

Bases de la documentación lingüística

Instituto Nacional de Lenguas Indígenas

E. Fernando Nava L.
Director General

Editores de la versión inglesa:

Jost Gippert
Nikolaus Himmelmann
Ulrike Mosel

John B. Haviland
José Antonio Flores Farfán
Coordinadores de la versión en español

de las ideologías. El problema para nosotros consiste en transmitir a nuestros estudiantes estas ideas de manera tan clara y contundente como lo es su formación en fonología, morfología y sintaxis. Espero que tengamos éxito en esta empresa. Así como los recientes avances en tipología lingüística han facilitado enormemente el reconocimiento de las estructuras lingüísticas que encontramos en la investigación de campo, los avances en el estudio de los procesos culturales pueden ayudarnos a organizar nuestra investigación y a trabajar más exitosamente como lingüistas y amigos, como colegas y defensores de las comunidades lingüísticas minoritarias.

Capítulo 6

La documentación del conocimiento léxico

John B. Haviland

Introducción

La lexicografía, es decir, la práctica de documentar tanto los significados como los usos de las "palabras" (literalmente, registrándolas "por escrito"), es quizás, gracias a sus productos, la rama más conocida de la lingüística para el público en general. También es una actividad antigua y sobre la que mucho se ha teorizado. En la trilogía boasiana para la descripción de una lengua, consistente en gramática, vocabulario y texto, es sin lugar a dudas la compilación del diccionario la parte más laboriosa. El proceso comienza con los primeros encuentros del aprendiz con el idioma y pareciera no acabar nunca. Peor aún, es una labor repleta de dudas con respecto a los límites del trabajo, tanto para el conjunto—¿cuándo se decide que las palabras básicas o más comunes de una variedad lingüística han sido capturadas y caracterizadas?—como para cualquier posible entrada del diccionario, dada la aparentemente infinita variedad de matices y ámbitos de las palabras y las formas, sin mencionar la idiosincrasia de las expresiones compuestas o derivadas. Además, a pesar de una buena dosis de especulación desde muchas tradiciones lingüísticas dispares sobre qué recursos metasemánticos se deben emplear para capturar los significados, a pesar de los múltiples modelos y ejemplos de los resultados de la elaboración de un diccionario y a pesar de la amplia experiencia, para la mayoría de nosotros, en la tarea usual de "explicar los significados de las palabras", sigue existiendo la posibilidad de que la duda nos asalte a cada paso: ¿fue suficiente lo que dijimos?, ¿olvidamos algo?, ¿entendimos bien siquiera esta simple palabra?

El presente capítulo introduce técnicas y conceptos relevantes en la producción de una base de datos léxicos como parte de un proyecto de documentación lingüística. Me concentro en una serie de obstáculos que son

fuentes de dudas para el lexicógrafo de campo, con algunas sugerencias, si no para superarlos, al menos para enfrentarlos. Mi cobertura es deliberadamente parcial. Me apoyo en buena medida en mi propio trabajo de campo en México y Australia para considerar tres cuestiones generales. En primer lugar, reviso algunas moralejas conocidas sobre la naturaleza del significado de las palabras, derivadas de conceptos de la filosofía lingüística, que son fáciles de olvidar al calor del trabajo lexicográfico. En segundo lugar, considero los metalenguajes semánticos planteados para tratar con diferentes tipos de elementos significativos, desde los "funcionales" hasta los léxicos, y desde las raíces hasta las bases léxicas. En tercer lugar, reviso ciertas técnicas para extraer sistemáticamente el conocimiento léxico. Obviaré en gran medida varios temas afines que son importantes: la variación léxica y cómo representarla (véase el capítulo 5), cuestiones ideológicas inevitablemente relacionadas con la promulgación de cualquier diccionario (véase nuevamente el capítulo 5 y las discusiones en Frawley *et al.* 2002) y otras cuestiones más generales de la teoría semántica léxica (las relaciones de sentido, los problemas de extensión vs. intensión, etc.), que subyacen a cualquier práctica lexicográfica, pero que están más allá del alcance de este trabajo. Comienzo con una revisión sumamente selectiva de los materiales publicados sobre el conocimiento léxico, especialmente en la medida en que son relevantes para la documentación de lenguas en peligro de extinción.

1. La lexicografía y sus productos

Además de una amplia bibliografía teórica sobre el significado, hay una tradición práctica de elaboración de diccionarios que ha hecho proliferar manuales e historias, así como ensayos sobre el oficio del lexicógrafo; pero éstos raramente ayudan al investigador de campo.

El lexicón, en la lingüística moderna, ha llegado a significar un depósito de hechos por lo demás anárquicos, un inventario de paquetes arbitrarios formados por pronunciaciones y puñados de rasgos semánticos distintivos. Es el lugar donde la lengua almacena sus idiosincrasias e irregularidades. La sistematicidad que pueda llegar a tener un lexicón concebido de esta forma proviene de los propios sistemas de rasgos distintivos, considerados como representantes de los patrones sintácticos y semánticos que subyacen a las formas léxicas superficiales. Para estudiar estos patrones está la acostumbrada provincia de la semántica léxica, que cataloga las diversas relaciones entre los sentidos de los miembros de diferentes subconjuntos de formas léxicas (Cruse 1986), las propiedades

sistemáticas de las clases de palabras superficiales o "partes del habla", los hechos de la estructura argumental, la diátesis y demás. La principal contribución a la teoría lingüística de parte de la lexicografía empírica ha sido la elucidación de las interrelaciones semánticas y sintácticas a nivel de la realización de la palabra individual (Levin 1993).

