84	79	84	82	83	86		71	71	48	42	58	43	41	46	36	40
TG	57	56	56	56	59	69	71		82	70	67	68	68	66	44	39	39
CA	63	56	63	62	66	70	71	82		68	67	87	68	66	44	41	37
AB	33	32	33	33	38	43	48	70	68		96	72	93	95	39	41	38
OX	26	25	26	27	32	40	42	67	67	96		71	93	99	37	39	33
TP	53	46	53	53	56	58	58	68	87	72	71		76	72	52	50	44
SP	29	28	28	30	33	41	43	68	68	93	93	76		94	44	42	40
CHA	25	24	25	26	32	39	41	66	66	95	99	72	94		38	40	34
AM	47	46	45	46	48	49	46	44	44	39	37	52	44	38		81	80
AG	31	34	30	31	34	36	36	39	41	41	39	50	42	40	81		77
VLR	38	35	37	38	38	35	40	39	37	38	33	44	40	34	80	77	

ANEXO 2: VARIABLES FONOLÓGICAS

1. Mantenimiento de la oposición /p'/ - /b/: (a) sí (b) no
2. Mantenimiento de la oposición /x/ - /h/: (a) sí (b) no
3. Mantenimiento de aspiración (/ʰ/) ante resonantes: (a) sí (b) no
4. Mantenimiento de aspiración (/ʰ/) ante consonantes ejectives: (a) sí (b) no
5. Mantenimiento de aspiración (/ʰ/) en posición preconsonántica interna (CVhCVC):
(a) sí (b) no
6. Mantenimiento de aspiración (/ʰ/) entre vocales: (a) sí (b) no
7. Mantenimiento de glotal (/ʔ/) entre vocales: (a) sí (b) no
8. Mantenimiento de glotal (/ʔ/) antes de consonante (/t/, /p/):
(a) sí (b) no (y glotalización de la consonante siguiente)
9. Secuencias /h/ + oclusiva/africada:
(a) sin cambio (b) glotalización de la oclusiva/africada
10. Asimilación del punto de articulación del prefijo s- ('A3') a distancia si la palabra contiene una palatal (por ejemplo: s- 'A3' + *machit* 'machete' → *xmachit* 'su machete' por la /tʃ/ de la segunda sílaba.): (a) no (b) sí
11. Mantenimiento de vocales reaspiradas homorgánicas en palabras frecuentes:
(a) sí (b) no
12. Uso innovador de la aproximante palatal /j/ para romper hiatos: (a) no (b) sí
13. Mantenimiento de /ʰ/ etimológica a final de raíz (*CVh) antes de consonante (con sufijos que empiezan por una consonante): (a) sí (b) no

Tabla 12: Variables fonológicas [* = ponderación]

	*	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VR
1. p' - b	14	a	a	a	a	a	a	a	a	a	a	b	a	a	b	a	a	a
2. x - h	12	a	b	b	a	b	b	b	b	b	b	b	b	b	b	b	b	b
3. h + Res.	7	a	b	a	a	b	b	b	b	b	b	b	b	b	b	b	b	b
4. h + Eyect.	5	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b
5. CVhCVC	5	a	b	a	a	a,b	a	a	a	a	a	a	a	a	a	a	a	a
6. VhV	9	a	a,b	a,(b)	a	a,b	a,(b)	a,b	b	b	b,(a)	b	b	b	b	b	b	a
7. V?V	6	a	a	a	a	a	a	a	a	a	a	b	a	a	b	a	a	a
8. V?C	6	a	a	a	a	a	a	a	a	a	a	b	a	a	b	a	a	a
9. h + Oc/Af	7	a	a	a	a	a	a	a	a	a	a	b	a	a,b	b	a	a	a
10. s > [ʃ]	8	a	a	a	a	a	b	a	b	b	b	b	b	b	b	b	b	b
11. VV > V	8	a	b	a,b	a	b	b	a	a	a	a	a	b	a	a	a,b	a,b	a,b
12. VV > VjV	5	b	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
13. CVh-C...	8	a	a	a	a	a	a,(b)	a	a	a,(b)	a	b	a	b	b	a,(b)	a	a

Tabla 13: Matriz cuantitativa del Índice de Similitud fonológica:

