

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ
ESCUELA DE POSGRADO



**HACIA UNA TIPOLOGÍA DE LOS FENÓMENOS DE VARIACIÓN
MORFOLÓGICA EN EL SHIPIBO-KONIBO: UNA CONTRIBUCIÓN PARA
SU TRADUCCIÓN AUTOMÁTICA.**

Tesis para optar el grado de Magíster en Lingüística que presenta

MARCELA ALEJANDRA CASTRO WILLIAMS

ASESOR

DR. ROBERTO DANIEL ZARIQUIEY BIONDI

JURADO

DR. JORGE IVÁN PEREZ SILVA
MG. FELIX ARTURO ONCEVAY MARCOS

Lima, 2018

Hacia una tipología de los fenómenos de variación morfológica en el shipibo-konibo: una contribución para su traducción automática

Marcela Alejandra Castro Williams
Pontificia Universidad Católica del Perú

RESUMEN EN CASTELLANO

La lengua shipibo-konibo (SK) es una de las más grandes de la Amazonía peruana. Debido a su numerosa población vernácula que alcanza los 23.000 hablantes, fue favorecida por un proyecto que busca proveer a las lenguas minoritarias del Perú de herramientas computacionales. Desarrollado en el marco del proyecto del Fondo Nacional de Desarrollo Científico Tecnológico y de Innovación Tecnológica (FONDECYT), la iniciativa tiene como meta desarrollar una plataforma para la traducción automática desde el shipibo-konibo al castellano y viceversa. Este proyecto llamado *Chana*, ha formado un equipo interdisciplinario de ingenieros y lingüistas con el objetivo de desarrollar el software y corpus necesario para implementar dicho traductor. En este contexto, la presente investigación propone una tipología que describe la lengua para los fines prácticos de la traducción automática y con ella, ayudar a la solución de problemas que se presentarán en el nivel de programación morfológico del traductor. La tipología expuesta busca clasificar las variaciones alomorfas en cuatro frentes, primero en su nivel lingüístico de condicionamiento en que se presenta la alomorfía, luego ofrece el número de variaciones formales que alcanza en el SK, también menciona el nivel de predictibilidad de dichas variaciones y finalmente describe la semejanza formal entre los alomorfos de un morfema. Estos elementos permitirán a los ingenieros del proyecto identificar las alomorfías por medio del inventario entregado y recomendar algunos ajustes que creo necesarios se deberán considerar en la programación en el nivel morfológico del TA y que servirán para solucionar los problemas cuando el traductor en línea deba traducir desde castellano al SK.

Palabras claves: *alomorfía, shipibo-konibo, traducción automática, tipología*

**Towards a typology of the phenomena of morphological variation in the
Shipibo-Konibo: a contribution for its automated translation.**

Marcela Alejandra Castro Williams

Pontificia Universidad Católica del Perú

RESUMEN EN INGLÉS

The Shipibo-Konibo (SK) language is one of the largest ones in the Peruvian Amazonia. Due to its numerous vernacular population that reaches 23,000 speakers, it was favored by a project which seeks to provide some minority languages in Peru with computational tools. Developed under the framework of the project of the Fondo Nacional de Desarrollo Científico Tecnológico y de Innovación Tecnológica (FONDECYT), the initiative has as goal to develop a platform for a machine translation (MT) from Shipibo-Konibo to Spanish and vice versa. This project called Chana has formed an interdisciplinary team of engineers and linguists with the goal of developing the necessary software and corpus for the implementation of such translator. In this context, the present investigation proposes a typology that describes the language with practical purposes for automated (MT) and, with this, help to solve problems that will come up at the level of the morphological programming of the translator. The typology exposed here looks to classify the allomorphic variations in four fronts, first the linguistic level of conditioning in which the allomorphy presents itself, then it offers the number of formal variations that reaches in the SK, it also mentions the level of predictability of such variations and finally it describes the formal similarity between the allomorphs of a morpheme. These elements will allow the engineers of the project to identify the allomorphies through the delivered inventory and to recommend some adjustments that I see necessary to be considered in the programming at the morphological level of the MT and that will be useful to solve the problems when the online translator should translate from Spanish to SK.

Keywords: *allomorphy, Shipibo-Konibo, machine translation, typology*



A la Iglesia Asamblea de Dios en Quilpué “Los Carrera”

Mi Antioquía

Hechos 13:1-3

Agradecimientos

Estaré profundamente agradecida de mi asesor, el Dr. Roberto Zariquiey, en primer lugar por recibirme en la maestría (y con ello creer en la interdisciplina) así como también en guiarme en esta investigación del proyecto Chana. Sus consejos y paciencia hicieron que la adquisición del conocimiento de la tipología lingüística fuera entretenida y también desafiante. Agradezco también, a Jorge Iván Pérez por haberme orientado con sus valiosas observaciones y sugerencias que han contribuido a mejorar la presente investigación.

Asimismo, agradezco el apoyo del "Concejo Nacional de Ciencia, Tecnología e Innovación Tecnológica" (CONCYTEC Perú) bajo el contrato 225-2015-FONDECYT por la beca recibida para participar de la investigación y realizar esta tesis. Y a Arturo Oncevay por su liderazgo y al compañerismo del equipo del proyecto Chana.

Agradezco a mi Pastor Juan Pineda por haber creído en este desafío de volverme lingüista, y a mi amada iglesia Quilpué Los Carrera en Chile. También a la iglesia peruana, Nazaret, que me adoptó en Lima, a sus Pastores, a sus jóvenes y a toda la congregación.

Agradezco a mis padres Fernando y Jeannette, por siempre estar orando y pendientes a pesar de la distancia. Agradezco su comprensión en estos años que no pude acompañarlos, y por todo el apoyo y cariño que siempre me brindan. Y a mi familia argentina en Lima, Elda y Silvia que me recibieron y cuidaron en este tiempo de estudio.

Y finalmente quiero agradecer a Dios por haber hecho posible el anhelo de venir al Perú y estudiar lingüística, una herramienta valiosa para los futuros proyectos de Traducción Bíblica. Si hasta aquí he llegado, ha sido por SU ayuda (1 S 7:12).

TABLA DE CONTENIDOS

Capítulo	Página
1. Introducción.....	8
1.1. Propósito y delimitación de objetivos.....	10
1.2. Metodología.....	10
1.2.1 Recopilación, identificación y análisis de la variación morfológica.....	11
1.2.2 Desarrollo de una tipología de la variación morfológica a partir de los rasgos resaltantes de los procesos de variación encontrados en el shipibo-konibo.....	12
2. Marco Teórico.....	13
2.1 La sílaba.....	14
2.1.1. La estructura silábica en el SK.....	15
2.1.2 El concepto de mora.....	16
2.1.3 El concepto de acento y pie métrico.....	17
2.1.4 La acentuación en el SK.....	18
2.1.5 El peso moraico de la sílaba en SK.....	19
2.2 Morfología.....	19
2.2.1 Noción de Palabra.....	19
2.2.2 Morfema.....	23
2.2.3 Afijo.....	25
2.2.4 Clíticos.....	27
2.2.5 Variación morfológica.....	29
2.2.5.1 Alomorfia.....	29
2.2.5.2 Suplección.....	33
2.3 La Traducción automática en lenguas minoritarias.....	36
2.3.1 Las lenguas minoritarias desde la perspectiva de la computación.....	37
2.3.2 La lingüística computacional y el procesamiento del lenguaje natural.....	39
2.3.3 La morfología en la traducción automática.....	43
3. Hacia una tipología de la variación formal en la formación de palabras.....	44

3.1 Nivel de condicionamiento.....	44
3.1.1 Dominio fonológico.....	44
3.1.2 Dominio morfofonológico.....	46
3.1.3 Dominio morfosintáctico.....	47
3.1.4 Dominio léxico.....	48
3.1.5 Variación sin motivación aparente (variación inmotivada).....	48
3.2 Variabilidad formal.....	50
3.3 Predictibilidad.....	50
3.4 Semejanza formal.....	53
4. La variación morfológica en el SK.....	56
4.1 Criterio 1: nivel de condicionamiento.....	57
4.3 Criterio 2: variabilidad formal.....	62
4.4 Criterio 3: predictibilidad.....	64
4.5 Criterio 4: semejanza formal.....	65
5. Contribuciones al desarrollo de un traductor automático castellano SK.....	68
5.1 Aplicaciones y sugerencias relacionadas con el nivel de condicionamiento	68
5.1.1 El contador de sílabas.....	69
5.1.2 Ejemplo de procedimiento de contar sílabas y asignar alomorfo	71
5.1.3 Contador de moras.....	73
5.1.4 Ejemplo de procedimiento de contar moras y asignar alomorfo.	74
5.2 Sugerencias en la variación formal y predictibilidad.....	77
5.3 Sugerencias en la semejanza formal.....	81
6. Conclusiones.....	82
Anexos.....	86
Anexo 1 Alfabeto fonético internacional AFI.....	86
Anexo 2 Criterio 1: nivel de condicionamiento.....	87
Anexo 3 Criterio 2: variación formal.....	91
Anexo 4 Criterio 3: predictibilidad.....	93
Anexo 5 Criterio 4: semejanza formal.....	97
Lista de abreviaturas.....	100
Referencias Bibliográficas.....	101

1 INTRODUCCIÓN

El Perú es un país con una gran diversidad lingüística. Cuenta con más de 47 lenguas nativas, distribuidas principalmente en la cordillera y en la selva. Dentro de ellas, la lengua shipibo-konibo (desde ahora SK) es una de las 21 lenguas que componen la familia lingüística Pano, una de las más grandes de la Amazonía peruana. Con cerca de 23.000 hablantes, el SK ha sido estudiado por Pilar Valenzuela en su tesis doctoral “Transitivity in Shipibo-Konibo Grammar” (2003b) y dicho trabajo es un referente para el estudio de la lengua. También el SK cuenta con una serie de otras investigaciones y artículos realizados por Pilar Valenzuela junto con otros colaboradores, como “Ergatividad escindida en wariapano, yaminawa y shipibo-konibo” (2000); “Causativization and Transitivity in Shipibo-Konibo”(2002a); “Relativization in Shipibo-Konibo: A typologically-oriented study” (2002b); “Evidentiality in Shipibo-Konibo, with a Comparative Overview of the Category in Panoan” (2003a); “Participant agreement in Panoan” (2005); “Applicative constructions in Shipibo-Konibo (Panoan)” (2010a); “Ergativity in Shipibo-Konibo” (2010b); “Argument Encoding and Pragmatic Marking of the ‘Ergative’ in Shiwilu (Kawapanan)” (2011b); “Multi-verb predicates and transitivity harmony in Shipibo-Konibo” (2011b); “Illustrations of the IPA: Shipibo”(2001a) y “El desarrollo de un marcador suprasegmental en chácobo (pano)” (2007).

Debido a su numerosa población vernácula, en el contexto de las lenguas amazónicas y nativas del Perú (la cuarta más grande después de los idiomas quechua, aimara y asháninka), y con el deseo del gobierno de acelerar la integración social de estas comunidades, es que fue favorecida por un proyecto que busca proveer a las lenguas amazónicas de herramientas computacionales. Desarrollado en el marco del proyecto del Fondo Nacional de Desarrollo Científico Tecnológico y de Innovación Tecnológica

(FONDECYT), la iniciativa tiene como meta desarrollar una plataforma para la traducción automática (TA) en línea desde el SK al castellano y viceversa. El nombre de este proyecto es *Chana*.

El proyecto *Chana*, consciente de las complejidades de este trabajo, ha formado un equipo interdisciplinario de ingenieros informáticos y lingüistas con el objetivo de desarrollar el software necesario para implementar dicho traductor, lo que supone la adecuada programación de las reglas gramaticales de este idioma, y la creación de un corpus bilingüe en el formato Treebank. En este contexto, la presente investigación aportará al trabajo lingüístico con la identificación y el inventario de las variaciones morfológicas del SK ya que la variación (sea esta regular o irregular) es siempre una fuente de dificultades para la traducción automática.

La presente tesis propone una tipología con el propósito principal de describir la lengua para fines prácticos de la traducción automática y con ella, ayudar a la solución de problemas en el nivel de programación morfológico del traductor. La tipología expuesta busca clasificar las variaciones alomórficas en cuatro frentes, primero en el nivel lingüístico de condicionamiento en que se presenta la alomorfia, luego presenta el número de variaciones formales que alcanza el SK, menciona el nivel de predictibilidad de dichas variaciones y finalmente la semejanza formal entre los alomorfos de un morfema. Estos elementos permitirán a los ingenieros del proyecto identificar las alomorfas por medio del inventario entregado y realizar los ajustes necesarios para solucionar los problemas al nivel del analizador morfológico cuando un texto quiera traducirse desde el castellano al SK.

1.1 PROPÓSITO Y DELIMITACIÓN DE LOS OBJETIVOS

El propósito de este estudio es mostrar los fenómenos de variación morfológica formal en el SK y proponer una tipología práctica para este fin. Así, los objetivos de nuestra tesis son listados a continuación:

- Realizar un compendio de los fenómenos de variación morfosintáctica del SK.
- Ofrecer una clasificación de dichos fenómenos.
- Sugerir una solución de traducción automática ante tales fenómenos.

El estudio está dividido en cinco capítulos. En este capítulo, además de la introducción ya ofrecida, discuto la metodología utilizada para la recolección de datos. El capítulo 2 presenta el marco teórico y explica algunos conceptos centrales de la sílaba, la morfología y la traducción automática. El capítulo 3 describe los criterios utilizados para la tipología propuesta en esta tesis. El capítulo 4 presenta el estudio de treinta alomorfos del SK y plantea una clasificación de los mismos a partir de la tipología esbozada en esta tesis. El capítulo 5 ofrece sugerencias y soluciones prácticas a nivel morfológico, que deberían implementarse en un traductor automático SK castellano. Finalmente, las conclusiones compendian los aportes de esta tipología.

1.2 METODOLOGÍA

Con el fin de alcanzar los objetivos propuestos anteriormente, esta tesis ha seguido el siguiente plan de trabajo, el mismo que consta de dos fases:

Fase 1: Recopilación, identificación y análisis de la variación en SK a partir de fuentes publicadas

Fase 2: Desarrollo de una tipología de la variación a partir de los rasgos resaltantes que encontramos en los fenómenos de variación formal del SK.

A continuación explicaré cada una de las fases.

1.2.1 Recopilación, identificación y análisis de la variación morfológica.

La recopilación de las variaciones morfológicas del SK las realicé revisando la tesis doctoral de Pilar Valenzuela (2003), desde el capítulo 3 hasta el capítulo 21 y luego el Diccionario Shipibo-Castellano (Loriot et al. 1993). Este trabajo se complementó con el estudio de otros estudios de la lengua SK de Pilar Valenzuela junto con otros colaboradores, como “Ergatividad escindida en wariapano, yaminawa y shipibo-konibo” (2000); “Causativization and Transitivity in Shipibo-Konibo”(2002a); “Relativization in Shipibo-Konibo: A typologically-oriented study” (2002b); “Evidentiality in Shipibo-Konibo, with a Comparative Overview of the Category in Panoan” (2003a); “Participant agreement in Panoan” (2005); “Applicative constructions in Shipibo-Konibo (Panoan)” (2010a); “Ergativity in Shipibo-Konibo” (2010b); “Argument Encoding and Pragmatic Marking of the ‘Ergative’ in Shiwilu (Kawapanan)” (2011b); “Multi-verb predicates and transitivity harmony in Shipibo-Konibo” (2011b); “Illustrations of the IPA: Shipibo”(2001a) y “El desarrollo de un marcador suprasegmental en chácobo (pano)” (2007). Cada fenómeno de variación morfológica encontrado fue anotado en una matriz de datos elaborada en Excel como una fila. Para cada fenómeno, se contrastó la forma regular con la forma variable y se añadió ejemplos de textos, con referencia a las páginas

relevantes de la fuente de donde se extrajo. Al final del proceso compilé una tabla con 30 filas, que reunían un total de 93 morfos¹ que utiliza el SK.

Se identificó cada morfema en cuanto a su funcionalidad y descripción gramatical y se analizó qué tipo de variación presentaban.

1.2.2 Desarrollo de una tipología de la variación morfológica a partir de los rasgos resaltantes de los procesos de variación encontrados en el SK.

La tipología debía permitir clasificar los diferentes patrones de variación encontrados en el SK, resaltando sus similitudes y diferencias y con el objetivo de ser útil para la tarea del TA. Después del análisis, se verificó que los aspectos más importantes de destacar en la variación debían ser: la complejidad o variabilidad formal, el nivel de condicionamiento, la predictibilidad y la semejanza formal. El detalle de cada uno de estos rasgos tipológicos es desarrollado en §4.

¹ Este listado general de morfemas son los recopilados en el proyecto del traductor automático por el equipo de lingüistas del que participo, que además sirvió para constatar y complementar mi propia lista.

2. MARCO TEÓRICO

Nuestro marco teórico recoge las estrategias conceptuales y analíticas necesarias para emprender una adecuada comprensión de los patrones de variación en SK y de los requerimientos que estos plantean para la tarea de traducir automáticamente esta lengua. Para ello, tenemos considerado necesario dividir nuestro marco teórico en tres componentes bien delimitados.

El primero se centra en la noción de sílaba y sus implicancias para comprender el sistema acentual del SK (§2.1). Para este fin es necesario también discutir la noción de mora, dado que el acento en esta lengua manifiesta un patrón sensible a peso silábico. Mora y sílaba son nociones centrales para analizar una buena parte de los fenómenos de variación morfológica encontrados en SK, por ello su tratamiento especial en este capítulo.

El segundo componente de este marco teórico proviene de las ideas de la morfología necesarias para describir y analizar los fenómenos de variación morfológica desde una perspectiva tipológica (§ 2.2). En dicha sección se presentan los conceptos de morfema, afijos y clíticos y cómo éstos se relacionan con sus variaciones formales, el tema central de esta tesis, a saber, la alomorfía. Además, señalamos los contextos que condicionan la alomorfía y cómo poder identificarla.

Finalmente, este marco teórico discute algunas nociones centrales para la comprensión de la traducción automática desde la perspectiva de la llamada Procesamiento de Lenguas Naturales (NLP) y la computarización de lenguas minoritarias (§2.3).

2.1 LA SÍLABA

La **sílaba** es la unidad de pronunciación mayor que un sonido y menor que una palabra (Crystal 2000: 511). Fonológicamente los segmentos de consonantes y vocales se agrupan en sílabas. La sílaba se puede definir, entonces, como un conjunto de segmentos agrupados en torno a un núcleo que por lo general es una vocal. El hecho de que los hablantes estén de acuerdo en cuanto al número de sílabas de una palabra demuestra que el concepto de sílaba tiene una realidad psicológica. Incluso al producir una palabra muy despacio se puede pronunciar sílaba a sílaba. (Hualde 2014: 195).

Dentro de una sílaba los segmentos se organizan con una **escala de sonancia**, de tal modo que el segmento con mayor sonancia ocupa el lugar central en la sílaba que se llama núcleo y los otros segmentos, a su izquierda o a su derecha, han de descender progresivamente en sonancia. El orden de sonancia se distribuye de mayor a menor, comenzando con las vocales, seguida de las consonantes deslizantes (o aproximantes) ², en tercer lugar están las consonantes líquidas (vibrantes y laterales), a continuación las consonantes nasales y finalmente las consonantes obstruyentes (fricativas, oclusivas y africadas) (Hualde 2014:196). Para identificar estos cinco grupos ver cuadro del Alfabeto fonético internacional AFI adjunto en Anexo 1.

La sílaba, por lo tanto, consiste en un conjunto de segmentos agrupados alrededor de un pico de sonancia, de tal forma, que los segmentos más cercanos al núcleo tienen un índice de sonancia decreciente (Hualde 2014:196).

La **estructura silábica** se encuentra formada por grupos de segmentos denominados núcleo, ataque (o arranque), coda y rima. El **núcleo** es el centro de la sílaba y el elemento mayor de sonancia dentro de ella. En español como en el SK (Valenzuela

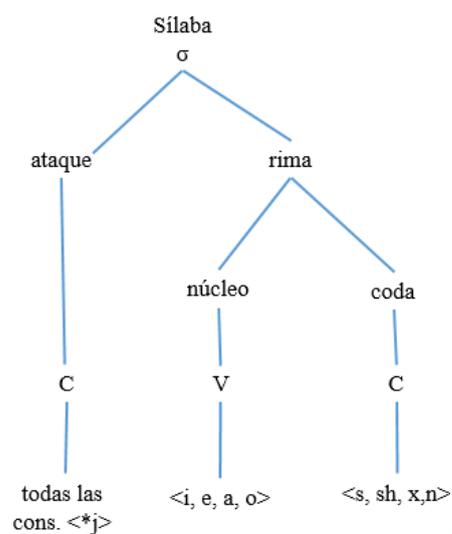
² Los nombres entre paréntesis son los análogos de acuerdo a la clasificación hecha por el Alfabeto Fonético Internacional (AFI).

2003b: 104), las vocales constituyen núcleos silábicos. El ataque (o arranque) es la consonante o grupo de consonantes que precede al núcleo dentro de la sílaba, como la consonante <p> de la sílaba *pe* en la palabra *pero*. En español está permitido el arranque con doble consonante si estas son oclusivas o <f> y va seguida de una líquida, como por ejemplo el grupo silábico *br* de la sílaba *bro* en la palabra *broma*. La combinación prohibida en el español es la <dl> (Hualde 2014: 197). La **coda** es lo que sigue al núcleo de la sílaba. En español, puede ser una consonante o una deslizante como por ejemplo, la [w] en la palabra *flau.ta* o la consonante [k] en la palabra *ac.to* (Hualde 2014: 197). Finalmente, el núcleo y coda se agrupan en una unidad superior denominada rima.

2.1.1. La estructura silábica en el SK

El SK exhibe el canon silábico (C)V(C) con las siguientes combinaciones: V, CV, VC y CVC. En la estructura silábica de la Figura 1, se ve que sólo las vocales pueden ser núcleo de la sílaba, al igual que el español. Respecto de las consonantes, todas pueden ser arranque de la sílaba cuando están en inicio de palabra. Las sílabas en medio de palabra admiten todas las consonantes excepto la <j>. Respecto de las codas, solo pueden componerla las consonantes sibilantes <s, sh, x > y la consonante nasal <n>. Siempre que una secuencia de dos consonantes se produce, cada consonante se asigna a una sílaba diferente. Cuando se presenta una secuencia de vocales cada una es interpretada como dos sílabas nucleares (Valenzuela 2003b: 105-106). Otras palabras con diferente estructura silábica pueden ser prestadas del español.