La lingüística de campo, antaño dominio de los lingüistas antropológicos, dio pie a gran parte del aparato conceptual subyacente en la semántica léxica. Las primeras teorías buscaban una analogía entre los rasgos fonológicos y los "componentes" del sentido en conjuntos estructurados de "terminología popular", desde el parentesco hasta la etnobotánica, desde los sistemas pronominales hasta las tipologías verbales. Los estudios clásicos de la "etnociencia" investigaban sistemas léxicos culturalmente elaborados, en particular en dominios "naturales" como la etnobotánica. Fuente de mayor inspiración empírica para la teorización semántica fueron, por ejemplo, las lenguas de los aborígenes australianos, célebres por su agudeza lingüística y genio creador. En dyirbal, por ejemplo, la semántica verbal y las propiedades del vocabulario especial "para la suegra", usado para mostrar respeto a los parientes políticos, llevaron a Dixon (1971) a postular una diferencia fundamental entre las palabras semánticamente básicas o "nucleares", que requieren algún tipo de descomposición en dimensiones subléticas significativas, y las palabras no nucleares, que podían ser *definidas* en términos de las palabras nucleares combinadas con otros recursos de la gramática. Los juegos rituales del lenguaje que aprendían los warlpiri y lardil iniciados sugerían que los etnolingüistas aborígenes habían desarrollado sofisticados análisis semánticos del vocabulario común (Hale 1971, 1982).

La referencia clásica en lexicografía es Zgusta (1971).¹ De especial interés para el lexicógrafo de campo es el libro de Frawley *et al.* (2002), una colección de ensayos escritos por lexicógrafos que trabajan con lenguas amerindias y que también consideran los problemas de *crear* una práctica lexicográfica en comunidades que carecen de ella.² Abarcan desde cuestiones teóricas de la semántica léxica (el carácter de la definición, el rango del conocimiento léxico que poseen los hablantes o que podría incluir un diccionario y la interrelación entre los hechos léxicos diacrónicos y sincrónicos) hasta las cuestiones de la forma representacional o cuestiones

¹Especialmente en referencia a la tradición europea escrita, son útiles los estudios de Landau (1984) y Svensén (1993). Véase también el manual en varios volúmenes editado por Hausmann *et al.* (1990-1991).

²Aunque idiomas como el náhuatl gozan de sus propias tradiciones de diccionarios que tienen siglos de antigüedad (Canger 2002; Amith 2002).

sociopolíticas en la elaboración de un diccionario (para quién se compila un diccionario y con qué propósitos, o qué tipos de categorías sociolingüísticas, géneros de habla especializados, formas léxicas específicas de género o clase, por ejemplo --habrán de distinguirse). Estos trabajos van mucho más allá de la limitada selección de temas que aquí se tratan.

El lingüista de campo no necesita ser un semantista, excepto "con fines prácticos", y la lexicografía al servicio de la documentación necesita encontrar un equilibrio entre aspiraciones opuestas. Por ejemplo, ¿en qué sentido la "exhaustividad" --como quiera que pueda esto definirse para una lengua amenazada-- es algo a lo que deba aspirarse? ¿Y qué decir sobre la mezcla de motivaciones teóricas y prácticas en los metalenguajes usados para representar la información léxica? En el campo, uno debe valerse de todos los trucos posibles: la elaboración de los diccionarios bilingües, por ejemplo, a menudo puede comenzar con listas de palabras ya existentes, ya sea en la lengua de origen o en la lengua meta, y no hay razón para encerrarse tras el purismo o estrictos principios metodológicos al generar lexemas para su incorporación en una base de datos léxicos.

Los diferentes productos lexicográficos reflejan diferentes puntos de partida y objetivos de los compiladores de bases de datos léxicos. Zgusta (1971) trata en capítulos separados los problemas relacionados con la compilación de diccionarios plurilingües (generalmente bilingües) y monolingües. El contraste y la elección de qué lenguas incluir en un diccionario multilingüe hacen surgir preguntas obvias: ¿para qué tipo de uso se produce una base de datos léxicos?, ¿qué conocimientos de parte del usuario se dan por sentados en su diseño?, ¿por qué lo produjo su compilador, por principio de cuentas? Permítanme reseñar diferentes tipos de diccionarios de campo relacionados con mi propia investigación en México y Australia. Me han sido especialmente útiles las introducciones a los dos diccionarios de tzotzil de Robert M. Laughlin (1975, 1988), uno moderno y el otro basado en una obra del siglo XVI.