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
PE		65	78	95	63	57	73	59	58	63	18	51	48	18	55	56	60
YA	65		79	60	93	79	82	71	69	73	30	79	60	30	74	76	71
CHI	78	79		83	87	81	89	77	75	80	36	77	66	36	78	80	78
BA	95	60	83		68	62	78	64	63	68	23	56	53	23	60	61	65
ST	63	93	87	68		88	90	79	77	81	38	87	68	38	82	84	79
GU	57	79	81	62	88		81	85	87	88	49	93	80	49	92	90	89
SB	73	82	89	78	90	81		89	87	91	48	81	78	48	84	86	81
TG	59	71	77	64	79	85	89		99	98	59	92	89	59	96	97	83
CA	58	69	75	63	77	87	87	99		97	63	91	93	63	97	96	82
AB	63	73	80	68	81	88	91	98	97		57	90	88	57	94	95	87
OX	18	30	36	23	38	49	48	59	63	57		51	71	100	60	56	42
TP	51	79	77	56	87	93	81	92	91	90	51		81	51	96	97	83
SP	48	60	66	53	68	80	78	89	93	88	71	81		71	90	86	72
CHA	18	30	36	23	38	49	48	59	63	57	100	51	71		60	56	42
AM	55	74	78	60	82	92	84	96	97	94	60	96	90	60		99	85
AG	56	76	80	61	84	90	86	97	96	95	56	97	86	56	99		86
VLR	60	71	78	65	79	89	81	83	82	87	42	83	72	42	85	86	

ANEXO 3: VARIABLES MORFOSINTÁCTICAS

1. Prefijo de primera persona ergativa ante consonante (1ERG):

(a) k- (b) j-

2. Orden del sufijo absolutivo y del sufijo de plural (ABS+PL):

(a) ABS-PL (b) PL-ABS

3. Sufijo de segunda persona plural absolutiva (2ABS.PL):

(a) -ex (b) -atik (c) -atex

4. Primera persona del plural exclusiva (EXCL):
 (a) *-yotik* (b) *-jo'tik* (c) *-kotik* (d) *-tikon* (e) inexistente
5. Sufijo de aplicativo más primera persona singular absoluta (APL+1ABS)
 (a) *-bon* (b) *-ben*
6. Sufijo de aplicativo más segunda persona singular absoluta (APL+2ABS):
 (a) *-beyat* (b) *-bat* (c) *-bet*
7. Sufijo de aplicativo más pasivo (APL+PAS):
 (a) *-bot* (b) *-bet*
8. Forma del verbo intransitivo en modo prohibitivo (VI PROHIB):
 (a) incompletiva (b) imperativa
9. Sufijo iterativo (ITER):
 (a) *-ilan / -ulan* (b) *-ila(y) / -ula(y)*
10. Sufijo distributivo (DISTR):
 (a) *-tik + ITER* (b) *-ti + ITER* (c) *-ta + ITER*
11. Construcción progresiva transitiva con objeto de tercera persona (PROG.O3):
 (a) Aux. + V. inaspectual (b) Aux. + ABS + V. infinitivo (c) Aux. + V flexionado
12. Construcción progresiva transitiva con objeto de 1^a/2^a persona (PROG.O1/2):
 (a) Aux. + V. inaspectual (b) Aux. + ABS + V. inaspectual (c) Aux. + V flexionado
 (d) Aux. + ABS + V. nominalizado poseído
13. Auxiliar de progresivo (PROG AUX):
 (a) *yakal, yak* (b) *nok'ol, no'* (c) *yipal* (d) *yak'al*
14. Reanálisis de la *k* del auxiliar *yak* de aspecto incompletivo como prefijo ante el prefijo *a-* de segunda persona ergativa (INC+2ERG):
 (a) sin reanálisis (b) reanálisis
15. Reanálisis de la *k* del auxiliar *yak* de aspecto incompletivo como prefijo ante el prefijo *y-* de tercera persona ergativa (INC+3ERG):
 (a) sin reanálisis (b) reanálisis
16. Direccional de acercamiento 'viniendo' (DIR venir):
 (a) *tal* (b) *tel* (c) *tael*
17. Direccional de alejamiento 'yendo' (DIR ir):
 (a) *bahel, bael* (b) *beel, bel* (c) *k'ajel, k'ael, k'al*
18. Voz inversa (sufijo *-on*) (INVERSO):
 (a) conservada (b) neutralizada a favor del pasivo
19. Construcción de verbo ligero agentivo (CVL AGENT):
 (a) ausente (b) presente