Figura 1 Estructura silábica del SK



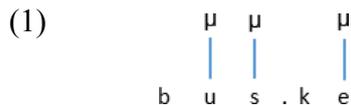
2.1.2 El concepto de mora

El concepto de mora está asociado a los modelos de fonología suprasegmental y a los tradicionales estudios de métrica. Hace referencia a una unidad mínima de tiempo o peso y se simboliza con la letra griega μ 'mi'. El análisis de los segmentos en moras se suele aplicar únicamente a la rima silábica, esto es el núcleo y la coda. Un recuento de moras se usa para describir las lenguas en las que hay una opción de sílaba pesada (de dos moras o bimoraicas) o sílabas ligeras (de una mora o monomoraicas) (Crystal 2000: 373).

El término "peso" es un concepto usado para distinguir niveles de prominencia silábica [...] Las sílabas pueden ser métricamente pesadas o ligeras: una sílaba ligera (o 'débil') es aquella cuya rima consta de un núcleo con una vocal breve solamente o seguido por una coda con una consonante breve como máximo — en términos de cantidad, una mora—; cualquier otro tipo será una sílaba pesada (o 'fuerte') —su cantidad será mayor que una mora—. Las sílabas con

estructura CVVC o CVCC se califican a menudo de "superpesadas"
(Crystal 2000: 427-428).

El ejemplo (1) muestra la palabra *busque*, de dos sílabas /*bus.ke*/ que contiene la sílaba pesada, *bus*, con dos moras o bimoraica y una sílaba ligera, *ke*, que contiene una sola mora o monomoraica (Hualde 2014:206)



2.1.3 El concepto de acento y pie métrico

El **acento** es el término usado en fonética para referirse al grado de fuerza usado en la producción de una sílaba. Martínez Celdrán (1998) muestra que la sílaba prominente o acentuada en castellano supera a las otras sílabas de la palabra en duración, tonalidad e intensidad (en ese orden). La intensidad es lo menos relevante. La distinción habitual se establece entre sílabas tónicas o acentuadas y átonas, donde las primeras son más prominentes que las últimas y aparecen marcadas en la transcripción con una línea vertical superindizada, ['] (Crystal 2000: 26).

En la teoría métrica se asume que el acento no es un rasgo distintivo segmental, sino más bien es una estructura rítmica, es decir, una manifestación lingüística del ritmo equivalente a la que se manifiesta en la música o en el verso (Morales-Front 2014: 238).

En latín el acento recaía sobre la penúltima sílaba si esta estaba cerrada por consonante, por ejemplo con *perfectum* [per.ˈfec.tum] ‘perfecto’, o si contenía una vocal larga, como la <ī> en *formica* [for.ˈmī.ca] ‘hormiga’. En caso contrario, el acento se situaba sobre la antepenúltima sílaba como en *tabula* [ˈta.bu.la] ‘tabla’. La regla de acentuación del latín

considera la distinción entre sílabas pesadas y ligeras y esta distribución se encuentra también en muchas otras lenguas (Hualde 2014:204-205).

Por otro lado, el concepto de **pie métrico** ayuda a representar las alternancias acentuales. Un pie describe la unidad rítmica en lenguas que muestran sincronía, esto es las lenguas en las que las sílabas acentuadas están distribuidas en intervalos aproximadamente regulares a lo largo de un enunciado (Crystal 2000: 428). Los pies métricos son estructuras binarias construidas sobre secuencias de sílabas o moras. En cada pie binario los constituyentes inmediatos son uno fuerte y el otro débil. El elemento fuerte se considera el núcleo de la estructura. El tipo de pie más común en las lenguas naturales es el troqueo silábico, esto es, un pie bisilábico con el núcleo a la izquierda (Morales-Front 2014: 238-240). En contraste está el yambo, pie métrico binario cuyo núcleo se encuentra en el margen derecho del mismo (Nuñez-Cedeño 2014: 400). La propuesta para formalizar el ritmo es la **parrilla métrica**, sistema formado por una secuencia de pies.

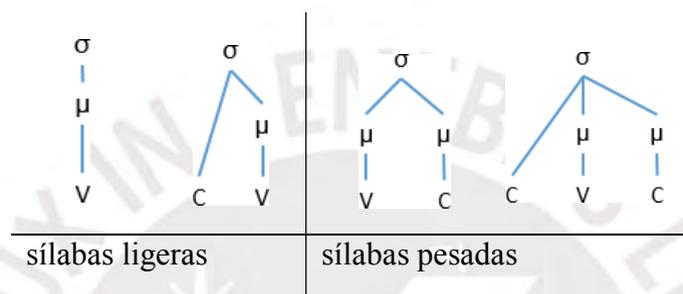
2.1.4 La acentuación en el SK

La acentuación no es completamente predecible en SK. Sin embargo, una gran cantidad de palabras siguen un patrón predecible. El acento primario cae en el primer pie de la palabra. La colocación del acento es trocaica pero sensible al peso de la sílaba; es decir, las palabras generalmente llevan su acento en la primera sílaba a menos que la segunda sea pesada. Si la segunda sílaba es pesada, el acento recaerá sobre ella incluso si la primera sílaba también es pesada (Valenzuela 2003b: 107-108).

2.1.5 El peso moraico de la sílaba en SK

Las palabras SK tienen un mínimo de dos moras. Por lo tanto, las palabras monosilábicas como *bi* 'mosquito nocturno' 'experimenta alargamiento de vocales. Cuando la palabra es bisilábica o mayor a eso, las sílabas con coda, como las de estructura VC y CVC son consideradas pesadas. Por último, la sílaba acentuada es considerada pesada y lleva doble mora. La figura 2 ilustra la asignación de moras de acuerdo a la sílaba posible en el SK.

Figura 2 Asignación de moras en las sílabas del SK



2.2 MORFOLOGÍA

2.2.1 Noción de Palabra

La palabra puede definirse como la unidad de expresión que los hablantes nativos reconocen de manera intuitiva y universal, tanto en la lengua hablada como en la escrita (Crystal 2000: 409). El lingüista generativista Katamba (1993) indica que la asunción de que las lenguas contienen palabras se da por sentada en la mayoría de las personas. Sin embargo, hay algunas dificultades para definirla (1993: 17). Las dificultades para aclarar la naturaleza de la noción de *palabra* se deben en gran medida al hecho de que el término

palabra se utiliza en una variedad de sentidos, tanto fonológicos como gramaticales, que generalmente no se distinguen (Katamba 1993: 17).

Para ayudar a entender el concepto de palabra, Haspelmath (2002:13-16) propone dos **definiciones** que son fundamentales para la morfología: el término *lexema* y la *word-form* traducido como “forma de palabra”.

El término **lexema**, definido como los elementos de vocabulario con un núcleo de significado común y que se enumeran en el diccionario (Katamba 1993: 18). Su nombre deriva de la idea del léxico mental que tiene cada persona en su mente. Son unidades abstractas y pueden considerarse como un conjunto de formas de palabras (Haspelmath 2002: 13). Por ejemplo, las formas inventadas para el hablante inglés ‘*pockling, pockle, pockles, pockled*’ son diferentes realizaciones del lexema *POCKLE*. Todas ellas comparten un significado básico, aunque, están deletreados y pronunciados de manera diferente (Katamba 1993:17).

Por otra parte, el término **forma de la palabra** se explica como una forma de realización física de ese lexema en el habla o en la escritura (Katamba 1993:17). Haspelmath (2002) le llama palabra de texto porque representa las palabras que están separadas por un espacio en la escritura. Las formas de palabra son formas concretas porque pueden ser pronunciadas y utilizadas en textos (2002:14). Así, cada forma de palabra pertenece a un lexema. Podemos ejemplificarlo con la palabra del inglés *see* ‘ver’: *see, sees, seeing, saw y seen*, que son cinco formas diferentes pero que constituyen un solo lexema (Katamba 1993:17).

Ilustraremos los dos conceptos a través del siguiente ejemplo. El lexema verbal *CORTAR* puede representar dos formas de palabras diferentes. En la oración del ejemplo (2 a) *corto* representa el verbo en tiempo presente, primera persona singular. En la

segunda oración, el ejemplo (2 b) la palabra *cortó* representa la forma verbal de tiempo pasado, tercera persona singular.

(2) a. Siempre corto el césped con la podadora.

b. María cortó el pelo a Juan.

Por otro lado, los criterios para identificar las palabras en el habla pueden ser varios. Crystal (2000) propone tres en su diccionario y Escandell (2011) aporta un cuarto criterio. Revisemos cada uno de ellos.

El primero de los criterios presentados por Crystal (2000) tiene que ver con la **estructura interna** de las palabras. Por su estructura interna, las palabras son las unidades lingüísticas más estables. Sus partes constituyentes disponen de poco margen para la reordenación, en comparación con la movilidad posicional de los constituyentes de la oración. Un ejemplo de palabra compleja es “desaprobación” donde *des-aproba-ción* no admite un reordenamiento. Por otro lado, la oración: *Juan y María se dirigieron al desierto* sí admite la ordenación *Al desierto se dirigieron Juan y María*. A este criterio Escandell (2011) le llama, también, **aislabilidad**.

Asimismo, su **ininterrumpibilidad o cohesión** es también un criterio central: no se pueden insertar elementos nuevos y/o pausas en el interior de una palabra en el habla normal, como tampoco permutar sus componentes. Vemos en el ejemplo (3) como la palabra *des-aproba-ción* se convierte en una palabra agramatical si se le inserta nuevos elementos como en (3 a), se le introduce una pausa en su interior (3 b) o se permutan sus componentes (3 c).

(3) a. inserción de elementos nuevos *des-aproba-s-ción

b. pausas al interior *des aprobación

- c. permutar sus componentes *aprobación-des

Otro de los criterios tiene que ver con la noción de **palabra mínima**, esto es, la unidad más pequeña que puede formar por sí mismo un enunciado completo. Las lenguas pueden tener distintos criterios de palabras mínimas. Por ejemplo, en SK (Valenzuela 2003), la palabra mínima debe durar dos moras. Cualquier elemento de menor duración **no puede ser una palabra** en esa lengua. En este criterio podemos incluir otros principios prosódicos como que en algunas lenguas cada palabra debe tener un acento o un tono.

Finalmente, Escandell aporta un último criterio que llama **movilidad posicional** en donde las palabras puedan aparecer en diferentes posiciones de la cadena hablada sin tener obligaciones por su naturaleza de ocupar posiciones fijas. (2011: 132).

En la misma línea, Dixon y Aikhenvald (2003), proponen que la noción de palabra puede entenderse desde dos perspectivas distintas: algunos de los criterios mencionados apuntan hacia la noción de palabra fonológica, mientras que otras nos ayudan a definir a la palabra desde una perspectiva gramatical. Entonces para Dixon y Aikhenvald (2003), la **palabra fonológica** consiste en la unidad fonológica más grande que la sílaba, la cual tiene al menos una propiedad fonológica definitoria elegida de las siguientes tres áreas:

- a) Características segmentales, como estructura interna silábica; realizaciones fonéticas en términos de estas, fenómenos de borde de palabra.
- b) Características prosódicas como asignación de acento y/o tono, así como otras características prosódicas como nasalización, retroflexión, vocalismo, etc.
- c) Reglas fonológicas: algunas reglas se aplican solo dentro de una palabra fonológica, otros se aplican específicamente a través de un límite de palabras fonológicas (Dixon y Aikhenvald 2003: 13).

Finalmente, la **palabra gramatical** consiste en una serie de elementos gramaticales que: siempre ocurren juntos, en lugar de esparcidos a través de la cláusula (criterio de cohesión). También que ocurren en orden fijo y finalmente tienen una coherencia y significado convencional. (Dixon y Aikhenvald 2003: 18-19).

2.2.2 Morfema

El morfema es definido como la unidad mínima más pequeña e indivisible de contenido semántico o función gramatical de las que se componen las palabras (Katamba 1993: 20). Un ejemplo dado por Payne (1997) ilustra esta definición. La palabra inglesa *dogs* ‘perros’ contiene dos morfemas: *dog* que encierra el contenido semántico principal de la expresión, y el sufijo *-s* que encierra el significado de pluralidad. La forma *dog* en sí misma no puede subdividirse en más unidades significativas, por ello es un morfema, una forma mínima (Payne 1997: 11).

Los morfemas se suelen clasificar en morfemas libres y morfemas ligados (Crystal 2000: 373). Un **morfema ligado** es un morfema que debe anexarse a algún otro morfema para poder quedar integrado de forma natural en el discurso (Payne 1997: 13).

Del ejemplo dado por Payne (1997) con la palabra *dogs*, el morfema *-s* es un morfema ligado ya que nunca puede pronunciarse por sí mismo. Por otro lado, *dog* es un **morfema libre** debido a que no tiene que unirse a alguna otra forma. (Payne 1997: 13). Una palabra que consta de un solo morfema (libre) es una palabra monomorfémica; su opuesto, es decir, una palabra que contiene dos o más morfemas es polimorfémica. (Crystal 2000: 373). Las palabras polimorfémicas están compuestas por lo general de una raíz y de uno o más afijos (a no ser que se trate de compuestos, que se obtienen mediante la combinación de dos o más raíces y no serán tratados en esta tesis).

Katamba (1993: 41) define a la raíz como el núcleo irreducible de una palabra. Payne (1997: 16) añade que dicha unidad es una forma no analizable y expresa el contenido léxico básico de una palabra. Una ilustración de esta idea la podemos tomar del inglés con el lexema WALK, que constituye la raíz de las palabras *walk*, *walks*, *walking*, *walked* (Katamba 1993:41). La raíz *walk* está presente en cada una de las palabras en forma íntegra pero ¿puede estar presente con alguna modificación en las diferentes manifestaciones del lexema? (Katamba 1993:41). Estos casos son llamados “supletivos” (ver sección §2.2.3.2). Un ejemplo de suplección es el lexema del inglés GOOD que tiene dos formas de palabras, a saber, *good* ‘bueno’ y *better* ‘mejor’. Ambas representan el mismo morfema pero no comparte una raíz morfológica común, pues sólo *good* es fonéticamente similar a GOOD.

Es importante mencionar que las raíces no son siempre morfemas libres. Por el contrario, existen raíces que no pueden constituir palabras totalmente comprensibles en sí mismas. A estas se les puede llamar raíces ligadas, tal como ocurre con los verbos del castellano, los mismos que necesitan más morfología para poder ser enunciados (*am- > amo, amas*, etc.) Entonces, las raíces ligadas requieren de alguna operación flexiva que involucre a un prefijo o a un sufijo (Payne 1997: 16).

Muchas raíces son **morfemas léxicos**, en tanto que llevan gran parte del contenido semántico de las palabras (*buey, pan, am- ‘amar’, com- ‘comer’, cruel, mal, hoy*) y pueden ser sustantivos, verbos, adjetivos, y adverbios (Katamba 1993: 41). Sin embargo, las raíces pueden ser también morfemas funcionales, tal como ocurre con los artículos o las preposiciones del castellano (Katamba 1993: 41-42).

2.2.3 Afijos

Un afijo es un morfema que solo ocurre cuando está unido a una raíz o tema (base) (Katamba 1993: 44). Por ello, los afijos son siempre morfemas ligados (Crystal 2000: 40). En una lengua, los afijos constituyen un grupo limitado y generalmente se clasifican en tres tipos dependiendo de su posición con respecto a la raíz de la palabra: prefijos, sufijos e infijos (Crystal 2000: 40). Payne (1997: 23) describe estos mismos tres procesos morfológicos, aunque añade un cuarto tipo de afijo: el suprafijo.

Los **prefijos** son afijos que se adjuntan al inicio de la raíz o base. Como lo vemos en (4), se adjuntan los prefijos *re-* en (4 a), *in-* en (4.b) y *zh-* en (4 c) a las raíces de las palabras en inglés, español y chipaya, respectivamente:

- | | | | |
|--------|-------------|-------------------------------|------------------|
| (4) a. | del inglés | <i>re-make</i> | ‘rehacer’ |
| b. | del español | <i>in-feliz</i> | ‘infeliz’ |
| c. | del chipaya | <i>zh-like-a</i> ³ | ‘¡cómete (eso)!’ |

Los **sufijos** son los afijos o morfemas ligados que se adjuntan después de la raíz o base. El ejemplo (5) muestra la sufijación de los morfemas ligados *-ed* en (5 a), *-ción* en (5 b) y *-wa* en (5.c) a las raíces de las lenguas inglés, español y aimara respectivamente:

- | | | | |
|--------|-------------|--------------------|--|
| (5) a. | del inglés | <i>walk-ed</i> | ‘caminó’ |
| b. | del español | <i>obliga-ción</i> | ‘obligación’ |
| c. | del aimara | <i>jani-wa</i> | ‘(me consta que) no’ (Cerrón-Palomino 2008: 103) |

³ El prefijo verbal *zh-* establece una referencia no-personal y está en franca obsolescencia en su lengua (Cerrón-Palomino 2006: 109 y 158).

La **infijación** consiste en la inserción de un afijo en la raíz misma o en los temas, como lo describen Katamba (1993: 44) y Payne (1997: 23). La infijación no es muy común en las lenguas romances, pero sí en las lenguas semíticas como el árabe y el hebreo y en las lenguas austronesias. A continuación, en el ejemplo (6 a) se le inserta el infijo *-um-* a la raíz *fikas* ‘ser fuerte’ del bontok, asimismo *-it-* a la raíz *Carlos* del español en (6 b).

(6) a. bontok (Filipinas) f-um-ikas ‘convertirse en fuerte’

b. español Carl-it-os ‘Carlitos’

Finalmente, la **suprafijación** o modificación de suprasegmentos es aquella en que el patrón tonal o acentual de una palabra señala una distinción semántica particular (Payne 1997: 23). El ejemplo (7 a) muestra el proceso de suprafijación en el que el patrón acentual distingue en el inglés nombre de verbo, y en (7 b) en el chakobo, la relación gramatical de nombres bisilábicos (Valenzuela e Iggesen 2007: 187).

(7) a. inglés *pérmit* ‘permiso’ *permit* ‘permitir’

b. chakobo joni ‘persona: absoluto’ joní ‘persona: ergativo’

2.2.4 Clíticos

El término clítico es definido por diversos autores como “un elemento morfológico que no tiene el conjunto completo de propiedades de una palabra independiente y que forma una unidad fonológica con la palabra que le precede” (Matthews 1997: 56, citado en Aikhenvald 2003: 42) o le sigue (ver también Anderson 2005: 10).

Dixon y Aikhenvald (2003: 25) recurren a la distinción entre palabra fonológica y palabra gramatical y definen a los clíticos como “palabras gramaticales que no

- (9) Jawen jema-*ra* ani iki.
 POS3 pueblo.ABS-EV grande COP
 ‘Su pueblo es grande.’ (Valezuela 2003b: 534 ejemplo 7b)

b) Selectividad y tipo de huésped

Los clíticos pueden adjuntarse a casi cualquier clase de palabras, o pueden ser selectivos en términos de la clase gramatical del *host*. Pueden adjuntarse a las frases verbales o a las frases nominales. También pueden ser no selectivos y adjuntarse a más de un tipo de frase (Aikhenvald 2003: 44).

Aikhenvald (2003: 45) menciona además que la posición de un clítico puede ser fija o libre. El primer grupo, el de los **clíticos de posición fija**, incluye clíticos de segunda posición (o Wackernagel) (Anderson 1992: 2000) o clíticos que aparecen al final o al principio de un constituyente (oración, cláusula o frase). En el caso del SK, encontramos clíticos de segunda posición como los enclíticos evidenciales *-ra* ~ *r-* ‘evidencial de primera mano’. El siguiente ejemplo (10) corrobora la segunda posición del evidencial de primera mano *-ra*, el que ocurre después de una coordinación de los sujetos de una frase nominal (Valenzuela 2003b:537 ejemplo 3)

- (10) [Papa betan tita]-*ra* wainko kabékonai
 ‘Padre y madre (-RA) van a la chacra.’

Por otro lado, los **clíticos flotantes** son aquellos que se unen a un constituyente particular del discurso, sin que medie ningún principio posicional (Dixon y Aikhenvald 2003: 46), como ocurre con los clíticos *-ribi* ‘también’ o *-bi* ‘enfático’ del SK.

2.2.5 Variación morfológica

2.2.5.1 Alomorfía

Los morfemas son unidades abstractas que se realizan en el habla mediante unidades discretas, conocidas como morfos. La relación suele recibir el nombre de exponencia o de realización. La mayoría de los morfemas se realizan mediante un único morfo, como algunos de los casos que hemos visto hasta este punto. No obstante, algunos morfemas se realizan mediante más de un morfo según su posición en una palabra u oración; tales morfos alternativos reciben el nombre de alomorfos o variantes morfémicas. Así pues, el morfema de {plural} en *pisos* y *panes* tiene los alomorfos representados fonéticamente por [-s] y [-es], respectivamente. En este caso, los alomorfos son el producto de la influencia del contexto fonético que presentan las formas del singular (una termina en vocal y la otra, en consonante) y el proceso se denomina de ‘condicionamiento fonológico’. La alomorfía es la variación de la forma de una unidad lingüística que no afecta su identidad funcional en la lengua. Es decir, la variación formal percibida no es lingüísticamente distintiva y no implica un cambio de significado (Crystal 2000: 45).

En el ejemplo (11) del inglés el sufijo plural -s presenta tres variantes, las formas [-s] en (11 a), [-əz] en (11 b) y [-z] en (11 c). Decimos que estas formas son alomorfos del morfema {plural} debido a que manifiestan una variación formal, pero esta variación no implica un cambio en su significado ni en su identidad funcional.

(11)	a.	[-s]	boy	boy-s	[ˈbɔɪ-s]
	b.	[-əz]	church	church-es	[ˈtʃɜː.tʃ-əz]
	c.	[-z]	dog	dog-s	[dɒg-z]

En quechua de Cusco el sufijo reportativo de primera mano *-mi* tiene dos variantes alomórficas formales: *-m* y *-mi*. El siguiente ejemplo (12 a) se adjunta *-m* después de vocal como en *runa* ‘hombre’ y se realiza *-mi* (12 b) después de consonante como en *p'unchaw* ‘día’ (Cerrón-Palomino 2008: 167).

- (12) a. *-m* *runa-m* ‘hombre’
 b. *-mi* *p'unchaw-mi* ‘día’

Causas de la alomorfía

Las causas identificadas por Varela (2005) como disparadoras de la alomorfía son de cinco tipos: fónica, morfofonológica, de acuerdo a la posición dentro de la palabra, gramatical y léxica.

La mayor parte de los alomorfos están condicionados por la fonología, es decir, es el contexto fónico el que determina una variante formal u otra. En el ejemplo (13) en español los afijos *-al* y *-ar* son derivadores adjetivales sobre bases nominales que al significado colectivo unen al del lugar (Varela 2005: 51). Si la palabra contiene < l > o < ll > el morfema en cuestión se realiza, por lo general, con el alomorfo *-ar* como se ve en (13 a); si no, se suele imponer el alomorfo *-al* como muestra (13 b) (Varela 2005: 23).

- (13) a. *-ar* *Oliv-ar* Contiene < l > o < ll >
 b. *-al* *Naranj-al* En los otros casos

En el ejemplo (14 a) del español el afijo *-ble* cuando está a final de palabra se realiza como *-ble*, pero como *-bil* cuando no está a final de palabra, es decir, cuando aparece seguido de otro afijo como lo vemos en (14 b) (Varela 2005: 23).

- (14) a. *-ble* *Disponi-ble* Final de palabra

- i. **tengan el mismo valor semántico**, es decir, las formas alternantes aporten a la palabra compleja el mismo significado (*similitud semántica*).
- ii. **tengan un parecido formal o fónico** lo suficientemente próximo como para que el hablante, sin conocimientos de la historia de la lengua, los identifique justamente como variantes del mismo morfema (*similitud fónica*) (Varela 2005)

De cumplirse la similitud semántica, pero no la fónica, entonces podríamos definir la forma como un caso de suplección.

En el ejemplo (17) se observa el verbo presente *voy* en español que manifiesta suplección, es decir, pese a ser miembros de un mismo paradigma y tener un parecido semántico no tienen un parecido fónico en el tiempo pasado:

(17)	Presente	 voy
	Pasado	 fui

Por otro lado, de cumplirse sólo la similitud fónica y no la semántica, estamos ante un caso de homofonía. En los ejemplos (18) del español: el sufijo *-ería* no es el mismo para ambos casos, pero no significan lo mismo:

(18)	bob- <i>ería</i> chul- <i>ería</i> conserj- <i>ería</i> hamburgues- <i>ería</i>	}	Da al adjetivo de la base el significado de <i>cualidad</i>
	hamburgues- <i>ería</i>	}	Da al sustantivo un significado <i>locativo</i>

2.2.3.2 Suplección

Una definición aproximada de suplección nos la entrega Mel'puk: "Para que los signos *X* e *Y* sean supletivos, su correlación semántica debe ser al máximo regular, mientras que su correlación formal es al máximo de irregular" (citado por Corbett 2009). En los ejemplos (19) del ruso e inglés, Corbett distingue dos categorías de suplección que son, una compleja (19 a) y otra parcial (o no máxima) como en (19 b):

- (19) a. Compleja, del ruso *rebēnok* 'niño' *det-i* 'niños'
- b. Parcial, del inglés *child* 'niño' *children* 'niños'

En la suplección compleja la raíz de la palabra cambia totalmente, mientras que en la suplección parcial, se aprecia aún una semejanza formal.

Categorías supletivas

Las categorías gramaticales que pueden manifestarse mediante formas supletivas son muchas. A continuación presentaré una muestra de algunas de ellas de acuerdo a Corbett (2009), prestándole atención al número, al caso, a la persona, al tiempo y al género.

Las palabras que significan 'persona' en **singular** son supletivas a **pluralizarse**, como lo muestra el ejemplo (20 a) en ruso como en el inglés (20 b), son supletivas.

- (20) idioma singular plural
- a. ruso *čelovek* *ljud-i*
- b. inglés *person* *people*

Por otro lado, el latín exhibe marcación de **caso** supletivo del nombre *Ioppiter* ‘Júpiter’ en el caso nominativo (en negrita), como el siguiente ejemplo (21). Se observa que todos los demás casos (genitivo, dativo, acusativo y ablativo) del latín mantienen la raíz *Iov-*, sin embargo el caso nominativo es una forma supletiva y cambia a *Iuppier*.

(21)	CASO	FORMA
	NOM	Iuppier
	GEN	Iov-is
	DAT	Iov-i
	ACU	Iov-em
	ABL	Iov-e

La **persona** parece ser la categoría con mayor frecuencia supletiva. Veamos los siguientes ejemplos, primero de los pronombres del ruso (22) y luego en SK (23).

Se observa en el ejemplo (22) del ruso, que la declinación nominativa es supletiva para la tercera persona (en negrita), a diferencia de las otras declinaciones en los caso acusativo, genitivo, dativo, instrumental y locativo, que mantienen evidentes correspondencias formales entre sí. .

(22)	Caso	3 Mas-SG	3 Fem-SG	3 Neu-SG	3 PL
	Nominativo	on	ona	ono	oni
	Acusativo	<i>ego</i>	<i>ee</i>	<i>ego</i>	<i>ix</i>
	Genitivo	<i>ego</i>	<i>ee</i>	<i>ego</i>	<i>ix</i>
	Dativo	<i>emu</i>	<i>ej</i>	<i>emu</i>	<i>im</i>
	Instrumental	<i>im</i>	<i>ej / eju</i>	<i>im</i>	<i>imi</i>
	Locativo	<i>nem</i>	<i>nej</i>	<i>nem</i>	<i>nix</i>

En el SK, como muestra el ejemplo (23) la primera persona singular tiene una raíz común *e-* en sus forma (23 a) absoluta, (23 b) ergativa, y (23 d) interesiva. Sin embargo en (23 c) el posesivo es supletivo por tener una raíz diferente (en negrita), que es *nokon*:

- (23)
- | | | | |
|----|------------|---------------------|---|
| a. | absolutivo | <i>e-a</i> | ‘yo’ sujeto de verbo intransitivo u objeto de verbo transitivo. |
| b. | ergativo | <i>e-n</i> | ‘yo’ sujeto de verbo transitivo |
| c. | posesivo | <i>nokon</i> | ‘mío, mi’ |
| d. | interesivo | <i>e-on</i> | ‘a mi’ |

En lo que toca al **tiempo**, los verbos de movimiento presentan con frecuencia patrones supletivos. El ejemplo (24) se enlista el presente y pasado del verbo *ir*, en (24 a) español presente *voy* y en pasado *fui*, en (24 b) del ruso en presente *idit* y en pasado *šel*, y en (24 c) inglés el presente *go* y en pasado *went*, cuyas raíces son diferentes.

- (24)
- | | | | |
|----|---------|-----------------------|--------------------|
| a. | español | presente <i>voy</i> | pasado <i>fui</i> |
| b. | ruso | presente <i>id-it</i> | pasado <i>še-l</i> |
| c. | inglés | presente <i>go</i> | pasado <i>went</i> |

Finalmente, si la lengua distingue **género** masculino, femenino o neutro, puede tener suplección como lo manifiesta el siguiente ejemplo (25) del griego clásico en los numerales. En este ejemplo observamos una raíz diferente para (25 a) masculino/neutro *év-* [en] y otra raíz (25 b) completamente diferente *μiα-* [mia] para el femenino.

- (25)
- | | | | |
|----|--------------------|-------------------|-------|
| a. | Masculino y neutro | <i>év-</i> [en] | ‘uno’ |
| b. | Femenino | <i>μiα-</i> [mia] | ‘una’ |

2.3 LA TRADUCCIÓN AUTOMÁTICA EN LENGUAS MINORITARIAS

Los hablantes del castellano tenemos a nuestra disposición una serie de recursos computacionales que nos brindan toda clase de información, como también entretenimiento en la web. Estos recursos digitales nos ayudan a conocer nuestro idioma a través, por ejemplo, de diversos textos educativos, de entretenimiento, judiciales, diccionarios, por nombrar algunos. Dichos recursos en castellano ayudan a registrar nuestra lengua, preservarla y traducirla a diferentes otros idiomas. Este es un privilegio con el que cuentan el castellano, lengua de amplia difusión, junto a otras lenguas indoeuropeas, pero que no cuentan otros idiomas del planeta debido a su falta de recursos digitales. Berment (2002) señala que menos del 1% de las lenguas que se hablan en el mundo están correctamente “computarizadas”. Al 99% restante⁴ se les llaman idiomas de baja densidad o *resource poor languages* (RPL) ‘lenguas de escasos recursos’, esto es, falta de datos digitalizados (Xia y Lewis 2008:1). La brecha entre los “lenguajes principales” que cuentan con un creciente aumento de recursos computacionales versus las lenguas minoritarias, que no los tienen, es un problema que enfrentan las regiones con una alta recepción de inmigrantes (Somers 1997). La necesidad de traducción, por ejemplo, de lenguas asiáticas, debido a la llegada de refugiados es una realidad que sufre Europa. Idiomas como el bengalí, el hindi, el urdu, el tamil y el cingalés, entre otros, son considerados por ellos como “lenguas minoritarias no indígenas” (NIML) y países como los que componen el Reino Unido han desarrollado proyectos de construcción de corpus electrónico para lenguas del Asia que se hablan en sus territorios (Berment 2002) buscando con ello solucionar problemas prácticos de comunicación. Las necesidades comunicativas de los migrantes son variadas, y van por un lado, desde la traducción de textos impresos de

⁴ Este 1% representa una baja densidad lingüística, pero una alta cantidad de hablantes, debido a que son las lenguas más habladas del mundo, a saber, inglés, chino, alemán, español, francés.

procesamientos legales, la provisión de servicios sociales, hasta el apoyo de traductores específicos en puestos de salud (Somers 1997). Esta problemática europea con las lenguas minoritarias no indígenas, también la vivimos en nuestro contexto sudamericano y específicamente con las lenguas nativas en el Perú. Al igual que los inmigrantes en Europa, para muchos nativos peruanos la “era de la información” está limitada para ellos, pues el uso de softwares y del internet, no satisfacen sus necesidades en términos de lenguaje y traducción automática (Bermet 2002).

2.3.1 Las lenguas minoritarias desde la perspectiva de la computación

En el Perú existen más de 47 lenguas originarias y cerca de cuatro millones de hablantes tienen una lengua nativa como L1 (Ministerio de Educación 2013). Todas estas lenguas son consideradas lenguas minoritarias, de acuerdo a las características que Forcada (2006) expone. Entre esas lenguas nativas del Perú, en la región de la Amazonía está el SK. Para poder entender por qué el SK es un idioma minoritario o de escasos recursos desde la perspectiva de la computarización, revisaremos las características que define Forcada.

Primeramente, una lengua minoritaria tiene un pequeño grupo de hablantes que saben leer y escribir en su idioma. El SK cuenta con una población de cerca de 22.500 hablantes y una asimetría en su literacidad, debido a que la población shipiba alfabetizada en su L1 es menor al 5% mientras que en castellano es cercana al 25%. Ambas cifras son bajas, considerando que la alfabetización en la población peruana nacional es del 94% (Ethnologue 2018). La segunda característica de una lengua minoritaria es que carece de un sistema de escritura único, una ortografía estable o una variedad estándar del idioma. Respecto a este punto, el SK logró un sistema ortográfico estable sólo desde el año 2007 cuando cambió su ortografía a la que tenía desde el año 1999 y a otras anteriores

(MINEDU 2013). Este cambio de ortografía perjudicó los textos publicados por el ILV antes de la norma, como el Diccionario Castellano Shipibo-Conibo (Loriot et al. 1993), la primera y segunda edición del Nuevo Testamento *DIOSEN JOI* (ILV 1983 y 2008), las “Lecciones para el aprendizaje del idioma shipibo-conibo” (Faust 1973), todos actualmente desactualizados⁵. Finalmente, las lenguas menores carecen de presencia en internet, de experiencia lingüística y de recursos legibles, como datos lingüísticos, corpus, etc. La realidad de esta lengua amazónica es que cuenta con cerca de 200 documentos en este idioma o referentes al SK, según Glottolog y de acuerdo con Ethnologue, cerca de 80 items lingüísticos⁶. Por lo tanto, el SK es un idioma que muestra una desventaja computacional frente al castellano y esta desventaja es también compartida por la gran mayoría de las lenguas originarias del Perú.

El camino para computarizar y preservar la lengua es una meta obligatoria y para ello se requieren múltiples trabajos (Berment 2002). Estos trabajos deben mirar a que los datos digitalizados puedan ser utilizados por la lingüística computacional.

En el siguiente apartado describiré las áreas involucradas en la preparación de los “datos” necesarios para digitalizar una lengua y para que estos datos puedan utilizarse en la traducción automática.

2.3.2 La lingüística computacional y el procesamiento del lenguaje natural

Hay muchas disciplinas relacionadas con la lingüística y las tecnologías computacionales, ellas son, por ejemplo, lingüística computacional, la lingüística informática, el procesamiento del lenguaje natural (NLP en inglés), la inteligencia artificial (IA), la

⁵ Solo el Nuevo Testamento fue actualizado publicándose el año 2012 con la nueva norma ortográfica.

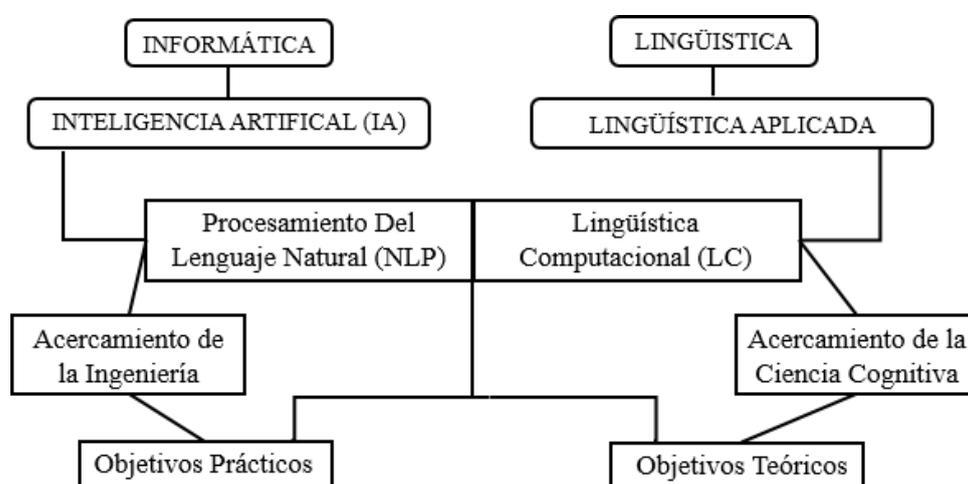
⁶ Artículos de investigación, de descripción de la lengua y textos en SK.

ingeniería lingüística, la industria de la lengua, etcétera. Todas ellas comparten el mismo objeto de estudio, esto es, el lenguaje, y difieren en la perspectiva que adoptan a la hora de abordarlo (Villayandre 2000: 204). Es por este motivo que creemos necesario identificar las áreas relevantes y que interactúan con el TA. Definiré los términos de lingüística computacional (LC), inteligencia artificial (IA) y el procesamiento del lenguaje natural (NLP) y delimitaré algunos de estos campos, con el objetivo de guiar a una mejor comprensión de las **disciplinas** involucradas y el trabajo interdisciplinario que conlleva la **realización** de un traductor automático (TA).

Comenzando por la **Inteligencia Artificial (AI)**, Villayandre afirma que “es la ciencia que trata de que las máquinas hagan la clase de cosas que hace la mente humana” y la define como “una de las subdisciplinas de la Informática. Su objetivo es diseñar ingenios artificiales que simulen el comportamiento inteligente humano”. Se refiere a los ingenios artificiales como a la “codificación en un programa informático las facultades cognitivas” (Villayandre 2010: 61).

Respecto de la **Lingüística Computacional (LC) y el Procesamiento del lenguaje natural (NLP)**, para Sandoval son sinónimos, pues ambos tratan el desarrollo de programas de computación que simulan la capacidad lingüística humana (Moreno Sandoval 1998 citado por Villayandre 2010: 205). Sin embargo, difieren en las disciplinas de donde proceden, situando a la LC vinculada al campo de la lingüística y el NLP al de la informática, como en la figura siguiente. Muchas veces ambos términos son usados indistintamente (Trask 1993:53 citado por Villayandres 2010). La Figura 3 esquematiza la propuesta de Sandoval. En ella se observan los dos grandes campos: la Informática y la Lingüística. De cada una de ellas se desprenden, por el lado de la informática, la Inteligencia artificial, y por el lado de la lingüística, la Lingüística Aplicada. Ambas convergen en dos disciplinas: la NLP y LC respectivamente.

Figura 3 Disciplinas y funciones en la TA



Otra buena definición que vincula la LC con el NPL es la entregada por Cunningham: “la lingüística computacional es parte de la ciencia del lenguaje que usa las computadoras como herramientas de investigación, mientras que el procesamiento natural del lenguaje es parte de la ciencia de la computación cuyo tema principal son los sistemas informáticos que procesan el lenguaje humano o natural” ((Cunningham 1999:4): citado por Villayandre 2010: 61).

Esto explica la necesidad de un equipo interdisciplinario para el proyecto, formado por el área de informática con ingenieros especializados en Inteligencia artificial y por otro lado lingüistas, reunidos en donde convergen ambos campos en la LC o en NLP.

Entonces, los lingüistas especializados en LC se encargan de la tarea de “preparar” el modelo lingüístico para que los ingenieros informáticos lo implementen en un código eficiente y funcional. Hay dos amplias categorías de enfoques en la morfología computacional: basada en reglas y corpus. El modelos lógico basado en gramáticas o

reglas consisten en la escritura reglas de reconocimiento de patrones estructurales, empleando un formalismo gramatical concreto. Estas reglas de traducción (RBMT) en combinación con la información almacenada en diccionarios morfológicos y bilingües, gramáticas y archivos de reglas de transferencia estructural, definen los patrones que hay que reconocer para resolver la tarea (buscar información, traducir, etc.).

Por otro lado, los **modelos probabilísticos del lenguaje natural** basados en datos tienen una aproximación a la inversa: los lingüistas recogen colecciones de ejemplos y datos (corpus paralelos alineados) y a partir de ellos se calculan las frecuencias de diferentes unidades lingüísticas (letras, palabras, oraciones) y su probabilidad de aparecer en un contexto determinado. Calculando esta probabilidad, se puede predecir cuál será la siguiente unidad en un contexto dado, sin necesidad de recurrir a reglas gramaticales explícitas. (Sandoval 2018).

Este método es el paradigma de “aprendizaje automático” que se ha impuesto en las últimas décadas en Inteligencia Artificial: los algoritmos infieren las posibles respuestas a partir de los datos observados anteriormente en el corpus, pero que en el caso de las lenguas minoritarias, estos corpus son difíciles de obtener y una vez que se obtienen son difíciles de preparar debido a las altas cantidades de texto en paralelo se necesitan para obtener resultados razonables (Zariquey 2018).

Berment (2002) propone en sus soluciones el reciclaje de diccionarios y la digitalización mediante OCR (reconocimiento óptico de caracteres), de los pocos textos existentes. El procedimiento consiste en la selección, la preparación (digitalización, corrección manual y traducción) y la alineación (juntar una frase en SK con su traducción al español) de un corpus paralelo de diferentes fuentes (Galarreta et al. 2017).

Para el proyecto Chana, se utilizó el diccionario Shipibo-Castellano (Loriot et al. 1993), textos educativos y textos religiosos. La alineación fue posible porque el diccionario contiene por cada entrada, una definición y por lo menos, dos oraciones en SK con su traducción, pero debido al cambio ortográfico descrito anteriormente, y a la imprecisión de la lectura OCR, hubo que revisarlo y corregirlo manualmente. En el caso de la Biblia, pudo alinearse cada versículo con una traducción en castellano. Se alinearon semiautomáticamente casi diez mil versículos con una biblia en español. El resultado de este trabajo se transformó en más de 13 mil reglas de traducción en el sentido SK-castellano para el diccionario y casi catorce mil frases de la Biblia en SK. De los textos educativo se extrajeron cerca de dos mil oraciones (Galarreta et al. 2017).

Otro de los métodos diferentes recomendados para comenzar un traductor en lenguas de escasos recursos, es el uso de analogía de idiomas, es decir, la adaptación de TA entre lenguas con similitud sintáctica o entre familias lingüísticas, como lo han podido realizar entre el español y el gallego y el catalán, y el japonés con el uigur que cuentan con una cercanía sintáctica (Berment 2002). Desgraciadamente este proceso no pudo aplicarse totalmente con el SK, al ser pionera en TA dentro de las lenguas amazónicas a las que pertenece y más bien, los trabajos en SK podrán favorecer en un futuro a las demás lenguas de la familia Pano. Por otro lado, Xian y Lewis (2008) recomiendan suplir la falta de *corpus* anotados y datos paralelos traducidos, mediante el uso del Texto interlineal glosado (IGT), que es un texto en el idioma meta glosado sintácticamente y traducido a una lengua de amplio uso. Ellos proponen la formación de una base de datos de texto interlineal, *Online Database of INterlinear text* ODIN, en sus siglas en inglés (Xian y Lewis 2008)

2.3.3 La morfología en la traducción automática

Es de gran importancia el análisis morfológico para las lenguas aglutinantes, primero por la complejidad de su variada morfología y por la importante información lingüística para su posterior procesamiento. El análisis morfológico juega un papel crítico en el desarrollo de aplicaciones de procesamiento de lenguaje natural (NLP). En la mayoría de las aplicaciones prácticas de la tecnología del lenguaje, el análisis morfológico se utiliza para realizar la segmentación de morfemas y como tarea final la identificación de características morfológicas. En los lenguajes morfológicamente complejos, el análisis morfológico también es un componente central en la recuperación de información, resumen de texto, respuesta a preguntas, traducción automática, etc. (Abate y Assolbie 2014: 1)

El anotador es la herramienta central en la generación de corpus paralelos de las lenguas aglutinantes. Sin embargo, la anotación de metadatos es una tarea manual muy laboriosa. Se requiere el uso de herramientas computacionales que simplifiquen algunas de las operaciones y a la vez que puedan ajustarse de manera apropiada a las necesidades particulares del idioma (Mercado et al. 2018).

El proyecto Chana utilizó un anotador propio, ChAnot, una herramienta de anotación basada en la web desarrollada para los idiomas indígenas y altamente aglutinantes peruanos, Esta herramienta es capaz de soportar un conjunto diverso de tareas de anotación lingüística, como el análisis morfológico, que son lema y segmentación morfológica, etiquetado de voz parcial, reconocimiento de entidad de nombre y análisis sintáctico (utilizando una interfaz BRAT1 (Stenetorp et al., 2012)) (Mercado et al. 2018).

3. HACIA UNA TIPOLOGÍA DE LA VARIACIÓN FORMAL EN LA FORMACIÓN DE PALABRAS

Para poder tipologizar la variación morfológica en SK propongo cuatro criterios: nivel de condicionamiento, variabilidad formal, predictibilidad y semejanza formal. Definiré cada una de ellas a continuación. Tal como se verá, estos criterios podrían tener aplicaciones tipológicas que trasciendan la descripción del SK y puedan ayudar a la lingüística computacional con la problemática que se presentan a nivel morfológico cuando las lenguas altamente aglutinantes.

3.1 NIVEL DE CONDICIONAMIENTO

Por nivel de condicionamiento nos referimos al ámbito gramatical en el que encontramos las causas o condicionamiento que producen la variación y que regulan su alternancia. Según Varela (1996) estos niveles de variación son el fonológico, el morfofonológico, gramatical y léxico (1996: 64). A continuación detallaré cada uno de ellos (ver también §2.5.1.1).