20. Reanálisis de formas transitivas infinitivas como pasivas finitas (PAS NF):

(a) ausente (b) presente

21. Partícula interrogativa (preguntas polares) (PART INT):

(a) *bal* (enclit. de 2ª pos.) (b) *me* (partícula inicial de oración)

Tabla 14: Variables morfosintácticas [* = ponderación]

	*	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
1. 1ERG	5	a	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b
2. ABS+PL	5	a	a	a	a	a	a	a	a	a	a	a	b	a	a	b	a	a
3. 2ABS.PL	5	a	a	a	a	a	a	a	a,b	a	a,b	a,b	a	a,b	b	a	a	c
4. EXCL	8	a	b	c	c	b	b	d	e	b,e	e	e	b	e	e	e	e	d
5. APL+1ABS	6	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a
6. APL+2ABS	8	b	b	b	b	b	b	b	b	c	c	b	c	c	b	a	b	c
7. APL+PAS	4	a	a	a	a	a	a	a	a	a	a	a	b	a	a	a	a	a
8. VI PROHIB	5	a,b	a	a	a	a	a	a	a	a	a	a	a	a	a	b	a,b	b
9. ITER	2	b	b	a	a	a	a	a	a,b	b	b	b	b	a	b	a	a	a
10. DISTR	2	a	a	a	a	a	a	a	a	b	b	b	a	a	b	c	c	c
11. PROG.O3	8	a	a	a	a	a	a	a,b	a	a,b	b	b	b,(a)	b,c	b	b,c	b	b,(c)
12. PROG.O1/2	6	a	a	a	a	a	a	a	a	a	a,b	a,b	a,b,(d)	c	a,b	c	d	d,(c)
13. PROG AUX	4	a	a	a	a,b	a	a	a	a	a	a	a	a	a	a	c	a,c	a,(d)
14. INC+2ERG	4	b	b	a,b	b	a	a,b	a,(b)	a	a	a	a	a	a	a	a	b	a
15. INC+3ERG	2	b	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a	a
16. DIR venir	3	a	a	a	a	a	a	a	a	b	b	b	a	a,c	b	a	a	a
17. DIR ir	3	a	a	a	a	a	a	a	a	b	b	b	b	a,c	b	a,c	d	a,d,(c)
18. INVERSO	4	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a
19. CVL AGENT	4	b	b	b	b	b	b	b	b	b	b	b	b	b	b	a	a	a
20. PAS NF	4	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b
21. PART INT	8	a	a	a	a	a	a	a	a	a	a	a	a	a	a	b	b	b

Tabla 15: Matriz cuantitativa del Índice de Similitud morfosintáctica

	PE	YA	CHI	BA	ST	GU	SB	TG	CA	AB	OX	TP	SP	CHA	AM	AG	VLR
PE	83	80	80	80	77	80	76	77	60	51	59	46	51	56	29	51	26
YA	83	89	89	89	94	97	85	85	74	60	68	63	60	65	33	56	30
CHI	80	89	97	91	92	89	88	70	60	68	55	64	65	37	56	34	
BA	80	89	97	87	89	85	84	66	56	64	51	60	61	35	57	31	
ST	77	94	91	87	99	88	89	76	62	70	65	66	67	39	54	36	
GU	80	97	92	89	99	89	88	75	60	68	63	64	65	37	56	34	
SB	76	85	89	85	88	89	86	73	66	74	60	69	71	42	60	48	
TG	77	85	88	84	89	88	86	75	73	81	56	75	79	44	59	35	
CA	60	74	70	66	76	75	73	75	90	82	73	75	79	41	51	42	
AB	51	60	60	56	62	60	66	73	90	92	68	81	90	43	55	44	
OX	59	68	68	64	70	68	74	81	82	92	60	73	98	43	63	36	
TP	46	63	55	51	65	63	60	56	73	68	60	60	57	34	37	33	
SP	51	60	64	60	66	64	69	75	75	81	73	60	71	59	56	53	
CHA	56	65	65	61	67	65	71	79	79	90	98	57	71	40	60	36	
AM	29	33	37	35	39	37	42	44	41	43	43	34	59	40	63	60	
AG	51	56	56	57	54	56	60	59	51	55	63	37	56	60	63	59	
VLR	26	30	34	31	36	34	48	35	42	44	36	33	53	36	60	59	