3.1.1 Dominio fonológico

Los fenómenos de variación fonológicamente motivados son aquellos que se producen a partir de un contexto fónico independiente a los morfemas en juego. Se trata por lo general de procesos regulares y automáticos (Varela 1996: 65). El ejemplo (26) del español muestra que el morfema negativo *IN-* tiene los siguientes alomorfos /*im-*/, /*in-*/, /*in̄-*/e /*in̄j-*/, producto de la asimilación del punto de articulación de la consonante inicial de la palabra que se prefija:

- (26) *im* *im*-posible Se asimila a la zona labial de la consonante /p/
imj *imj*-faltable Se asimila a la zona labio dental de la consonante /f/
in *in*-tachable Se asimila a la zona alveolar de la consonante /t/
inj *inj*-korexible Se asimila a la zona velar de la consonante /k/

El siguiente ejemplo (27) es tomado del griego clásico. Las raíces $\lambda\alpha\beta$ - [lab] y $\mu\alpha\theta$ - [maθ] sufren la inserción de una consonante nasal en el tiempo presente indicativo y en el imperfecto, respectivamente. Esta consonante nasal incorporada varía asimilándose al punto de articulación de la consonante que le sigue. En el caso de $\lambda\alpha\beta$ - se incorpora la nasal labial μ motivada por la labial β y en el de $\mu\alpha\theta$ - se incorpora la nasal alveolar ν motivada por la alveolar θ (Berenguer 1999: 153).

- (27) $\lambda\alpha\beta$ - $\lambda\alpha$ - μ - $\beta\alpha\nu\omega$ ‘tomo, cojo’ : la nasal se vuelve μ [m] junto a β [b]
 $\mu\alpha\theta$ - $\mu\alpha$ - ν - $\theta\alpha\nu\omega$ ‘aprendo’ : la nasal se vuelve ν [n] junto a θ [θ]

En ambos casos, (26) del español y (27) del griego, vemos el mismo proceso fonológico, esto es, la asimilación de punto de asimilación por parte de una nasal en base a la consonante contigua.

Los ejemplos vistos hasta ahora son de variación condicionada fonológicamente, es decir, a nivel de segmento. También hay variación condicionada al nivel prosódico o suprasegmental, como el siguiente del SK

El sufijo verbal *-katit* ‘pasado distante’ varía su forma de acuerdo a la cantidad de moras que tenga el verbo al que se adjunta. Así, después de verbos con un número par

de moras, se realiza como *-katit-ai*. Pero, si la cantidad de moras del verbo es impar, entonces se manifiesta como *-kati-ai*. Los ejemplos de (28) muestran su alomorfo en (28 a) con un verbo *rete-* ‘llegar’ de dos moras, esto es, con número par de moras y en (28 b) el verbo *pi-* ‘comer’ que tiene una mora, es decir, un número impar de moras (Valenzuela 2003b: 136 ejemplo 61)

- (28) a. *rete-katit-ai* ‘asesinado hace muchos años (incompletivo)’
b. *pi-kati-ai* ‘comimos hace muchos años (incompletivo)’

3.1.2 Dominio morfofonológico

El dominio de variación morfofonológica se relaciona con el fonológico visto anteriormente, debido a que su contexto y el cambio producido pueden establecerse en términos fonológicos, pero no son fruto de la aplicación de una regla fonológica automática y regular; sino que son motivados por morfemas específicos en contextos fónicos determinados (Varela 1996: 65). Como lo dice Cerrón-Palomino, los procesos morfofonológicos son aquellos en que los procesos fonológicos dependen de la estructura interna de los morfemas o de las formas gramaticales de la lengua. Agrega además que la morfofonología es “la bisagra que relaciona estrechamente los niveles fonológico y gramatical de una lengua. Así, pues, tales procesos requieren, para su aplicación, de una información gramatical previa, a diferencia de los procesos fonológicos que tienen una aplicación mecánica y hasta cierto punto ‘ciega’ respecto del resto de la gramática” (2008: 78). Agrega que “las lenguas aglutinantes, que hacen uso prolijo de afijos, son teóricamente más proclives a presentar una morfofonémica compleja; y son también importantes los factores supra-segmentales de acento, ritmo y entonación” (Cerrón-Palomino 2008: 78).

El ejemplo (29) es del quechua: los sufijos verbales *-yku* ‘acción esmerada’ y *-rqu* ‘acción terminante’ cambian su vocal *u* en *a* cuando van seguidos, directa o indirectamente, de los sufijos direccionales *-mu*, y *-pu*. (Cerrón-Palomino 2008: 86).

- (29) *-yku* *apa-yka-pu-sqa-n-ku* * *apa-yku-pu-sqa-n-ku* ‘ellos lo estaban llevando’
-rqu *apa-rqa-mu-y* * *apa-rqu-mu-y* ‘¡tráelo de una vez!’

3.1.3. Dominio morfosintáctico

De acuerdo a Varela (2005: 66): “Los alomorfos de dominio morfosintáctico no son automáticos ni están regulados por su fonética. Se trata de alternancias que se rigen por rasgos de índole gramatical, como las categorías sintácticas o gramaticales de un determinado morfema”. Dentro de este grupo, se incluyen también aquellas alternancias que están reguladas por factores morfológicos como la clase conjugacional del verbo como lo veremos en el ejemplo (30) del español. El morfema del pretérito imperfectivo de la primera y la tercera persona tiene dos alomorfos, *-ba* e *-ía*, que alternan de acuerdo al modelo de conjugación del verbo (Varela 2005: 66).

- (30) *-ba* *ama-ba* Primera conjugación
-ía *corr-ía* Segunda conjugación
ven-ía. Tercera conjugación

En este siguiente ejemplo (31) del SK tomamos una variación motivada por el dominio sintáctico.

Pero también nos encontramos con alomorfas que no podemos clasificar en ninguno de los cuatro niveles lingüísticos descritos anteriormente.

3.1.5 Variación sin motivación aparente (variación inmotivada)

Decimos que un alomorfo se comporta sin motivación aparente o con variación inmotivada, cuando se sustituye un morfema por otro **en un determinado contexto** sin ningún condicionamiento lingüístico⁷ y sin cambio de significado de la palabra (Crystal 2000: 338).

Un ejemplo de variación inmotivada del SK, son los alomorfos del evidencial especulativo *-mein*, que se realiza también como *-main* indistintamente. Vemos en la primera oración (32) que el especulativo se sufixa al verbo *pi-* ‘comer’ (después de *-ax*) y pierde su nasal final. En (33) el especulativo se adjunta a *bi* ‘tomar, coger’ y cambia su forma ahora a *-main*. (Valenzuela 2003b: 548 ejemplo 25 y 26 respectivamente).

- (33) Nato jiwí bimi pi-ax-**mei** (-ra) sipan
 Este árbol fruta:ABS comer-PSSS-**SPECL**(-EV) sardinas:ABS
 noya-ke.
 volar-CMPL
 ‘Habiendo comido la fruta de este árbol, ¿quizás volaron las sardinas?’

- (34) Nato joxo koríki miin-kan-ni
 Este blanco dinero:ABS enterrar-PL-REM:NMLZ

 bi-ax-**main**, no-a yometso i-ti iki.
 tomar-PSSS-**SPECL** 1P-ABS volverse.ladrón hacer.I-INF AUX

‘Parece que si tomamos esta moneda que fue enterrada en la antigüedad, podríamos resultar ladrones’.

⁷ Sin embargo, algunos lingüistas atribuyen esta inmotivación a factores sociolingüísticos, como cuando se habla enfatizando, o un estilo de habla cuidado versus un estilo informal (Crystal 2000: 339).

En resumen, en lo que respecta a los dominios lingüísticos donde se produce la variación, es posible establecer los siguientes tipos de variación por factor condicionante: fonológico, morfofonológico, morfosintáctico, léxico e inmotivado. Todos estos dominios resultan relevantes para el SK, salvo el léxico.

3.2 VARIABILIDAD FORMAL

La variabilidad formal o complejidad la podemos definir como todo lo *que se compone de elementos diversos*, y se opone a lo simple. Haré referencia, con este término a la cantidad de formas o alomorfos que tiene un paradigma de variación. La complejidad de dicho paradigma puede ser múltiple; podemos encontrar paradigmas con solo dos formas (binarios) o con muchas formas (incluso más de diez formas). En el SK en forma particular, encontramos cinco tipos de variabilidades formales.

En los siguientes ejemplos del SK se aprecian diferentes grados de variabilidad alomórfica. Por ejemplo, el sufijo verbal negador *-yama* tiene dos formas *-yama ~ -ama* mientras que el enclítico plural ergativo *-baon* tiene tres formas, *-baon ~ -boan ~ -boon*. El enclítico polifuncional *forma -n* tiene once formas, a saber, *-n ~ -en ~ -an ~ -in ~ -man ~ -kan ~ -tan ~ -ten ~ -nin ~ -ton ~ -tonin*.

Para esta tesis, hemos encontrado cinco tipos de variaciones formales en el SK (Ver § 4.2).

3.3 PREDICTIBILIDAD

Cuando un morfema presenta alomorfos, estos pueden ser totalmente **predecibles** si sus diversos condicionamientos son precisos y se cumplen sin excepción. Si los

condicionamientos donde se produce la alomorfía no son precisos, es decir, si no podemos determinar qué motiva la alternancia, entonces diremos que **no es predecible**.

Pero no todos los condicionamientos son totalmente predecibles o no predecibles. En la práctica, y según observé en el SK, los morfemas que presentan una alta variabilidad formal también muestran una **predictibilidad mixta**. En el ejemplo (35) un subconjunto de los once alomorfos del enclítico polifuncional *forma -n* se alternan entre sí en un condicionamiento fonológico específico. Este condicionamiento es el siguiente: todas las palabras con dos sílabas y con la segunda sílaba trabada con una consonante sibilante, se les epentiza una vocal /e/, /a/. Si la última sílaba contiene una /i/, entonces se le puede epentizar, esta misma vocal (Valenzuela 2003b:120). Este alomorfo puede tomar hasta tres formas en este determinado nivel de condicionamiento, pero la alternancia entre ellas es impredecible, pueden usarse por el hablante en forma indistinta. Vemos en (35 a) la palabra *kamox* ‘serpiente’⁸ en su forma absoluta. Para formar su ergativo entonces se le adjunta los alomorfos *-en* y *-an* indistintamente. Por otro lado, en (35 b) la palabra *mawis* ‘hormiga’ en su forma absoluta, cuenta con la vocal /i/ en su última sílaba, por lo que se le puede adherir los tres enclíticos de la forma *-n*, estos son *-en*, *-an* e *-in*.

- (35) a. *kamox* *kamox-en* ~ *kamox-an* ‘clase de serpiente’
b. *mawis* *mawis-en* ~ *mawis-an* ~ *mawis-in* ‘clase de hormiga’

Con el fin de clasificar en grados de predictibilidad en la variación morfológica, propongo una escala de tres niveles. A continuación presento los detalles de cada uno de ellos:

- i. **Totalmente predecible:** Un alomorfo es totalmente predecible si sus condicionamientos son precisos y se cumplen sin excepción. Por ejemplo,

⁸Una clase de serpiente muy venenosa conocida en la Amazonía como *shushupe*.

cuando los alomorfos se presentan en distribución complementaria. En el SK el clítico evidencial reportativo de primera mano *-ra* tiene alomorfía predecible en sus tres formas: *r-* solo aparece ante *iki*, como se muestra en el ejemplo (36), *-a* aparecen cuando la palabra termina en sibilante (36) y *-ra* en todos los otros contextos como lo vemos en (37) (Valenzuela 2003b: 534, 537).

(36) Nokon jane *r-iki* Inkan Soi.
 POS1 nombre.ABS **EV-COP** Inkan Soi.
 Mi nombre es Inka Soi. (Valenzuela 2003b: 536 ejemplo 2a)

(37) No-n papa nokó-ke-tian oin-ax-*a* no-a
 1P-GEN padre:ABS llegar-P-DS ver-PSSS-**EV** 1P-ABS
 raro-ke.
 volverse.feliz-CMPL
 ‘Cuando vimos que nuestro padre llegó, nos alegramos’
 (Valenzuela 2003b: 537 ejemplo 5)

(38) Jawen jema-*ra* ani iki.
 POS3 pueblo.ABS-**EV** grande COP
 ‘Su pueblo es grande.’ (Valenzuela 2003b: 534 ejemplo 7b)

- ii. **Parcialmente predecible:** ocurre cuando un alomorfo en un determinado nivel de condicionamiento puede tomar distintas alternancias y ninguna de ellas puede predecirse. Ejemplo de ello es uno de los grupos alomórficos del enclítico polifuncional *forma -n* como lo vimos en los ejemplos (35 a) y (35 b).
- iii. **Impredecible:** Nunca es posible saber el nivel de condicionamientos donde se produce la alomorfía. Estos casos de variación inmotivada lo vimos en §

3.1.5. En el SK, el clítico especulativo se manifiesta de dos formas y no se conoce aún su motivación. Los ejemplos (33) y (34) de § 3.1.5 nos muestra el clítico evidencial especulativo con sus dos alomorfos *mein* ~ *main*. También lo vemos en (39) y (40) con el sufijo nominal distributivo *-tibi* que alterna con *-tiibi* sin conocerse el condicionamiento que lo motiva. Por lo tanto los alomorfos *-tibi* ~ *tiibi* y el enclítico *-mein* ~ *-main* son impredecibles.

(39) Apo-n-ra jema-**tiibi** jawéki-bo meni-ke.
 Presidente-ERG-EV pueblo-**cada** cosa-PL:ABS dar-CMPL
 ‘El presidente hizo una donación a cada aldea.’

(40) Nete-**tibi**-ra e-a nashi-ai paro-n.
 Día-**todos**-EV 1-ABS bañar-INC río-LOC
 ‘Todos los días me baño en el río.’

3.4 SEMEJANZA FORMAL

Se define como semejanza formal o semejanza a la relación que sostienen aquellos alomorfos de un morfema que se asemejan o se parecen formalmente entre sí. En el caso de no haber ninguna semejanza formal entre ellos, podemos decir que es una suplección compleja, como lo vimos en § 2.2.3.2.

No hay un criterio establecido sobre los límites de la variación del alomorfo hasta convertirse en un paradigma supletivo. Algunos autores, como Corbett (2009: 21), discuten los límites de la suplección y se la plantean como un asunto de grados: “si la suplección es un asunto de grados, ¿cuál sería el punto final de una irregularidad? ¿Hay un claro corte entre los elementos que son simplemente irregulares por un lado y los elementos suplementarios por el otro lado?”

Con el fin de operativizar el criterio de semejanza, esto es, acercarme lo más razonablemente posible a un valor binario⁹, propongo una clasificación simple de los parecidos o diferencias entre los alomorfos de un paradigma. Por un lado, tenemos casos donde todos los alomorfos son muy parecidos (cuando varían solamente en relación a una consonante o una vocal) y, en el otro extremo, cuando el alomorfo cambia totalmente y no mantiene ninguna consonante o vocal de su otra forma. A continuación detallaré esta clasificación con la propuesta de tres categorías.

- i. **Todos parecidos entre sí:** decimos que todos son parecidos entre sí cuando el paradigma solo varía en una vocal o consonante. Ilustrando con el SK (41) encontramos que el sufijo verbal asociativo en su alomorfía varía en una sola vocal. En cambio, en (42) el prefijo partes del cuerpo ‘nalgas’ varía en su consonante.

(41) *-kin ~ -kiin* se parecen en toda su forma y varían solo en una vocal, se añade la /i/ a la segunda forma.

(42) *chi- ~ tsi-* mantienen la /i/ y varían en una sola letra, la consonante /ch/ que alterna con /ts/

- ii. **Parcialmente parecidos entre sí:** decimos que un morfema tiene alomorfos parcialmente parecidos entre sí cuando el paradigma varía en más de una vocal o consonante. Podemos ver el caso de esta variación parcial en el enclítico intensificador (43) que tiene dos alomorfos y varía en su primera sílaba CV

(43) *-yora ~ -(h)ira* varían la CV /yo/ e /hi/ y mantienen el parecido en *ra*

⁹ Considero que esta clasificación es casi binaria, pues la distancia entre “todos parecidos” a “parcialmente parecidos” depende de una sola letra. Por otro lado, el “ningún parecido” depende de que no haya ninguna similitud. Pensando en la programación computacional, estos valores son más sencillos de programar que una clasificación más elaborada.

- iii. **Ningún parecido entre sí:** las formas del paradigma no son derivables fonológicamente. Este es el caso de la suplección §2.2.3.2. Tenemos suplección en el ejemplo (43) con el sufijo verbal venitivo

(44) – *jo* ~ *-be* No son derivables entre sí.

Estos son los criterios que la presente tesis propone como básicos para la creación de una tipología de la variación morfológica en el SK. Los criterios seleccionados se condicen con el estudio de la variación en otras lenguas y por lo tanto podrían tener un potencial significativo para el desarrollo de una aproximación tipológica interlingüística al fenómeno de la variación y también para los fines de esta tesis, a saber, identificar los problemas a los que se enfrenta la traducción automática. En lo que sigue, emplearemos esta tipología para ofrecer un estudio sistemático de la variación en SK, a partir de los 30 casos de variación morfológica identificados para la lengua.

4. VARIACIÓN MORFOLÓGICA EN EL SK

En este capítulo analizaremos la variación morfológica en el SK a partir de un inventario de 30 morfemas que presentan paradigmas de variación de distinto tipo. Tal como ya se propuso en el §3, clasificaremos estos 30 casos a partir de una tipología de la variación morfológica basada en cuatro criterios que ya fueron explicados en dicho capítulo, y que forman parte de la propuesta de tipología de la variación morfológica presentada en esta tesis.

Metodología y datos

El inventario utilizado para esta tesis se obtuvo a partir de la revisión de la Tesis doctoral de Pilar Valenzuela (2003), del diccionario Shipibo-Castellano del ILV (Loriot, Lauriault y Day 1993) y los artículos de Pilar Valenzuela. Este inventario consta de 30 morfemas que muestran un total de 93 alomorfos (las listas completas, organizadas a partir de cuatro tablas de acuerdo a los cuatro criterios de organización de esta tipología, se presenta como anexo a esta tesis). A lo largo de este capítulo, presentamos para cada sección tablas estadísticas y gráficos circulares que ilustrarán la distribución de los distintos tipos de variación encontrados en nuestro corpus.

Debo aclarar que para realizar este estudio se tomaron dos grupos de datos de acuerdo a la necesidad de análisis de los cuatro criterios seleccionados. Para los criterios de *variación formal*, *predictibilidad* y *semejanza formal* se revisaron todos los alomorfos, esto es, las 93 variaciones formales que se recogieron en nuestros datos, agrupados a partir de los 30 morfemas que los generaron. Sin embargo, para analizar los dominios de variación, se clasificó cada uno de los alomorfos en forma individual y se contabilizó estadísticamente cada uno de ellos en cada dominio, lo que dio un valor superior a 93 *tokens*. Este resultado se debió a que algunos alomorfos encontraron la causa de su

variación en más de un dominio y por lo tanto su doble nivel de condicionamiento fue contabilizada dos veces. El total de *tokens* que se contabilizaron para el estudio del nivel de condicionamiento fue de 121.

4.1 CRITERIO 1: NIVEL DE CONDICIONAMIENTO

El criterio de nivel de condicionamiento, está relacionado con el nivel lingüístico en el que ocurre la variación y que vimos en § 3.1. Los dominios lingüísticos donde ocurre la alomorfia incluyen la variación con motivación *fonológica*, la variación con motivación *morfofonológica*, la variación con motivación *morfosintáctica* y la variación con motivación *léxica*. También encontramos alomorfos que se incluyen en el ítem de *variación inmotivada* debido a que para ellos no se ha podido identificar motivación ninguna.

Por fines prácticos de ocurrencias en el SK¹⁰, subcategorizamos la clasificación de dominios. El nivel fonológico se ha dividido en fonológico-suprasegmental y fonológico-segmental. El nivel morfosintáctico también se divide en dos, distinguiendo el nivel morfológico y el nivel sintáctico. Los alomorfos con condicionamiento morfofonológico, se reinterpretan como variaciones en el dominio fonológico-segmental y también como variaciones en el dominio morfológico. Respecto del dominio léxico, para el caso del SK no tiene relevancia, ya que no encontramos ningún caso de este tipo en nuestro corpus.

¹⁰ Estos fines prácticos tienen que ver con el alto porcentaje de alomorfos encontrados en el SK condicionados por el nivel fonológico suprasegmental (por ejemplo, la cantidad de moras de su raíz) y también por el nivel fonológico-segmental (por ejemplo, condicionadas por sibilante que la precede), en contraste con la baja cantidad de alomorfos en los otros niveles. Es por ello que vimos conveniente abrir los niveles y de esta forma parecerse aún más a los niveles de estudio de la lingüística.

Los 30 morfemas que son la base de este estudio, tienen un total de 93 alomorfos. Sin embargo, los niveles de condicionamiento de estos morfemas llegan a 121 *tokens*. Es decir, que un alomorfo tiene más de un dominio de condicionamiento. Esta relación entre alomorfos y nivel de condicionamiento no es biunívoca, y presenta un “exceso de dominios de condicionamiento”. La explicación es sencilla y es el resultado de la separación de niveles explicada anteriormente. Por un lado, algunos alomorfos condicionados fonológicamente, internamente deben su variación a motivos suprasegmentales y también segmentales, como lo revisaremos en el ejemplo (45). Por otro lado, los alomorfos condicionados en el dominio morfofonológico deben su condicionamiento al dominio fonológico-segmental y morfológico, como lo muestran los ejemplos (46) y (47).

En (45) vemos el enclítico polifuncional *forma -n*, que ocurre en el nivel fonológico-segmental y suprasegmental. El alomorfo *-man* elegirá raíz que tenga dos sílabas y termine en /n/. Para adjuntarse eliminará la /n/ y adjuntará el morfema *-man* (Valenzuela 2003b: 120 ejemplo 38).

(45)	forma ABS	<i>forma -n</i>	glosa
a.	<i>amen</i>	<i>ame-man</i>	‘ronsoco’
b.	<i>manan</i>	<i>mana-man</i>	‘montaña’

Del mismo modo, los alomorfos condicionados morfofonológicamente (§3.1.2) internamente están alternando en dos niveles de condicionamiento, el fonológico-segmental y el morfológico. Revisemos el caso de dos alomorfos del ablativo¹¹ *-nkonია* y *~mea* en los ejemplo (46) y (47) respectivamente. El ablativo será *-nkonია* si el

¹¹ El caso ablativo se forma con los casos locativo-ablativo, lo que resulta en un "apilamiento de casos" (Valenzuela 2003b: 228). Las terminaciones ablativas son siete, *-mea*, *me ~ -(ai)nkonიაx* ~ *-kea* ~ *-ainoa* ~ *-nkonია* ~ *-nkoxon* ~ *-oa(x)*

locativo¹² que lo compone termina en *-nko*. Entonces se le añadirá *-nia*. Pero, si el locativo de la raíz se configura desde la *forma -n*, entonces pierde la última nasal y adjunta *-mea*. (Valenzuela 2003b: 228).

En estos dos ejemplos, el alativo elegirá su alomorfo de acuerdo al locativo que lo compone y a la vez, ese locativo es elegido por las moras o sílabas que tiene la palabra raíz. En (46) vemos que a *wetsa* ‘otro’ se le añade *-nia* después del locativo *-nko*¹³. Contrasta con (47), la palabra *manan* ‘montaña’ que añade *-mea* debido a que la palabra *mana* toma su locativo de la *forma -n*, como se explica en el ejemplo (45 b)

- (46) Jepe Ian ainbo-bo iki wetsa-*nko-nia*-bo keská-ma.
 Jepe Ian mujer-PL:ABS COP otro-LOC-**ABL**-PL SIML-NEG
 ‘Las mujeres del pueblo de San Francisco no son como las de otros pueblos.’
- (47) No-n Inka iki mana-*ma-mea*-ma....
 1P-GEN Inka:ABS COP montaña:LOC-**ABL**-NEG
 ‘Nuestro Inka no es de los Andes....’

Resultados del análisis del Criterio 1: nivel de condicionamiento

Después de analizar cada nivel de condicionamiento a los 93 alomorfo (ver anexo-2), encontramos los siguientes resultados globales presentados en la Tabla 1. En ella se observa una tendencia hacia la variación fonológico-segmental con 50 (41,3 %) *tokens* del total. Es seguida por la variación fonológica suprasegmental con 31 (25,6 %) *tokens* del universo. Juntas, los niveles de condicionamiento suprasegmentales y segmentales (que llamaremos simplemente fonológicos) suman el 66,9% del universo, tal como lo

¹² El locativo en SK tiene dos manifestaciones, primero en su forma locativa-alativa, los alomorfos son *-nko* ~ *-ain* ~ *-ainko*. Sin embargo también toman la *forma -n* cuando el sustantivo al que se adjuntará tiene dos sílabas y termina en sílaba abierta sin acentuar (Valenzuela 2003b: 207)

¹³ La palabra *wetsa* ‘otro’ toma su alomorfo por el número par de moras.

muestra la Tabla 2. El tercer grupo que les sigue en porcentaje estadístico es el de variación inmotivada con 28 *tokens* (23,1%) del total. Por último los nivel de condicionamiento morfológica y sintáctico son los menos frecuente con solo 8 (6,6%) y 4 (3,3%) *tokens* y porcentajes del total, respectivamente. La Tabla 2 representa el resumen de los cuatro niveles agrupados en dos, a saber, el fonológico suprasegmental y fonológicos segmental, llamado fonológico, por una parte y los niveles morfológicos y sintácticos, por otro. Claramente el nivel fonológico es el mayoritario. El detalle de cada alomorfo clasificado según su nivel de condicionamiento se encuentra en el Anexo 2.

Tabla 1 Criterio 1: Nivel de condicionamiento

Nivel de condicionam.	Fonológico Suprasegmental	Fonológico Segmental	Morfológico	Sintáctico	"Inmotivado"	Total
Ocurrencias	31	50	8	4	28	121
Porcentajes	25,6%	41,3%	6,6%	3,3%	23,1%	100%

Tabla 2 Criterio 1 Nivel de condicionamiento reagrupados

Nivel de condicionam.	Fonológico	Gramatical	Inmotivado	Total
Ocurrencias	81	12	28	121
Porcentajes	66,9%	9,9%	23,1%	100%

El Gráfico 1 muestra estas tendencias en la forma de una representación tipo circular. Se observa en color verde el nivel de condicionamiento mayor, el fonológico segmental (41,3%). Luego, en azul, encontramos la el nivel de condicionamiento suprasegmental, representa un poco más de la cuarta parte del total. Se destaca en color anaranjado “la inmotivada”, que representa casi la cuarta parte del total (23,1%). Por último, los dominios con las más bajas ocurrencias son el morfológico (en rojo) y el sintáctico (con solo una pequeña fracción del gráfico circular y en amarillo). El Gráfico 2 combinamos los dominios tal como se hizo en la Tabla 2.

Gráfico 1 Criterio 1: Niveles de condicionamiento

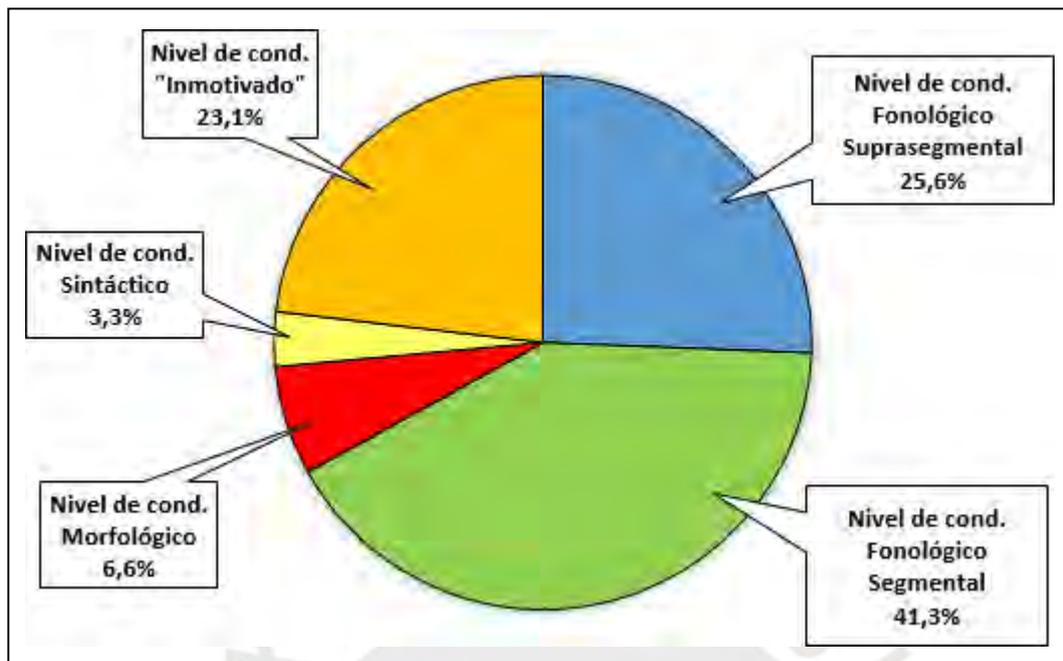
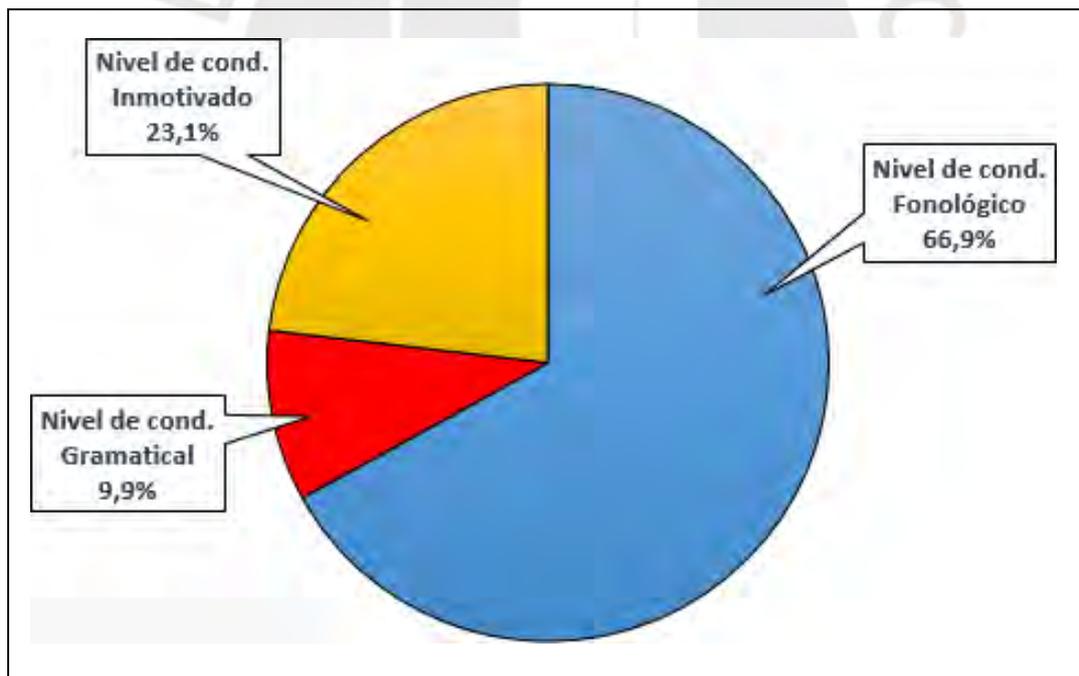


Gráfico 2 Criterio 1: Niveles de condicionamiento agrupados



En resumen, el nivel de condicionamiento de la alomorfia está motivado principalmente por el dominio fonológico, a saber, a nivel suprasegmental y segmental.

4.2 CRITERIO 2: VARIABILIDAD FORMAL

Como lo vimos en la § 3.2, la variabilidad formal es entendida en esta tesis como el número de alomorfos que manifiesta cada morfema. El resultado del análisis de los 30 morfemas (y sus 93 alomorfos) lo podemos ver en la Tabla 3.

Tabla-3 Criterio 2: Variabilidad formal

Número de alomorfos por morfemas	#2	#3	#7	#9	#11	Total
Número de morfemas	19	7	2	1	1	30
Porcentajes	63,3%	23,3%	6,7%	3,3%	3,3%	100,0%

En ella, observamos que el SK tiene cinco grupos de variabilidad formal: tenemos 19 casos en que el morfema tiene **dos alomorfos**. Ellos representan la mayor frecuencia con un 63,3% de la muestras. Dentro de ese grupo de morfemas tenemos el caso de enclítico despreciativo *-isi ~ -osi*; el sufijo verbal *-yat ~ -ya*; o el enclítico especulativo *-mein ~ -main*, etcétera.

Luego le siguen los casos con **tres alomorfos** por morfema, siete casos que representan con el 23,3% del universo. Ejemplo de este grupo es el enclítico propietivo con los alomorfos *-ya ~ -sha ~ -a*,

El siguiente grupo es el de los morfemas que tienen **siete** alomorfos por morfema, que representa el 6,7% del total y son solo dos, el sufijo verbal intransivizador reflexivo *-(t) ~ -kaa(t) ~ -φ ~ -mee(t) ~ -(k)ee(t) ~ -(k)ii(t) ~ -(k)oo(t)*.

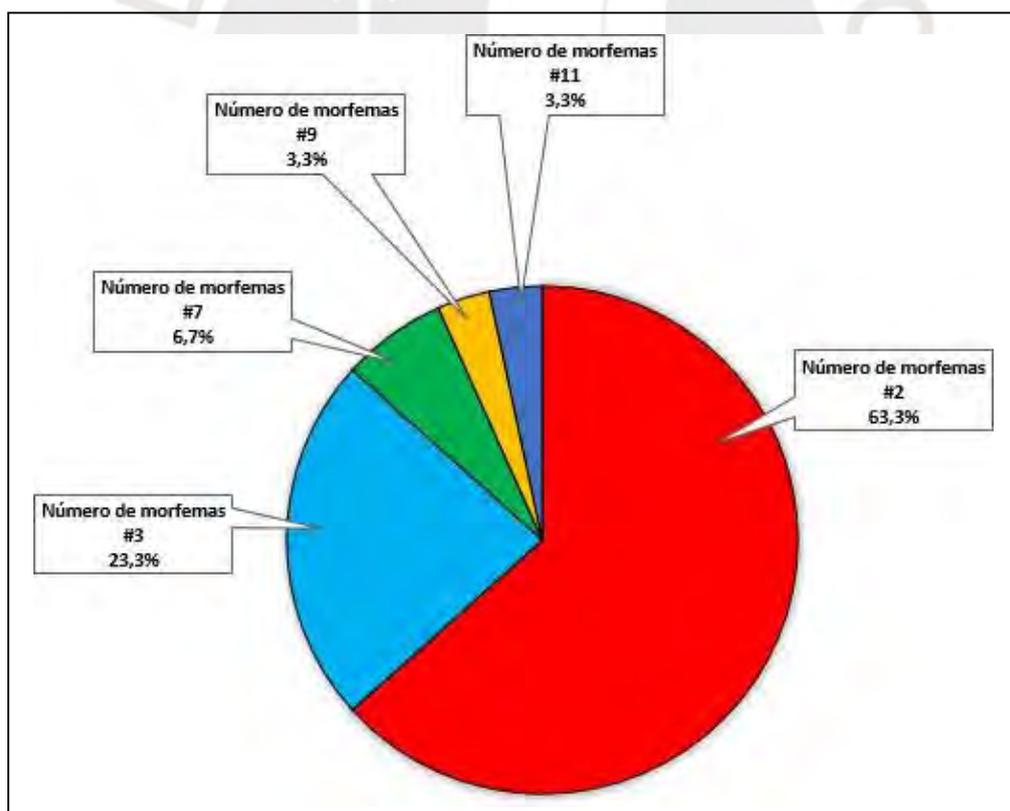
Los últimos son los morfemas con variabilidad formal de **nueve** alomorfos por morfema y **once** alomorfos por morfema. De ambos solo hay un morfema por grupo, y cada uno representa el 3% del total. El paradigma con nueve alomorfos es sufijo verbal malefactivo *-anan ~ -ananan ~ -(a)naan ~ -(e)naan ~ -(i)naan ~ -(o)naan ~ -naan ~ -nan*

~ -an y con once alternancias el enclítico polifuncional *forma -n* con -n ~ -en ~ -an ~ -in ~ -man ~ -kan ~ -tan ~ -ten ~ -nin ~ -ton ~ -tonin.

El Gráfico 3 ilustra la distribución de la Tabla 3 en una representación de gráfico circular. Apreciamos la superioridad de la variación binaria (en color rojo) con respecto a los casos de mayor variabilidad formal: tres alomorfos (azul) que representa casi la cuarta parte del universo y los casos de siete, nueve y once alomorfos (colores verde, amarillo y morado, respectivamente) con un porcentaje muy menor respecto al de los anteriores.

El detalle de la lista de alomorfos clasificados según su variabilidad formal se encuentra en el Anexo 3.

Gráfico 3 Criterio 2: Variabilidad formal



En resumen, el SK es una legua con una variabilidad formal binaria y con escasos alomorfos de gran variabilidad.

4.3 CRITERIO 3: PREDICTIBILIDAD

El tercer criterio de organización tipológica es la predictibilidad, tal como se vio en §3.3. Cuando un morfema presenta dos o más alomorfos, estos pueden ser **totalmente predecibles** si sus niveles de condicionamientos son precisos y se cumplen sin excepción. Por otro lado, decimos que es **parcialmente predecible** cuando un alomorfo, en un determinado nivel de condicionamiento, puede tomar distintas alternancias y ninguna de ellas puede predecirse. Por último, la variación es **impredecible** cuando nunca es posible saber el nivel de condicionamientos donde se produce esta alomorfía.

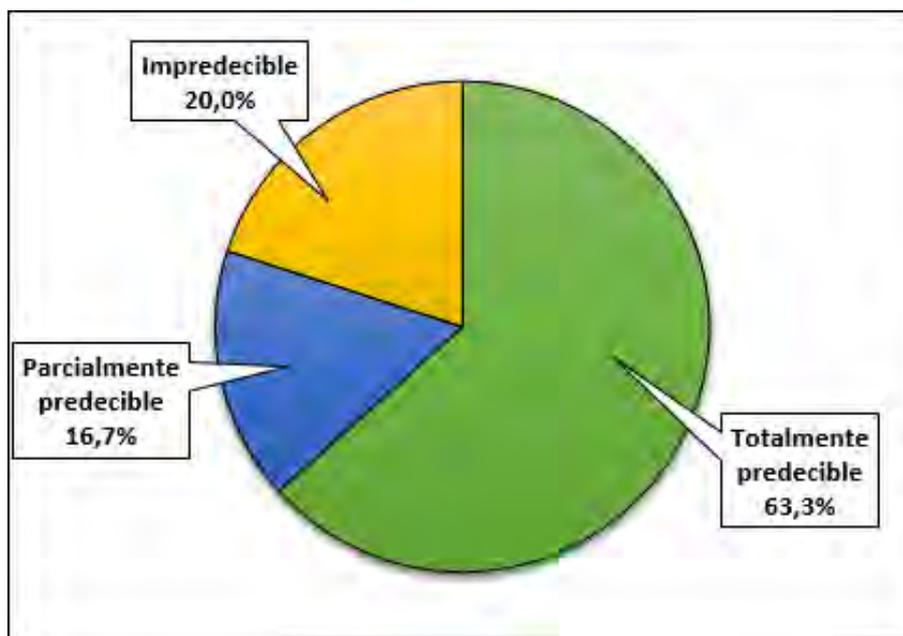
En la Tabla 4 podemos observar los resultados. Un universo de 19 alomorfos (63,3%) presenta una variación formal predecible. Luego, la predictibilidad parcial comparte con la impredecibilidad valores muy cercanos. La primera, los alomorfos parcialmente predecibles cuentan 5 casos (16,7%) y la segunda, los morfemas con alomorfía impredecibles son 6 casos (20%). Los alomorfos clasificados según su predictibilidad se encuentran en el Anexo 4.

Tabla-4 Criterio 3: Predictibilidad

Predictibilidad	Totalmente predecible	Parcialmente predecible	Impredecible	Total
Morfemas	19	5	6	30
Porcentajes	63,3%	16,7%	20,0%	100%

El Gráfico 4 muestra estas proporciones. En verde se observa las casi dos terceras partes de los alomorfos totalmente predecibles. Luego en amarillo y verde, encontramos las proporciones restantes que, como ya se dijo, son muy semejantes (impredecibles y parcialmente predecibles, respectivamente).

Gráfico 4 Criterio 3: Predictibilidad



En resumen, el SK muestra una tendencia a que sus alomorfos sean totalmente predecibles. Sin embargo, el 36% de los casos son parcialmente predecibles o impredecibles.

4.4 CRITERIO 4: SEMEJANZA FORMAL

Presentamos la semejanza formal en § 3.4, donde afirmamos que los alomorfos de un morfema son **todos parecidos entre sí** cuando el paradigma solo varía en una vocal o consonante. Luego decimos que son **parcialmente parecidos entre sí**: cuando el paradigma varía en más de una vocal o consonante. Finalmente no tienen **ningún parecido entre sí** cuando los alomorfos del paradigma no tiene nada en común entre ellos, como en el caso de la suplección.

En el SK la semejanza formal presenta una clara tendencia hacia una similitud formal entre los alomorfos. En la Tabla 5 observamos que el 73,3% de los alomorfos son parecidos entre sí, es decir, solo varía una consonante o vocal entre ellos. Luego, están los casos de alomorfos parcialmente parecidos, es decir que tiene dos o más consonantes o vocales diferentes, que representan el 23,3 %. Finalmente los casos en los que los alomorfos no presentan ningún parecido entre sí alcanzan solo al 3,3 % del universo de nuestro corpus. Este único caso es una suplección total. Los alomorfos clasificados según su semejanza formal se encuentran en el Anexo 5.

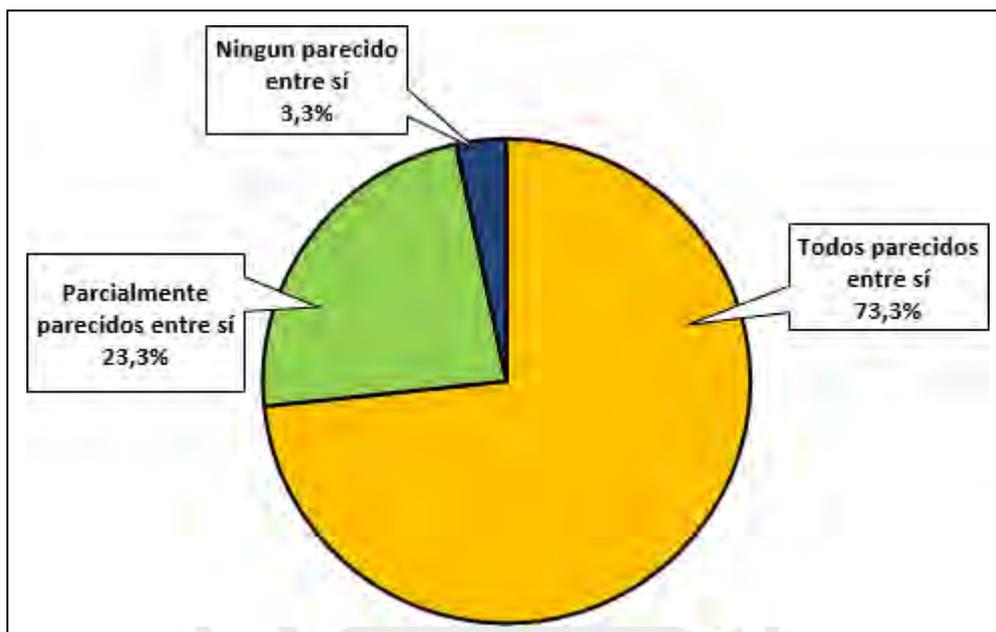
Tabla 5 Criterio 4: Semejanza formal

Semejanza formal	Todos parecidos entre sí	Parcialmente parecidos entre sí	Ningún parecido entre sí	Total
Morfemas	22	7	1	30
Porcentajes	73,3%	23,3%	3,3%	100 % ¹⁴

El Gráfico 5 muestra esta distribución, con un alto porcentaje de alomorfos parecidos entre sí (en amarillo) seguido de un porcentaje verde para los que manifiestan solo una similitud parcial. Finalmente, el único caso de ningún parecido entre las formas aparece en azul.

¹⁴ La suma de porcentajes resulta 99.9% y se aproxima automáticamente a 100%

Gráfico 5 Criterio 4: Semejanza formal



En resumen, con respecto al criterio de semejanza formal, el SK muestra una tendencia a que sus alomorfos sean parecidos entre sí, variando solo una vocal o consonante en cada paradigma en esta categoría.

5. CONTRIBUCIONES AL DESARROLLO DE UN TRADUCTOR AUTOMÁTICO CASTELLANO SK

En esta tesis hemos presentado una tipología de la variación formal y que luego aplicamos a la alomorfía del SK. En este capítulo exploraremos las consecuencias prácticas de nuestro estudio para la tarea de diseñar un traductor automático castellano-SK.

5.1 APLICACIONES Y SUGERENCIAS RELACIONADAS CON EL NIVEL DE CONDICIONAMIENTO

En la aplicación de nuestra tipología en § 4.1 encontramos, bajo el primer criterio¹⁵, que el dominio fonológico (66,9 %) es el nivel de condicionamiento que genera la mayor alomorfía del SK. El nivel de condicionamiento fonológico-segmental (41,3 %) no presenta problemas, ya que la alternancia depende de la identificación de sílabas o consonantes previas a las que se adjuntará el alomorfo. Sin embargo, el desafío del TA, en su tarea de traducir, estará en el otro 25,6% de los alomorfos que se presentan en el nivel fonológico-suprasegmental. Decimos que es un desafío porque el TA deberá, por ejemplo, silabificar y contar moras de la palabra raíz a la que adjunte el alomorfo correspondiente.

Con el fin de poder predecir qué alomorfo debe asignar el TA, cuando el trabajo de traducir sea desde el sentido Castellano al SK, será necesario implementar algunos procedimientos automáticos previos.

¹⁵ El resultado de la aplicación de estos criterios se encuentra en el Anexo 2, específicamente en la “explicación del nivel de condicionamiento” donde se detalla los criterios para la selección de cada alomorfo.

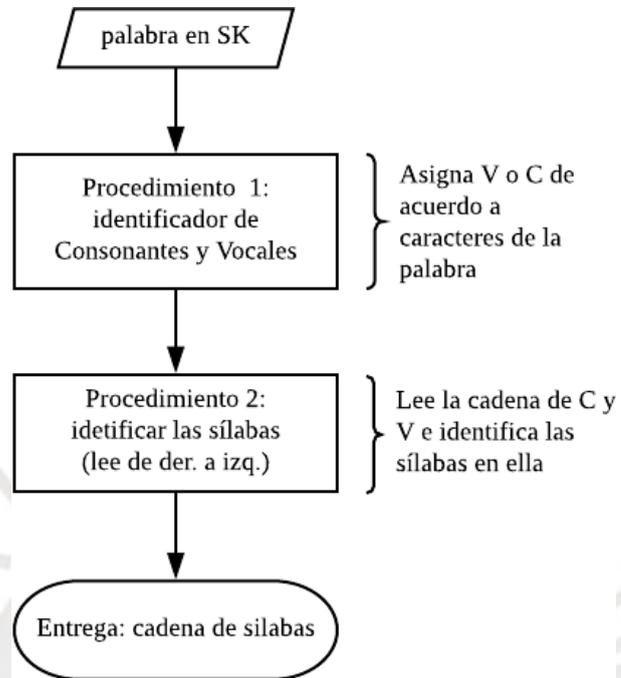
Para encontrar estos procedimientos, busqué en el análisis de los 93 alomorfos enlistados en el Anexo 2 y encontré la repetición de diez procesos que el TA debe implementar para seleccionar la alomorfía que corresponda. Estos procedimientos son los siguientes: contar moras, contar sílabas, identificar el último carácter, identificar la paridad de sílabas o moras, identificar sustantivos derivados, identificar el tipo de sílabas (trabada o abierta), identificar sufijos específicos, identificar sustantivos derivados, leer el último carácter.

Tomando los procedimientos más recurrentes, a saber, el contador de sílabas y moras, detallaré a continuación las sentencias que deben realizar, con su respectivo diagrama de flujos y aplicando cada uno de estos procedimientos a un morfema a modo de ejemplo.

5.1.1 El contador de sílabas

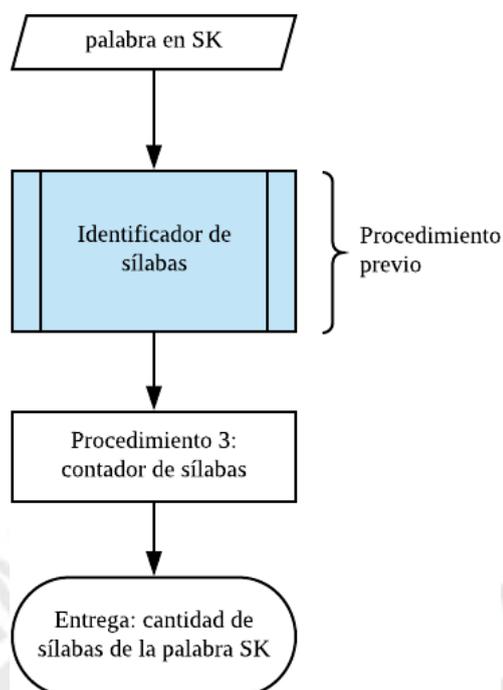
Para seleccionar el alomorfo correspondiente, el TA deberá identificar cada sílaba que contenga la raíz en SK. El trabajo consta de dos procedimientos, el primero consiste en tomar la raíz de la palabra en SK e identificar las vocales y consonantes dejando una cadena de caracteres con valores C y V. El segundo procedimiento es la silabificación de la cadena de caracteres, basado en el modelo de sílabas posibles descrita en §2.1.1. Dicho modelo identifica solo cuatro formas silábicas V, CV, VC y CVC. El TA deberá leer de derecha a izquierda la cadena de C y V, y luego agrupar las sílabas. Todo este procedimiento está ilustrado en la Figura 4.

Figura 4 Identificador de sílabas



Una vez terminada la silabificación, el TA deberá contar las sílabas encontradas entregando como resultado un valor numérico. El procedimiento está diagramado en flujos en la Figura 5.

Figura 5 Contador de sílabas



5.1.2 Ejemplo de procedimiento de contar sílabas y asignar alomorfo

Los morfemas *-anan*, *forma -n*, *-t* y *-taanan*, utilizan el procedimiento de contar sílabas, entre otros, para asignar sus alomorfos. Ilustraremos este procedimiento en los verbos *reteti* ‘pelear’ y *xoati* ‘rascar’ para la asignación del alomorfo del sufijo verbal recíproco *-anan* en cada uno de ellos.

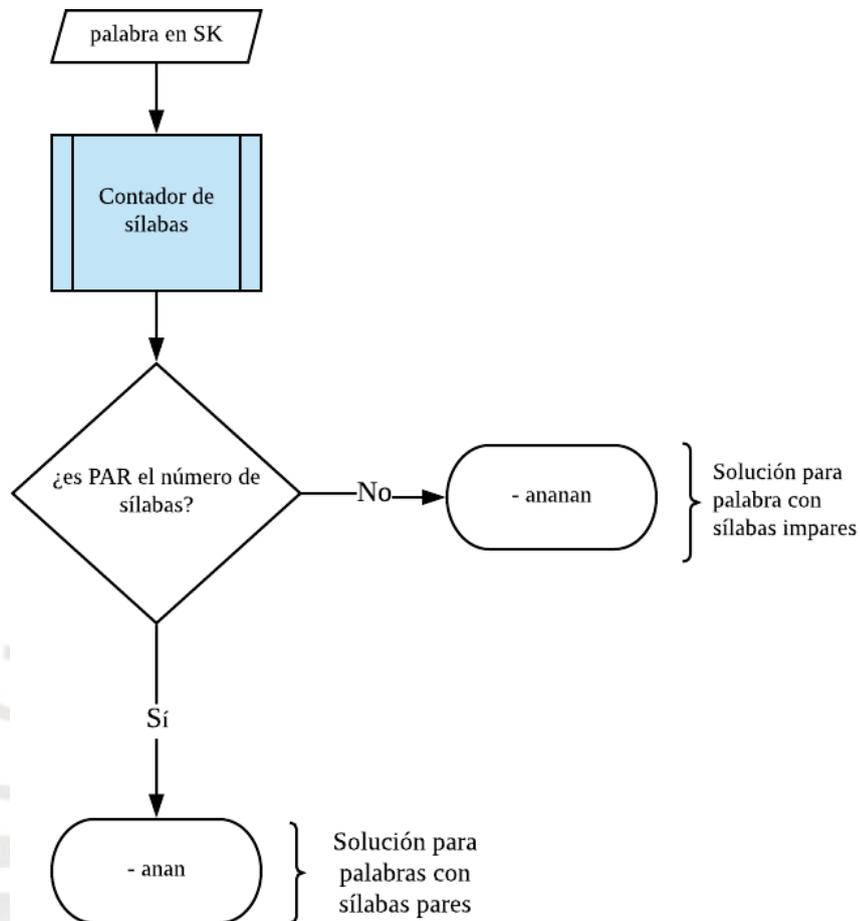
El primer procedimiento que realiza el contador de sílabas es identificar en la palabra SK las vocales y consonantes, en estos casos de los verbos *reteti* y *xoati* sin su forma infinitiva, *rete-* y *xo-* respectivamente. El verbo *rete-* lo transforma en un equivalente a CVCV y el verbo *xo-* en CV. El segundo procedimiento realiza la silabificación, es decir, identificar las sílabas de cada palabra leyendo de derecha a izquierda. Para el verbo *rete-* identifica primero la sílaba CV y luego CV, ambas abiertas. Para el verbo *xo-* identifica CV, también sílaba abierta. Luego, como tercer procedimiento

cuenta las sílabas encontradas en cada caso. En el verbo *rete-*, las sílabas CV.CV suman 2 sílabas y en el verbo *xo-*, la sílaba CV cuenta solo 1. Finalmente debe responder la pregunta “¿es par el número de sílabas?”. La respuesta para el primero, *rete-* es “sí” (2 es par), y se le asigna el alomorfo *-anan*. En el segundo caso la respuesta es “no”, pues 1 es impar y se le asigna el alomorfo *-ananan*. De este modo las formas recíprocas de los verbos quedarían como *reteanan* ‘pelearse el uno al otro’ y *xoananan* ‘rascarse el uno al otro’. Podemos comprobar estas asignaciones alomórficas con los ejemplo (48) y (49) del diccionario Castellano SK (Loriot 1993: 90) y el ejemplo 125 de la gramática SK (Valenzuela 2003: 814) respectivamente.

- (48) Ja atapa ramera rete-*anan*-ai
 Esas gallinas dos pelear-REC-INC
 ‘Esas dos gallinas están peleándose’
- (49) Oa ramera maponko xo-*ananan*-ai.
 Estos dos:ABS-EV cabeza-LOC rascar-REC-INC
 ‘Esos dos rascaron uno al otro sus cabezas’

La figura 6 muestra este mismo procedimiento en diagramas de flujo. Se observa que utiliza el contador de sílabas que además utiliza el identificador de sílabas.

Figura 6 Procedimiento de asignación del alomorfo
sufijo verbal recíproco *-anan* ~ *-ananan*

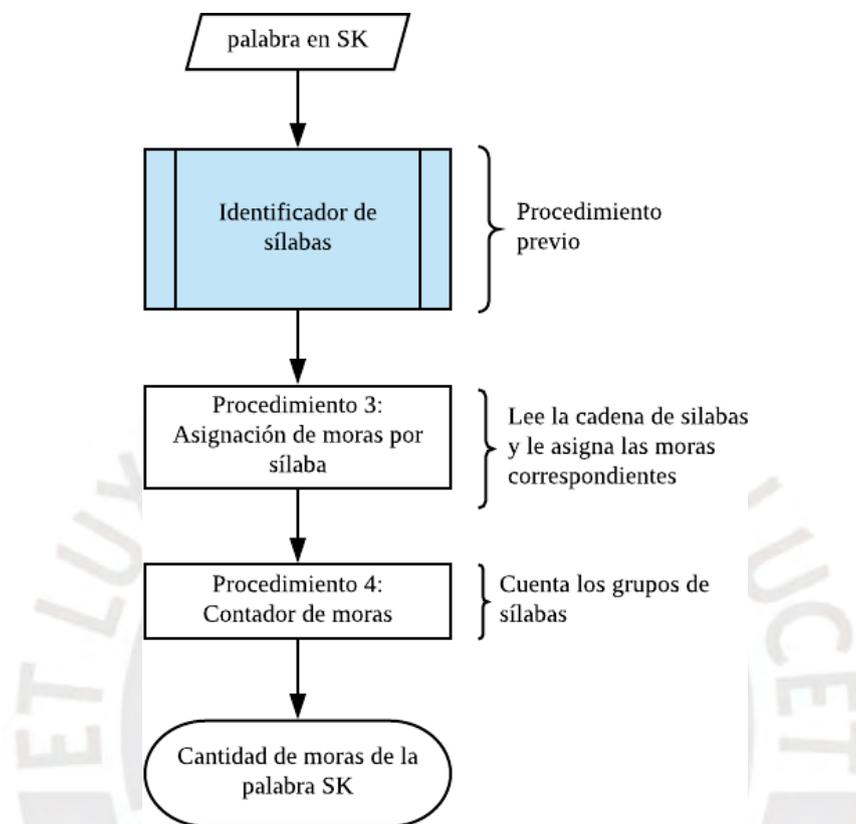


5.1.3 Contador de moras

También será necesario que por cada palabra raíz traducida desde el español al SK se le cuente el número de moras, para el uso de los alomorfos *-ain*, *-anan*, *-katit*, *-kin* y *-riba*. Basado en el identificador de sílabas, el programa debe revisar la estructura de cada sílaba y asignarle un valor moraico de acuerdo al siguiente criterio: CV y V se les considera sílabas abiertas y ligeras, su valor moraico es uno; las sílabas trabadas VC y CVC, como las sílabas ligeras acentuadas son consideradas pesadas con valor moraico de dos. Una vez que tenga identificadas las sílabas con sus respectivos pesos moraicos, sumará los

valores de cada sílaba y entregará la cuenta de ellos. Cada procedimiento es ilustrado en la Figura-7

Figura 7 Contador de moras



5.1.4 Ejemplo de procedimiento de contar moras y asignar alomorfo

Utilizaremos dos verbos *biti* ‘traer’ y *oroti* ‘cultivar’ para asignarles el sufijo verbal asociativo *-kin*, sometiéndolos a los procedimientos de los diagramas de flujo de las Figuras 5 y Figura 7. Este sufijo verbal asociativo tiene el nivel de condicionamiento en el nivel suprasegmental de acuerdo al número de moras que presente la raíz al que se sufijs. Si dicha raíz tiene un número par de moras se le adjunta el alomorfo *-kin*. En cambio, si la raíz tiene un número impar de moras, se le adjunta el alomorfo *-kiin* (Valenzuela 2003b: 105). Una vez que probemos el contador de moras, comprobaremos

el resultado con las oraciones aplicativas del Diccionario Castellano SK del ILV (Loriot et al. 1993: 349).

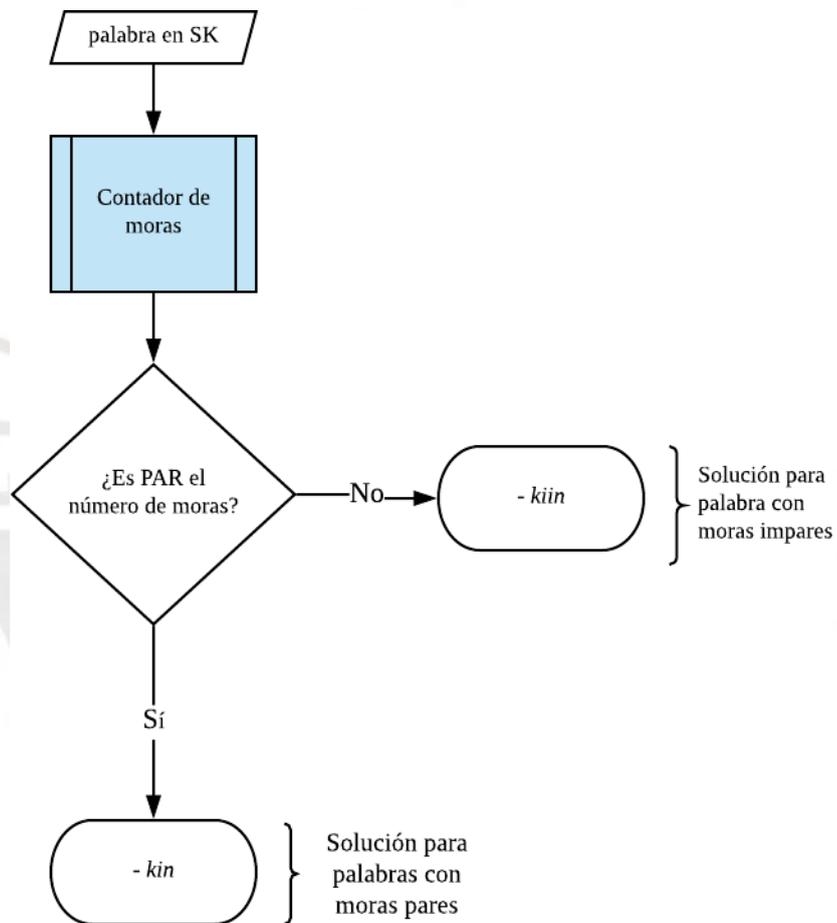
Primeramente el TA debe realizar el procedimiento de identificar las consonantes y vocales de la palabra, en estos casos de los verbos *biti* y *oroti* sin su forma infinitiva, *bi-* y *oro-* respectivamente. El verbo *bi-* lo transforma en un equivalente a CV y el verbo *oro-* en VCV. El segundo procedimiento realiza la silabificación, es decir, identifica las sílabas de cada palabra leyendo de derecha a izquierda. Para el verbo *bi-* identifica la sílaba CV (abierta) y para *oro-* identifica, primero CV y luego V, ambas sílabas abiertas. Luego, como tercer procedimiento asigna las moras a cada sílaba identificada. La palabra *bi-* es una sílaba CV ligera pero acentuada, y por lo tanto le asigna 2 moras convirtiéndose en una sílaba acentuada y pesada. Para la palabra *oro-* con las sílabas identificadas V y CV, le asigna a V que es una sílaba abierta y ligera pero acentuada, 2 moras y a CV le asigna 1 mora, pues es sílabas abierta y ligera. Finalmente el cuarto procedimiento es contar las moras de cada palabra. En el caso de *bi-* las moras contadas son 2 y para el otro verbo, *oro-*, la cuenta de moras es 3. Este resultado es sometido al identificador de paridad y debe responder la pregunta “¿es PAR el número de moras?”. El verbo *bi-* responderá “sí” a dicha pregunta, porque tiene 2 moras y le será asignado el alomorfo *-kin*. Por otro lado, para el verbo *oro-* que contiene 3 moras, la respuesta a la paridad será “no” y se le asignará el alomorfo *-kiin*. Así las formas asociativas de los verbos *biti* será *bikin* y la de *oroti* será *orokiin*, como lo muestra el ejemplo del diccionario (Loriot 1993:349) en (50) y (51) respectivamente.

- (50) Maistoro-ra en pei *bi-kin*-ai xobo-a-ti
Maestro-EV yo-ERG hoja traer-ASSOC-INC casa-CAUS-INF
‘Estoy ayudando al maestro a *traer* hojas para hacer su casa’

(51) Eanra nokon bake oro-**kiin**-ke jawen wai
 Yo-ABS-EV mi hijo cultivar-ASSOC-CMPL su chacra
 ‘He ayudado a mi hijo a cultivar su chacra’

Se puede observar el procedimiento de asignación alomorfa del morfema *-kin* ya descrito, en la Figura 8.

Figura 8 Procedimiento de asignación del alomorfo sufijo verbal del asociativo *-kin ~kiin*



5.2 SUGERENCIAS EN LA VARIABILIDAD FORMAL Y PREDICTIBILIDAD

Respecto al segundo criterio, el de **variabilidad formal**, el SK es principalmente y mayoritariamente una lengua en la que los paradigmas de variación morfológica constan de dos alomorfos (63,3%). Por ello, el traductor automático debería estar en condiciones de procesar paradigmas binarios. La baja cantidad de paradigmas de variabilidad formal mayores a tres, muestra que los casos de alta variabilidad (y por lo tanto más complejos y difíciles de programar) son escasos, aunque no por ellos poco importantes. De igual forma su **predictibilidad** (el tercer criterio) alcanza un 63,3%, lo que indica que los casos especiales de predictibilidad parcial o impredecibles son de baja frecuencia (36,7%). Pero estas bajas estadísticas no quieren decir que los alomorfos con una alta variabilidad formal y niveles de condicionamiento parcialmente predecibles sean alomorfos de poco uso. Al contrario, muchos de ellos son los más funcionales¹⁶, como el caso de enclítico polifuncional *forma -n* que revisaremos a continuación.

El caso del enclítico polifuncional *forma-n* será un proceso difícil de tratar debido a que gran parte de sus once alomorfos tienen un nivel de condicionamiento acotado pero a la vez alternan entre sí sin motivación aparente.

El enclítico polifuncional *forma -n* es uno de los más utilizados con funciones de caso ergativo, genitivo, instrumental, locativo y otras funciones oblicuas (Valenzuela 2003b: 118). Valenzuela clasifica sus niveles de condicionamientos en ocho grupos. A continuación enlistaré estos grupos, con su respectivo ejemplo (con la primera palabra en

¹⁶ El caso puntual de este enclítico es tema de otra tesis del proyecto Chana: *Desambiguación de morfemas polifuncionales en la traducción automática de lenguas minoritarias: el caso del enclítico =n en el shipibo-konibo*, a cargo de Renzo Ego-Aguirre.

su forma absoluta) y propondré al final de cada ejemplo una solución a la TA como una sugerencia.

Grupo 1: Para palabras de dos sílabas livianas (abiertas). Se le epentiza *-n*. En (52) vemos la forma absoluta del sustantivo *bake* al que se le aplica la *forma -n*. Este alomorfo es el totalmente predecible.

(52) *bake* *bake-n* ‘niño’

Grupo 2: Palabras con dos sílabas, la segunda pesada y trabada con una consonante sibilante. A esta forma se le epentiza una vocal /-e/ o /-a/. La /i/ es admitida solo si la palabra termina en /iS/. Es una forma de armonía vocálica. En la palabra *mawis* ‘hormiga’ (53) en su forma absoluta, cuenta con la vocal /i/ en su última sílaba, por lo que se le puede adherir los tres enclíticos de la forma *-n*, estos son *-en*, *-an* e *-in*.

(53) *mawis* *mawis-en* ~ *mawis-an* ~ *mawis-in* ‘clase de hormiga’

Pese a estar en un nivel de condicionamiento predecible, esto es, palabras de dos sílabas con la segunda sílaba trabada terminada en sibilante, estas tres alomorfos alternan en tres formas: *-en*, *-an* e *-in* (si la última sílaba es /i/) lo que hace que este subconjunto sea parcialmente predecible. Ante casos como el descrito, el TA deberá ser programado en forma mixta. Por un lado, debe predecir el contexto y luego basarse en estadísticas de su corpus. De no tenerlos, entonces deberá elegir una forma como la usual. Sugiero una aproximación de armonía vocálica, es decir, aproximarse fonológicamente a las vocales previas de la raíz.

Grupo 3: agrupa los sustantivos con dos sílabas terminados en /n/. La nasal final /n/ cambia a /m/, se le epentiza una /a/ a la *-n*. En el ejemplo (54) vemos lo predecible que es este alomorfo.

(54) *nonon* *nono-man* ‘pato’

Grupo 7: Las palabras nominales que llevan el plural *-bo* (ABS). En este caso, la *forma -n* ocurre como *-baon*, *-boan*, e incluso *-boon*, como se observa en (59)

(59) ochiti-*bo* ochiti-*baon* ~ ochiti-*boan* ~ ochiti-*boon* ‘perros’

Esta alomorfía es impredecible, y la solución puede ser, por un lado, las construcciones estadísticas más comunes o por otro lado, optar por uno de los tres alomorfos como el canónico.

Grupo 8: Los derivados nominales exhiben la *forma-n* como: *-ton* ~ *-tonin*. Se necesita más investigación para tener en cuenta la distribución de estas formas alternativas.

Podemos ver ambos alomorfos, *-ton* en (60) y *-tonin* en (61).

(60) westiora ‘uno’ westiora-*ton*

(61) wisoshaman ‘muy negro’ wisoshaman-*tonin*

En este grupo la alomorfía es parcialmente predecible debido a que tiene un dominio de variación acotado, los derivados nominales. La solución puede ser, por un lado, las construcciones estadísticas más comunes o por otro lado, optar por uno de los dos alomorfos como el canónico.

Nuestra propuesta es que el traductor se enfoque en un primer momento en aquellos casos de formas totalmente predecibles (63,3%) que deberían ser fácilmente reducibles a reglas susceptibles de representarse en códigos de computadora. Los casos de formas no predecibles y parcialmente predecibles deberán abordarse de forma cuidadosa. Cuando es parcialmente predecible deberá buscar el alomorfo que más se ajuste a las reglas asociadas a término, pero siempre habrá imprecisión en ello, como lo vimos en los grupos parcialmente predecibles, a saber, los grupos 2, grupo 4 y grupo 8 de la *forma -n*.

Así mismo, si un grupo de alomorfos no son predecibles, el traductor deberá buscar ejemplos previos para elegir una variedad o sencillamente tener un alomorfo como predeterminado. Es probable que buena parte de las formas parcialmente predecibles o impredecibles requiera un tratamiento léxico. En otras palabras, es probable que sea más económico tratar alomorfos como sinónimos si su distribución es poco o nada predecible.

5.3 SUGERENCIAS EN LA SEMEJANZA FORMAL

Con relación a la semejanza formal, esta podría tener menor relevancia que los otros fenómenos presentados aquí cuando pensamos en un trabajo de traducción en el sentido Castellano al SK. Pero su aplicación puede aportar en la corrección ortográfica en el sentido de traducción SK al castellano. Cuando un individuo escriba el morfema puede cometer el error de elegir la forma más simple, y el corrector podrá aproximarse con facilidad a la forma correcta con una alta probabilidad de acertar, debido a que el 73,3% de los alomorfos son totalmente parecidos entre sí, esto quiere decir que varían entre sí en solo una letra o vocal.

6. CONCLUSIONES

Después del presente estudio, es posible ofrecer una tipología de los fenómenos de variación morfológica en el SK y gracias a esta clasificación, vislumbrar las facilidades y problemas que presentará la tarea de traducción.

De acuerdo a los cuatro criterios utilizados en esta tesis, podemos decir, en primer lugar, que el **nivel de condicionamiento** de la alomorfía está motivado principalmente por el dominio fonológico, a saber, a nivel suprasegmental y segmental.

Luego, de acuerdo al criterio de **variabilidad formal**, el SK muestra patrones de baja variabilidad que favorece los casos de paradigmas binarios. Esto implica que los morfemas con variación en el SK tienen en su gran mayoría dos representaciones alomórficas. Solo en escasas veces encontramos paradigmas de alta variabilidad, como en el caso de las once representaciones alomórficas del enclítico polifuncional *forma-n*.

Por su parte, con relación al criterio de **predictibilidad** concluimos que el SK es una lengua de alomorfía mayormente predecible, debido a que casi las dos terceras partes de su alomorfía tienen niveles de condicionamientos precisos y que se cumplen sin excepción.

Finalmente, del análisis de la **semejanza formal** podemos afirmar que el SK tiende a presentar paradigmas alomórficos en que los alomorfos presentan semejanza entre sí, lo que facilita su reconocimiento.

En resumen podemos concluir que **el SK es una lengua con variación formal principalmente binaria, generalmente predecible, mayormente en lo que respecta al condicionamiento fonológico de tipo prosódico y cuyos alomorfos presentan una alta semejanza entre sí.**

Por otro lado, el proceso de traducción automática tiene dos sentidos, por un lado un sujeto puede ingresar construcciones en SK que el traductor debe convertir a formas gramaticalmente construidas del castellano. Por otro lado, puede darse una situación en la que los sujetos ingresen construcciones en castellano y que el traductor deberá convertir a construcciones gramaticales en SK. Creemos que nuestra tesis ayudará a comprender los desafíos que plantea la automatización del proceso de traducción del castellano al SK.

Los casos de variación morfológica plantean desafíos de distintos grados de complejidad en este proceso (traducción castellano > SK). La idea es que nuestro traductor debería poder elegir el alomorfo correspondiente para cada caso de forma exitosa, a fin de producir estructuras morfológicas felices en la lengua meta.

En ese sentido, al abordar el problema de la variación, un traductor simultáneo SK-castellano como el que viene preparando el proyecto *Chana*, se enfrentará principalmente a formas con todas estas características: el alomorfo prototípico del SK tiene parecido entre sí, es predecible, forma parte de un paradigma binario y tiene un nivel de condicionamiento fonológico. Teóricamente, la computarización de la variación morfológica en el SK no debería ser tan compleja en el contexto de un traductor basado en reglas. El traductor no tendrá grandes problemas para generar formas morfológicamente felices, ya que buena parte de la variación morfológica en el SK son causadas por el **nivel de condicionamiento fonológico-segmental** (41,3%) y que encuentra un correlato directo en la forma escrita de las palabras. Una de las consecuencias de que el SK sea principalmente y mayoritariamente una lengua en la que los paradigmas de variación morfológica conste de dos alomorfos, es que el traductor automático debería estar en condiciones de procesar paradigmas binarios. Muchos de los paradigmas binarios del SK son además predecibles y ello facilitaría aún más el proceso.

La predictibilidad es un criterio central para un modelo de traductor automático. Como lo vimos, las lenguas de pocos recursos computacionales como el SK, necesitan incorporar el uso de reglas y combinarlo con los métodos estadísticos. Un traductor de una lengua minoritaria debe ser necesariamente un traductor (parcialmente) basado en reglas. Por ello, la existencia de patrones de variación no predecibles constituye un desafío enorme para un traductor de este tipo. En ese sentido, nuestro estudio ha mostrado que solo un 36,7 % de los casos de variación morfológica en el SK no son totalmente predecibles (y muchos de estos son en realidad parcialmente predecibles). Nuestra propuesta es que el traductor no tendrá problemas en enfocarse en aquellos casos de formas totalmente predecibles (63,3 %) que deberían ser fácilmente reducibles a reglas susceptibles de representarse en códigos de computadora.

Sin embargo, la mayor utilidad de esta tipología radica en nos permitirnos examinar con detención **los problemas con los que se enfrentará el TA** y en base a ellos presentar una **contribución** en el marco del proyecto Chana.

Los aportes a la tarea del traductor automático son el resultado del análisis de los porcentajes donde **los criterios tipológicos presentan baja frecuencia** pero potenciales problemas. Del criterio de nivel de condicionamiento, aquellas situaciones donde el nivel de condicionamiento es el fonológico-suprasegmental (25,6%); del criterio de variación formal, los alomorfos que presentan una alta variabilidad (3%); y del criterio de predictibilidad los que son , parcialmente predecibles (16,7%) o impredecibles (20%).

Bajo el criterio de **nivel de condicionamiento** pudimos identificar uno de los dominios lingüístico que causa alomorfía, a saber, el fonológica suprasegmental. La relevancia de este hallazgo radica en que el nivel suprasegmental regula la alternancia alomorfa basada en factores como la cantidad de sílabas o moras que contenga la raíz de la palabra. Es por ello que, para estos casos, se propuso programar un identificador de

sílabas como también, los contadores de sílabas y moras, que ayudarán a seleccionar el alomorfo preciso y construir una traducción correcta para cerca del 26% de los casos.

También pudimos reconocer los casos bajo el criterio de **variabilidad formal** alta, como ocurre con el enclítico polifuncional *forma -n*, que presentará un desafío al trabajo de traducción. Este enclítico exhibe ocho grupos de alomorfos, de los cuales tres son **parcialmente predecibles** (Grupo 2, 4 y 8), en cuyo caso el TA deberá elegir el alomorfo a asignar que más se ajuste a las reglas asociadas al término pero a la vez utilizando su base de datos debido a la alternancia. Las sugerencias presentadas son el ajuste por armonía vocálica, o también puede nombrar un alomorfo como la forma canónica. Siempre habrá imprecisión en ello

Finalmente, en los casos en el grupo de alomorfos sea **impredecible** (20%) alomorfía sea, entonces el TA dependerá de su base de datos que contiene ejemplos previos para elegir una variedad o podrá resultar, como también puede sencillamente tener un alomorfo como predeterminado. Es probable que buena parte de las formas impredecibles requieran un tratamiento léxico. En otras palabras, es probable que sea más económico tratar alomorfos como sinónimos.

ANEXOS

Anexo 1 Alfabeto fonético internacional AFI

THE INTERNATIONAL PHONETIC ALPHABET (revised to 2015)

CONSONANTS (PULMONIC)

© 2015 IPA

	Bilabial	Labiodental	Dental	Alveolar	Postalveolar	Retroflex	Palatal	Velar	Uvular	Pharyngeal	Glottal
Plosive	p b			t d		ʈ ɖ	c ɟ	k ɡ	q ɢ		ʔ
Nasal	m	ɱ		n		ɳ	ɲ	ŋ	ɴ		
Trill	ʙ			r					ʀ		
Tap or Flap		ⱱ		ɾ		ɽ					
Fricative	ɸ β	f v	θ ð	s z	ʃ ʒ	ʂ ʐ	ç ʝ	x ɣ	χ ʁ	ħ ʕ	h ɦ
Lateral fricative				ɬ ɮ							
Approximant		ʋ		ɹ		ɻ	j	ɰ			
Lateral approximant				l		ɭ	ʎ	ʟ			

Symbols to the right in a cell are voiced, to the left are voiceless. Shaded areas denote articulations judged impossible.

CONSONANTS (NON-PULMONIC)

Clicks	Voiced implosives	Ejectives
⦿ Bilabial	ɓ Bilabial	ʼ Examples:
ǀ Dental	ɗ Dental/alveolar	pʼ Bilabial
ǃ (Post)alveolar	ɟ Palatal	tʼ Dental/alveolar
ǂ Palatoalveolar	ɠ Velar	kʼ Velar
ǁ Alveolar lateral	ɣ Uvular	sʼ Alveolar fricative

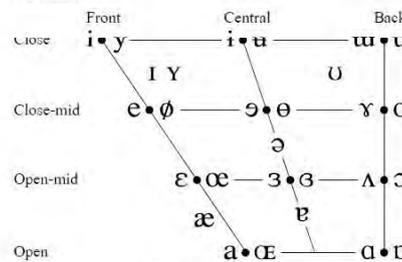
OTHER SYMBOLS

ʌ Voiceless labial-velar fricative	ɕ ʑ Alveolo-palatal fricatives
ʋ Voiced labial-velar approximant	ɺ Voiced alveolar lateral flap
ɥ Voiced labial-palatal approximant	ɥ Simultaneous ʃ and X
ħ Voiceless epiglottal fricative	Affricates and double articulations can be represented by two symbols joined by a tie bar if necessary.
ʕ Voiced epiglottal fricative	ts̺ k̺p̺
ʡ Epiglottal plosive	

DIACRITICS Some diacritics may be placed above a symbol with a descender, e.g. **ŋ̥**

◌̥ Voiceless	ŋ̥ ɖ̥	◌̚ Breathily voiced	ɓ̚ ɗ̚	◌̪ Dental	t̪ d̪
◌̜ Voiced	ʂ̜ ʐ̜	◌̝ Creaky voiced	ɓ̝ ɗ̝	◌̫ Apical	t̫ d̫
◌̙ More rounded	ɔ̙	◌̜ Labialized	t̜w d̜w	◌̜ Nasalized	ẽ
◌̘ Less rounded	ɔ̘	◌̞ Palatalized	t̞j d̞j	◌̚ Nasal release	d̚n
◌̟ Advanced	ɥ̟	◌̜ Velarized	t̜ʷ d̜ʷ	◌̚ Lateral release	d̚l
◌̠ Retracted	ɛ̠	◌̜ Pharyngealized	t̜ʕ d̜ʕ	◌̚ No audible release	d̚ʰ
◌̠ Centralized	ẽ̠	◌̜ Velarized or pharyngealized	ɬ̜		
◌̘ Mid-centralized	ẽ̘	◌̜ Raised	ɛ̜ (ɹ̜ = voiced alveolar fricative)		
◌̙ Syllabic	ŋ̙	◌̜ Lowered	ɛ̜ (β̜ = voiced bilabial approximant)		
◌̚ Non-syllabic	ɸ̚	◌̜ Advanced Tongue Root	ɛ̜		
◌̚ Rhoticity	ɻ̚ ʟ̚	◌̜ Retracted Tongue Root	ɛ̜		

VOWELS



Where symbols appear in pairs, the one to the right represents a rounded vowel.

SUPRASEGMENTALS

ˈ Primary stress	ˈfounəˈtʃən
ˌ Secondary stress	
ː Long	eː
ˑ Half-long	eˑ
◌̥ Extra-short	ẽ̥
◌̚ Minor (foot) group	
◌̚ Major (intonation) group	
◌̚ Linking (absence of a break)	

TONES AND WORD ACCENTS

LEVEL	CONTOUR
ẽ̥ or ʔ Extra high	ẽ̥ or ʔ Rising
é High	é Falling
ē Mid	ē High rising
è Low	è Low rising
ẽ̥ Extra low	ẽ̥ Rising-falling
↓ Downstep	↗ Global rise
↑ Upstep	↘ Global fall

Typeset using Times 10 (L) and Helvetica (A) fonts. IPA 4.0 (2015)

Anexo 2: Criterio 1: Nivel de condicionamiento

N°	Paradigma	Alomorfos	Fonológico Suprasegmental	Fonológico Segmental	Morfológico	Sintáctico	Sin motivación aparente	Explicación dominio de condicionamiento	
1	ain	-ain	Si					En número impar de moras	
2		-nko	Si					En número par de moras	
3		-ainko	Si		Si			En sustantivos derivativos de mora par	
4	anan	-anan	Si					En número par de moras	
5		-ananan	Si					En número impar de moras	
6		-(a)naan		Si				Luego de /a/	
7		-(e)naan		Si				Luego de /e/	
8		-(i)naan		Si				Luego de /i/	
9		-(o)naan		Si				Luego de /o/	
10		-naan					Si	Variación inmotivada	
11		-nan					Si	Variación inmotivada	
12		-an					Si	Variación inmotivada	
13		ananan	-ananan	Si					En número de las sílabas impar
14			-anan	Si					En número de las sílabas par
15		ax	-ax		Si				En todos los casos menos ante /a/
16	-x			Si				Antes de una /a/	
17	baon	-baon					Si	Variación inmotivada (Grupo 7 § 5.2)	
18		-boan					Si	Variación inmotivada (Grupo 7 § 5.2)	
19		-boon					Si	Variación inmotivada (Grupo 7 § 5.2)	
20	be	-be				Si	Si	Aparentemente supletivo, se prefiere en el plural	
21		-jo				Si	Si	Aparentemente supletivo, se prefiere en el singular	
22	cha	-cha					Si	Variación inmotivada	

N°	Paradigma	Alomorfos	Fonológico Suprasegmental	Fonológico Segmental	Morfológico	Sintáctico	Sin motivación aparente	Explicación dominio de condicionamiento
23		-chaa					Si	Variación inmotivada
24	chi	chi-					Si	Variación inmotivada
25		tsi-					Si	Variación inmotivada
26	ibat	-ibat		Si				Cuando le sigue una vocal
27		-iba		Si				Cuando le sigue una consonante
28	ina	ina		Si				Antes de consonantes
29		inat		Si				Antes de vocales
30	isi	-isi		Si				Todos excepto cuando termina en / i/
31		-osi		Si				Cuando la raíz termina en /i/
32	katit	-káti	Si					Raíz con moras impares
33		-kátit	Si					Raíz con moras pares
34	kin	-kin	Si					En número par de moras
35		-kiin	Si					En número impar de moras
36	mea	-mea , me		Si				En palabras que terminan con una consonante nasal (esta nasal se pierde)
37		-(ai)nkonia x			Si			En palabras cuya terminación es "ainoa", esta se reemplaza por "ainkonia"
38		-kea		Si				En palabras que acaban en vocal o en /x/
39		-ainoax		Si	Si			En palabras cuyo LOC/ALAT acaba en "ain"
40		-nkoniax		Si	Si			En palabras que cuya marca de LOC tendría que ser "nko"
41		-nkoxon					Si	Variación inmotivada
42		-oa(x)				Si		En palabras que tienen el sufijo "meran"
43	mein	-mein					Si	Variación inmotivada
44		-main					Si	Variación inmotivada

N°	Paradigma	Alomorfos	Fonológico Suprasegmental	Fonológico Segmental	Morfológico	Sintáctico	Sin motivación aparente	Explicación dominio de condicionamiento	
45	forma -n	-n	Si					Forma subyacente (Grupo 1 en § 5.2)	
46		-en	Si	SI			SI	Palabra de dos sílabas, con la segunda sílaba trabada con una sibilante. Alternancia con <i>-an</i> e <i>-in</i> (Grupo 2 en § 5.2)	
47		-an	Si	Si			SI	Palabra de dos sílabas, con la segunda sílaba trabada con una sibilante. Alternancia con <i>-en</i> e <i>-in</i> (Grupo 2 en § 5.2)	
48		-in	Si	Si			SI	Palabra de dos sílabas, con la segunda sílaba trabada con una sibilante y cuando la sílaba anterior termina en /iS/. Alternancia con <i>-en</i> y <i>-an</i> -in (Grupo 2 en § 5.2)	
49		-man	Si	Si				Con dos sílabas terminadas en /n/. Cambia la /n/ a /m/ y se añade <i>-an</i> . (Grupo 3 en § 5.2)	
50		-kan	Si					Palabra terminada en sílaba acentuada y que no contiene k (Grupo 4 en § 5.2).	
51		-tan	Si	SI			SI	Palabra terminada en sílaba acentuada. Alterna con <i>-kan</i> y <i>ten</i> (Grupo 4 en § 5.2)	
52		-ten	Si	SI			SI	Palabra terminada en sílaba acentuada. Alterna con <i>-kan</i> y <i>-tan</i> . (Grupo 4 en § 5.2)	
53		-nin	Si					Palabra tetrasilábica (Grupo 5 en § 5.2)	
54		-ton				Si		Si	Los derivados nominales. Alterna con <i>-tonin</i> (Grupo 8 en § 5.2)
55		-tonin				Si		Si	Los derivados nominales. Alterna con <i>-ton</i> (Grupo 8 en § 5.2)
56	o	-o		Si				Forma subyacente	
57		-yo		Si				Después de segmento sibilante	
58		-sho		Si				Después de /x/	
59	ra	-a		Si				Luego de /x/, la /r/ se elide	
60		-ra		Si				Forma subyacente	
61		r-				Si		Ante la copula <i>iki</i>	

N°	Paradigma	Alomorfos	Fonológico Suprasegmental	Fonológico Segmental	Morfológico	Sintáctico	Sin motivación aparente	Explicación dominio de condicionamiento	
62	riba	-ribi	Si					En palabras con moras pares	
63		-riba	Si					En palabras con moras impares	
64	ronki	-ronki				Si		En cláusulas principales	
65		-ki				Si		En cláusulas subordinadas	
66		-onki		Si				Luego de /x/, la /r/ se elide	
67	ska	-ska		Si				Antes de consonante	
68		-skat		Si				Antes de vocal	
69	t	-(t)	Si	Si			Si	Raíces verbales de dos sílabas terminadas en sílaba abierta, cuando el siguiente sufijo comienza con una vocal. De lo contrario cae la / t/ dejando la mora.	
70		-kaa(t)	Si	Si			Si	Raíces verbales de dos sílabas terminadas en sílaba abierta y cuando el siguiente sufijo comienza con una vocal. De caso contrario cae la / t/	
71		-φ			Si				Antes de un morfema que comience en consonante
72		-mee(t)			Si				Raíces terminadas en /n/ suprimen esta consonante y agregan "-mee(t)"
73		-(k)ee(t)	Si	Si					Raíces verbales de dos sílabas terminadas en sílaba abierta (se sospecha de una armonía vocálica)
74		-(k)ii(t)	Si	Si					Raíces verbales de dos sílabas terminadas en sílaba abierta (se sospecha de una armonía vocálica)
75		-(k)oo(t)	Si	Si					Raíces verbales de dos sílabas terminadas en sílaba abierta (se sospecha de una armonía vocálica)
76	taanan	-taanan	Si					En palabras con sílabas impar	
77		-tanán	Si					En palabras con sílabas par	
78	tibi	-tibi					Si	Variación inmotivada	
79		-tiibi					Si	Variación inmotivada	

N°	Paradigma	Alomorfos	Fonológico Suprasegmental	Fonológico Segmental	Morfológico	Sintáctico	Sin motivación aparente	Explicación dominio de condicionamiento
80	toshi	-toshi		Si				Antes de consonante
81		-toshit		Si				Antes de vocales
82	ya	-ya		Si				Después de cualquiera, menos /i/ o /s/
83		-sha		Si				Después de /i/
84		-a		Si				Después de /s/
85	yama	-yama		Si				Después de cualquiera menos /i/ o /s/
86		-ama		Si				Después de /i/ o /s/
87	yat	-yat		Si				Cuando le sigue una vocal
88		-ya		Si				Cuando le sigue una consonante
89	yora	-yora					Si	Variación inmotivada
90		-(h)ira		Si			Si	Variación inmotivada (si se usa antes de /n/ se agrega una /h/ antes)
91	yosma	-yosma	Si	Si				Después de vocal no acentuada menos /i/
92		-osma		Si				Después de /i/
93		-isma	Si	Si				Cuando la raíz termina en una vocal acentuada

Anexo 3: Criterio 2: Variabilidad formal¹⁷

Paradigma	Alomorfos	Variabilidad formal	Categoría gramatical	Categoría semántica
ananan	-ananan ~ -anan	2	Sufijo verbal	Reciproco
ax	-ax ~ -x	2	Enclítico	Concordancia con S
be	-be ~ -jo	2	Sufijo verbal	Venitivo
cha	-cha ~ -chaa	2	Enclítico	Aditivo
chi	chi- ~ tsi-	2	Prefijo	Nalgas
ibat	-ibat ~ -iba	2	Sufijo verbal	Ayer, hace unos pocos días.
ina	ina ~ inat	2	Sufijo verbal	Yendo arriba
isi	-isi ~ -osi	2	Enclítico	Despreciativo
katit	-káti ~ -kátit	2	Sufijo verbal	Pasado distante
kin	-kin ~ -kiin	2	Sufijo verbal	Asociativo
mein	-mein ~ -main	2	Enclítico	Especulativo
ribi	-ribi ~ -riba	2	Enclítico	También
ska	-ska ~ -skat	2	Enclítico	Comparativo
taanan	-taanan ~ -tanán	2	Enclítico	Sujetos idénticos
tibi	-tibi ~ -tiibi	2	Sufijo nominal	Distributivo
toshi	-toshi ~ -toshit	2	Sufijo verbal	Repentino / sorpresivo

¹⁷ Esta tabla no contiene numeración correlativa debido a que fue ordenada por su valor de variabilidad y no por el orden alfabético en que se presentan los otros criterios.

yama	-yama ~ -ama	2	Sufijo verbal	Negador
yat	-yat ~ -ya	2	Sufijo verbal	Mañana
yora	-yora ~ -(h)ira	2	Enclítico	Intensificador
ain	-ain ~ -nko ~ -ainko	3	Enclítico	Locativo / allativo
baon	-baon ~ -boan ~ -boon	3	Enclítico	Plural ERG
o	-o ~ -yo ~ -sho	3	Enclítico	Locativo aproximativo
ra	-a ~ -ra ~ r-	3	Clítico	Evidencial directo
ronki	-ronki ~ -ki ~ -onki	3	Enclítico	Reportativo 'dicen'
ya	-ya ~ -sha ~ -a	3	Enclítico	Propietivo
yosma	-yosma ~ -osma ~ -isma	3	Sufijo verbal	Adverbializador "uno que nunca"
mea	-mea , me ~ -(ai)nkoniak ~ -kea ~ -ainoax ~ -nkoniak ~ -nkoxon ~ -oa(x)	7	Enclítico	Ablativo
t	-(t) ~ -kaa(t) ~ -φ ~ -mee(t) ~ -(k)ee(t) ~ -(k)ii(t) ~ -(k)oo(t)	7	Sufijo verbal	Intransitivizador
anan	-anan ~ -ananan ~ -(a)naan ~ -(e)naan ~ -(i)naan ~ -(o)naan ~ -naan ~ -nan ~ -an	9	Sufijo verbal	Malefactivo
n	-n ~ -en ~ -an ~ -in ~ -man ~ -kan ~ -tan ~ -ten ~ -nin ~ -ton ~ -tonin	11	Enclítico	Forma -n

Las categorías semánticas ennegrecidas tienen una traducción directa.

Anexo 4. Criterio 3: Predictibilidad

N°	Paradigma	Alomorfos	Explicación dominio de condicionamiento	Totalmente predecible	Parcialmente predecible	Nada predecible	Explicación predictibilidad parcial
1	ain	-ain	En número impar de moras	Si			
2		-nko	En número par de moras				
3		-ainko	En sustantivos derivativos de mora par				
4	anan	-anan	En número par de moras		Si		Un conjunto de alomorfos (-naan, -nan y -an) sin motivación aparente
5		-ananan	En número impar de moras				
6		-(a)naan	Luego de /a/				
7		-(e)naan	Luego de /e/				
8		-(i)naan	Luego de /i/				
9		-(o)naan	Luego de /o/				
10		-naan	Variación inmotivada				
11		-nan	Variación inmotivada				
12		-an	Variación inmotivada				
13	ananan	-ananan	En número de las sílabas impar	Si			
14		-anan	En número de las sílabas par				
15	ax	-ax	En todos los casos menos ante /a/	Si			
16		-x	Antes de una /a/				
17	baon	-baon	Variación inmotivada (Grupo 7 § 5.2)			Si	
18		-boan	Variación inmotivada (Grupo 7 § 5.2)				
19		-boon	Variación inmotivada (Grupo 7 § 5.2)				
20	be	-be	Aparentemente supletivo, se prefiere en el plural		Si		Prefiere el plural
21		-jo	Aparentemente supletivo, se prefiere en el singular				
22	cha	-cha	Variación inmotivada			Si	

N°	Paradigma	Alomorfos	Explicación dominio de condicionamiento	Totalmente predecible	Parcialmente predecible	Nada predecible	Explicación predictibilidad parcial
23		-chaa	Variación inmotivada				
24	chi	chi-	Variación inmotivada			Si	
25		tsi-	Variación inmotivada				
26	ibat	-ibat	Cuando le sigue una vocal	Si			
27		-iba	Cuando le sigue una consonante				
28	ina	ina	Antes de consonantes	Si			
29		inat	Antes de vocales				
30	isi	-isi	Todos excepto cuando termina en / i/	Si			
31		-osi	Cuando la raíz termina en /i/				
32	katit	-káti	Raíz con moras impares	Si			
33		-kátit	Raíz con moras pares				
34	kin	-kin	En número par de moras	Si			
35		-kiin	En número impar de moras				
36	mea	-mea , me	En palabras que terminan con una consonante nasal (esta nasal se pierde)		Si		Con un alomorfo (-nkoxon) sin motivación aparente
37		-(ai)nkoniax	En palabras cuya terminación es "ainoa", esta se reemplaza por "ainkonionia"				
38		-kea	En palabras que acaban en vocal o en /x/				
39		-ainoax	En palabras cuyo LOC/ALAT acaba en "ain"				
40		-nkoniax	En palabras que cuya marca de LOC tendría que ser "nko"				
41		-nkoxon	Variación inmotivada				
42		-oa(x)	En palabras que tienen el sufijo "meran"				
43	mein	-mein	Variación inmotivada			Si	
44		-main	Variación inmotivada				

N°	Paradigma	Alomorfos	Explicación dominio de condicionamiento	Totalmente predecible	Parcialmente predecible	Nada predecible	Explicación predictibilidad parcial
45	forma -n	-n	Forma subyacente (Grupo 1 en § 5.2)		Si		Conjuntos de alomorfos (Grupo 2: -en, -an e -in; grupo 4: -tan y -ten; grupo 8: -ton, -tonin) alternando sin motivación aparente.
46		-en	Palabra de dos sílabas, con la segunda sílaba trabada con una sibilante. Alternancia con <i>-an</i> e <i>-in</i> (Grupo 2 en § 5.2)				
47		-an	Palabra de dos sílabas, con la segunda sílaba trabada con una sibilante. Alternancia con <i>-en</i> e <i>-in</i> (Grupo 2 en § 5.2)				
48		-in	Palabra de dos sílabas, con la segunda sílaba trabada con una sibilante y cuando la sílaba anterior termina en /iS/. Alternancia con <i>-en</i> y <i>-an</i> -in (Grupo 2 en § 5.2)				
49		-man	Con dos sílabas terminadas en /n/. Cambia la /n/ a /m/ y se añade -an. (Grupo 3 en § 5.2)				
50		-kan	Palabra terminada en sílaba acentuada y que no contiene k (Grupo 4 en § 5.2).				
51		-tan	Palabra terminada en sílaba acentuada. Alterna con <i>-kan</i> y <i>ten</i> (Grupo 4 en § 5.2)				
52		-ten	Palabra terminada en sílaba acentuada. Alterna con <i>-kan</i> y <i>-tan</i> . (Grupo 4 en § 5.2)				
53		-nin	Palabra tetrasilábica (Grupo 5 en § 5.2)				
54		-ton	Los derivados nominales. Alterna con <i>-tonin</i> (Grupo 8 en § 5.2)				
55		-tonin	Los derivados nominales. Alterna con <i>-ton</i> (Grupo 8 en § 5.2)				
56	o	-o	Forma subyacente	Si			
57		-yo	Después de segmento sibilante				
58		-sho	Después de /x/				
59	ra	-a	Luego de /x/, la /r/ se elide	Si			
60		-ra	Forma subyacente				
61		r-	Ante la copula <i>iki</i>				
62	riba	-ribi	En palabras con moras pares	Si			

N°	Paradigma	Alomorfos	Explicación dominio de condicionamiento	Totalmente predecible	Parcialmente predecible	Nada predecible	Explicación predictibilidad parcial
63		-riba	En palabras con moras impares				
64	ronki	-ronki	En cláusulas principales	Si			
65		-ki	En cláusulas subordinadas				
66		-onki	Luego de /x/, la /r/ se elide				
67	ska	-ska	Antes de consonante	Si			
68		-skat	Antes de vocal				
69	t	-(t)	Raíces verbales de dos sílabas terminadas en sílaba abierta, cuando el siguiente sufijo comienza con una vocal. De lo contrario cae la / t/ dejando la mora.				Con un grupo de alomorfos (-t) y -kaa(t) en alternancia sin motivación aparente.
70		-kaa(t)	Raíces verbales de dos sílabas terminadas en sílaba abierta y cuando el siguiente sufijo comienza con una vocal. De caso contrario cae la / t/				
71		-φ	Antes de un morfema que comience en consonante				
72		-mee(t)	Raíces terminadas en /n/ suprimen esta consonante y agregan "-mee(t)"				
73		-(k)ee(t)	Raíces verbales de dos sílabas terminadas en sílaba abierta (se sospecha de una armonía vocálica)				
74		-(k)ii(t)	Raíces verbales de dos sílabas terminadas en sílaba abierta (se sospecha de una armonía vocálica)				
75		-(k)oo(t)	Raíces verbales de dos sílabas terminadas en sílaba abierta (se sospecha de una armonía vocálica)				
76	taanan	-taanan	En palabras con sílabas impar	Si			
77		-tanán	En palabras con sílabas par				
78	tibi	-tibi	Variación inmotivada			Si	
79		-tiibi	Variación inmotivada				
80	toshi	-toshi	Antes de consonante	Si			

N°	Paradigma	Alomorfos	Explicación dominio de condicionamiento	Totalmente predecible	Parcialmente predecible	Nada predecible	Explicación predictibilidad parcial
81		-toshit	Antes de vocales				
82	ya	-ya	Después de cualquiera, menos /i/ o /s/	Si			
83		-sha	Después de /i/				
84		-a	Después de /s/				
85	yama	-yama	Después de cualquiera menos /i/ o /s/	Si			
86		-ama	Después de /i/ o /s/				
87	yat	-yat	Cuando le sigue una vocal	Si			
88		-ya	Cuando le sigue una consonante				
89	yora	-yora	Variación inmotivada			Si	
90		-(h)ira	Variación inmotivada (si se usa antes de /n/ se agrega una /h/ antes)				
91	yosma	-yosma	Después de vocal no acentuada menos /i/	Si			
92		-osma	Después de /i/				
93		-isma	Cuando la raíz termina en una vocal acentuada				

Anexo-5. Criterio 4: Semejanza formal

N°	Paradigma	Alomorfos	Todos parecidos entre sí	Parcialmente parecidos entre sí	Ningún parecido entre sí
1	ain	-ain / -nko / -ainko		Si	
2	anan	-anan / -ananan / -(a)naan / -(e)naan / -(i)naan / -(o)naan / -naan / -nan / -an	Si		
3	ananan	-ananan / -anan	Si		
4	ax	-ax / -x	Si		
5	baon	-baon / -boan / -boon	Si		
6	be	-be / -jo			Si
7	cha	-cha / -chaa	Si		
8	chi	chi- / tsi-	Si		
9	ibat	-ibat / -iba	Si		
10	ina	ina / inat	Si		
11	isi	-isi / -osi	Si		
12	katit	-káti / -kátit	Si		
13	kin	-kin / -kiin	Si		
14	mea	-mea , me / -(ai)nkoniax / -kea / -ainoax / -nkoniax / -nkoxon / -oa(x)		Si	
15	mein	-mein / -main	Si		
16	n	-n / -en / -an / -in / -man / -kan / -tan / -ten / -nin / -ton / -tonin		Si	
17	o	-o / -yo / -sho	Si		
18	ra	-a / -ra / r-	Si		
19	ribi	-ribi / -riba	Si		
20	ronki	-ronki / -ki / -onki		si	
21	ska	-ska / -skat	Si		
22	t	-(t) / -kaa(t) / -φ / -mee(t) / -(k)ee(t) / -(k)ii(t) / -(k)oo(t)		Si	
23	taanan	-taanan / -tanán	Si		
24	tibi	-tibi / -tiibi	Si		

25	toshi	-toshi / -toshit	Si		
26	ya	-ya / -sha / -a	Si		
27	yama	-yama / -ama	Si		
28	yat	-yat / -ya	Si		
29	yora	-yora / -(h)ira		Si	
30	yosma	-yosma / -osma / -isma		Si	



LISTA DE ABREVIATURAS

1	primera persona singular
1P	primera persona plural
ABS	absolutivo
ABL	ablativo
ASSOC	asociativo
AUX	auxiliar
CAUS	causativo
CMPL	aspecto completivo
COP	cópula
ERG	ergativo
EV	evidencial directo
HYS2	segundo evidencial reportativo
GEN	genitivo
I	intransitivo (orientación al objeto)
IMP	imperativo
INC	aspecto incompletivo
INF	infinitivo
LOC	locativo
NEG	negativo
NMLZ	nominalizador
PL	plural
POS1	posesivo de primera persona singular
POS3	posesivo de tercera persona singular
P-DS	evento previo, diferentes sujetos
PSSS	evento previo igual sujeto orientado a S
REC	recíproco
REM	pasado remoto
SIML	similitivo
SPECL	especulativo

REFERENCIAS BIBLIOGRÁFICAS

ABATE, Mesfin y Yaregal ASSABIE

- 2014 “Development of Amharic morphological analyzer using memory-based learning”. En PRZEPIÓRKOWSKI Adam y Maciej OGRODNICZUK (editores). *Advances in Natural Language Processing.* Springer, Cham. pp 1-13.

ANDERSON, Stephens

- 1992 *A Morphous Morphology*. New York: Oxford University Press.
- 2005 *Aspects of the Theory of Clitics*. New York: Oxford University Press.

AIKHENVALD, Alexandra

- 2003 “Typological parameters for the study of clitics, with special reference to Tariana”. En DIXON R. M. W. y Alexandra Y. AIKHENVALD (editores). *Word. A cross-linguistic typology*. New York: Cambridge University Press.

BENDER, Emily M. y D. Terence LANGENDOEN

- 2010 “Computational Linguistics in Support of Linguistic Theory”. En *Linguistic Issues in Language Technology*, 3 . pp 1-31.

BERENGUER, Jaime

- 1999 *Gramática griega*. Barcelona: Bosch.

BERMENT, Vicent

- 2002 “Several directions for minority languages computerization”. En *Proceedings of the 19th international conference on Computational linguistics-Volume 2*. pp. 1-5.

CERRÓN-PALOMINO, Rodolfo

- 2006 *El chipaya o la lengua de los hombres del agua*. Lima: Fondo Editorial de la Pontificia Universidad Católica del Perú.
- 2008 *Quechumara Estructuras paralelas del Quechua y del aimara*. La Paz: Plural editores.

CORBETT, Greville

- 2007 “Canonical typology, suppletion and possible words”. En *Language*, 83. pp. 8-42
- 2009 “Suppletion: Typology, markedness, complexity”. En STEINKRÜGER, MKP (editor). *On Inflection*. Berlin : Mouton de Gruyter , pp. 25-40.

CRYSTAL, David

- 2000 *Diccionario de lingüística y fonética*. España: Octaedro.

DIXON R. M. W. Y Alexandra AIKHENVALD

- 2003 “Word: a typological framework”. En DIXON, R. M. W. y Alexandra Y. AIKHENVALD (editores). *Word. A cross-linguistic typology*. New York: Cambridge University

ESCANDELL, María Victoria

2011 *Invitación a la lingüística*. Editorial: Universitaria Ramón Areces.

ESPICHÁN, Alexandra y Arturo ONCEVAY

2017 “A low-resourced peruvian language identification model Workshop”.
En *Proceedings of the SIMBig 2017 Track on Applied Natural Language Processing, ANLP 2017*. Springer. In-press.

2018 “Language identification with scarce data: A case study from Peru”.
Conferencia *In Information Management and Big Data: Fourth Annual International Symposium, SIMBig 2017*, Lima, Peru, September 4-6, 2017, Revised Selected Papers. Springer International Publishing. (In-press).

FÁBREGAS, Antonio

2013 *La morfología. El análisis de la palabra compleja*. Madrid: Síntesis.

FAUST, Norma

1973 *Lecciones para el aprendizaje del idioma shipibo-conibo* (No. 1). Instituto Lingüístico de Verano.

FORCADA, Mikel.

2006 “Open-source machine translation: an opportunity for minor languages”. En *Proceedings of the Workshop Strategies for developing machine translation for minority languages*, LREC Vol. 6. pp. 1-6.

GALARRETA, Ana Paula, Andrés MELGAR y Arturo ONCEVAY

2017 “Corpus creation and initial SMT experiments between Spanish and Shipibo-konibo”. Conferencia. In Proceedings of the International Conference on Recent Advances in Natural Language Processing, RANLP 2017

HASPELMATH, Martin

2002 *Understanding Morphology*. London: Arnold.

HUALDE, José Ignacio

2014 “La entonación”. En NUÑEZ-CEDENO, Rafael, Sonia COLINA y Travis BRADLEY (editores). *Fonología generativa contemporánea de la lengua española*. Georgetown University Press. pp 267-290.

KATAMBA, Francis

1993 *Morphology*. London: Palgrave Macmillan. pp 17-19;

LORIOT, James, Erwin LAURIAULT, and Dwight DAY.

1993 *Diccionario shipibo-castellano*. Yarinacocha, Pucallpa: Ministerio de Educación del Perú y El Instituto Lingüístico de Verano.

MAGUIÑO, Diego, Arturo ONCEVAY y Marco SOBREVILLA

2018 “WordNet-Shp: Towards the building of a lexical database for a Peruvian minority language”. Conferencia In Proceedings of the

Eleventh International Conference on Language Resources and Evaluation (LREC 2018). (In-press).

MARTÍNEZ CELDRÁN, Eugenio

1998 *Lingüística: teoría y aplicaciones*. Elsevier España.

MERCADO Rodolfo, José PEREIRA, Marco SOBREVILLA y Arturo ONCEVAY

2018 “ChAnot: An intelligent annotation tool for indigenous and highly agglutinative languages in Peru” Conferencia. En *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC 2018). (In-press).

MINISTERIO DE EDUCACIÓN

2013 “AHORA, CON ESTAS LETRAS VAMOS A ANDAR” Sistematización del proceso de normalización de los alfabetos de las lenguas originarias del Perú. Lima

MORALES-FRONT, Alfonso

2014 “El acento”. En NUÑEZ-CEDENO, Rafael, Sonia COLINA y Travis BRADLEY (editores). *Fonología generativa contemporánea de la lengua española*. Georgetown University Press. pp 235-264.

MORENO, Antonio

2017 “Procesamiento del lenguaje natural ¿qué es?” Instituto de Ingeniería del Conocimiento. Consulta Consulta 18 junio de 2018 <<http://www.iic.uam.es/inteligencia/que-es-procesamiento-del-lenguaje-natural/>>

MUNGUÍA, Irma, Marta MUNGUÍA y Guilda ROCHA

2004 *Larousse: Gramática Lengua española*. México D.F.: Larousse

NEVINS, Andrew

2010 “Phonologically-Conditioned Allomorph Selection”. En EWEN, Colin, Elizabeth HUME, Marc VAN OOSTENDORP, Keren RICE. (editors). *The Blackwell Companion to Phonology*. 4. pp. 2357-2382.

NUÑEZ-CEDENO, Rafael

2014 “Glosario”. En NUÑEZ-CEDENO, Rafael, Sonia COLINA y Travis BRADLEY (editores). *Fonología generativa contemporánea de la lengua española*. Georgetown University Press. pp 389-401.

PAYNE, Thomas

1997 *Describing Morphosyntax: Guide for Field Linguists*. New York: Cambridge University Press,

SOMERS, Harold

1997 “Machine translation and minority languages”. En *Translating and the Computer*.

VALENZUELA, Pilar

- 2000 “Ergativity escindida en wariapano, yaminawa y shipibo-konibo”. En H. van der Voort and S. van de Kerke (eds.). *Indigenous languages of Lowland South America*. Leiden: University of Leiden. pp. 111-128.
- 2002a. “Causativization and Transitivity in Shipibo-Konibo”. En SHIBATANI, Masayoshi (ed.). *The grammar of causation and interpersonal manipulation*. Amsterdam: John Benjamins. pp. 417-483.
- 2002b. “Relativization in Shipibo-Konibo: A typologically-oriented study”. LINCOM Studies in Native American Linguistics 42. München: Lincom Europa.
- 2003a “Evidentiality in Shipibo-Konibo, with a Comparative Overview of the Category in Panoan”. En AIKHENVALD, A. Y. y R. M. W. DIXON (eds.). *Studies in Evidentiality*. Amsterdam: John Benjamins. pp , 33-61.
- 2003b *Transitivity in Shipibo-Konibo Grammar*. Tesis doctoral Ph.D. grammar in linguistics, University of Oregon, Eugene.
- 2005 “Participant agreement in Panoan”. En HIMMELMANN, N. P. y E. SCHULTZE-BERNDT (eds.). *Secondary predication and adverbial modification: The typology of depictives*. Oxford: Oxford University Press. pp , 259-298.
- 2010a “Applicative constructions in Shipibo-Konibo (Panoan)”. En *International Journal of American Linguistics* 76. pp 101-144.
- 2010b “Ergativity in Shipibo-Konibo”. En S. Gildea y F. Queixalós (eds), *Ergativity in Amazonia*. Amsterdam: John Benjamins. pp. 65-96.
- 2011a “Argument Encoding and Pragmatic Marking of the ‘Ergative’ in Shiwilu (Kawapanan)”. En *International Journal of American Linguistics* 77: pp91-120
- 2011b “Multi-verb predicates and transitivity harmony in Shipibo-Konibo”. En MUYSKEN, P.C. y A. Y. AIKHENVALD (eds.). *Multi-verb Constructions: A view from the Americas*. Studies in the Indigenous Languages of the Americas, Amsterdam: Brill. 185-212.

VALENZUELA, Pilar, Luis MÁRQUEZ PINEDO e Ian MADDIESON.

2001a. “Illustrations of the IPA: Shipibo”. En *Journal of the International Phonetic Association* 31.2: pp 281-85.

VALENZUELA, Pilar M. y Oliver A. IGGESSEN

2007 “El desarrollo de un marcador suprasegmental en chácobo (pano)”. En ROMERO-FIGUEROA, Andrés, Ana Fernández GARAY, y Ángel CORBERA MORI (eds.). *Lenguas Indígenas de América del Sur: Estudios descriptivo-tipológicos y sus contribuciones para la Lingüística teórica*. Caracas: Universidad Católica Andrés Bello. pp,187-199

VARELA, Soledad

1996 *Fundamentos de la morfología*. Madrid: EDITORIAL SÍNTESIS

2005 *Morfología léxica: la formación de palabras*. Madrid: Gredos.

VILLAYANDRE, Milka

2000 “La lingüística computacional y sus aplicaciones”. En *Revista Estudios Humanísticos. Filología* Número 22. pp 195-220.

2010 *Aproximación a la lingüística computacional*. Tesis doctoral. León: Universidad de León.

XIA, Fei y William LEWIS

2008 “Repurposing Theoretical Linguistic Data for Tool Development and Search”. En *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*.

ZARIQUIEY, Roberto

2018 Separatas curso 'SEMINARIO DE MÉTODOS COMPUTACIONES

PARA EL ESTUDIO DE LENGUAS AMERINDIAS' PUCP