En lo que llamo la tradición colonial, la recolección de vocabularios invariablemente era una vocación imperialista, a menudo un subproducto de la exploración y la conquista. Los exploradores recogían flora y fauna, y a menudo también palabras. Un tanto menos inocentes eran las listas de palabras creadas explícitamente para ayudar en la conversión, la conquista y el control. Los diccionarios de los idiomas indígenas del Nuevo Mundo elaborados por los frailes o los vocabularios vernáculos destinados a los burócratas coloniales en África e India representaban una "documentación" descaradamente instrumental, comúnmente de aquellas lenguas que con el

paso del tiempo corrían el riesgo de desaparecer a consecuencia de la expansión colonial original. Estas listas de palabras decididamente no estaban hechas "para" los hablantes de los idiomas así documentados. La tradición de los misioneros sigue produciendo muchos diccionarios de campo, y leerlos da cierta idea de los propósitos y las poblaciones beneficiadas por esta práctica lexicográfica en particular. En Chiapas, México, el Instituto Lingüístico de Verano --una organización protestante que traduce el Nuevo Testamento-- ha publicado muchos diccionarios de lenguas indígenas de la región (Delgaty y Ruiz 1978, para el tzotzil; Aulie y Aulie 1978, para el chol, por mencionar sólo dos), muy utilizados incluso por hablantes que no comparten las creencias religiosas de los traductores. Estos diccionarios están sutilmente permeados por el comentario metacultural y la ideología religiosa.

Aquí tenemos, por ejemplo, una traducción de la entrada en Aulie y Aulie (1978) de la palabra chol *ajaw*, reflejo de una raíz que significa "señor, amo, Dios" en otros idiomas mayas. De acuerdo con los Aulie, la palabra chol significa "espíritu malo de la tierra", y luego comentan:

Lo llaman *lak' tat* "nuestro padre". Se cree que una persona puede hacer un pacto con él. Dicha persona puede hacer peticiones al espíritu a favor o en contra de otra persona. La persona que establece dichas relaciones con el *ajaw* es llamado "sacristán". Si un hombre o una mujer ofenden al sacristán, éste llama al espíritu para maldecir a la otra persona, que en poco tiempo morirá.

Aquí la voz de los lexicógrafos y su acento ideológico subyacente están claramente expuestos. Como se ve, para los Aulie no hay ninguna disonancia aparente entre la paráfrasis propuesta, "espíritu malo de la tierra" y la locución alterna "nuestro padre" (con un prefijo de primera persona del plural inclusivo). Aun el "ellos" del comentario ("lo llaman") hace referencia a otras personas que no son los autores del diccionario (aunque tal vez sean las mismas que lo utilizan). Obsérvese finalmente un interesante contraste en la emergencia de voces distintas. La posibilidad de "hacer un pacto" con el *ajaw* se cita como algo que "se erree" (algo que supuestamente "ellos" creen). En cambio, las consecuencias de que el hipotético *sacristán* (el término mismo es un préstamo del español que se introdujo al chol durante la conversión católica de los hablantes de chol después de la conquista) llame al *ajaw* reciben una categoría epistemológica diferente: "la otra persona ... en poco tiempo morirá". El diccionario

incorpora, pues, posiciones diferentes y quizás mutuamente contradictorias ante las creencias y prácticas choles, dentro de la misma entrada léxica.

Un tanto distinta es la tradición lexicográfica "etnolingüística", cuyos orígenes inmediatos se encuentran en la investigación etnográfica. Regresando a Los Altos de Chiapas, el exhaustivo diccionario de Laughlin del tzotzil contemporáneo de Zinacantán (1975) tiene la forma de un diccionario bilingüe tradicional (tzotzil-inglés). La primera sección ofrece extensas paráfrasis en inglés de palabras tzotziles, tanto derivadas como simples, ordenadas según sus supuestas raíces subyacentes. Luego sigue un índice en inglés a la sección en tzotzil. El diccionario de Laughlin tiene más de 35 mil entradas del tzotzil al inglés, por lo que es uno de los diccionarios más amplios de una lengua indígena americana. Sin embargo, es un diccionario bilingüe en tzotzil e inglés, lo que limita su uso directo al reducido grupo de personas que hablan ambos idiomas.³ También es un diccionario con evidentes sesgos dialectales e incluso de género, pues documenta la forma en que hablaban tzotzil los hombres de mediana edad durante las décadas de 1960 y 1970 exclusivamente en el municipio de Zinacantán, lo que puede considerarse una variante minoritaria de lo que desde entonces se ha convertido en una lengua indígena dominante en Los Altos de Chiapas, con un número mucho mayor de hablantes de otros dialectos. Así pues, la elección de la variedad lingüística de un diccionario refleja los accidentes de la investigación que lo sustenta, y no tanto un plan con principios lexicográficos o sociolingüísticos. Además, como este diccionario agrupa las entradas según una raíz teórica subyacente (una forma que no ocurre en el habla, sino que tiene una "realidad" meramente psicológica y no tanto superficial) y las despoja de todo tipo de afijos—esto es, las lematiza—buscar una palabra se convierte en un desafío analítico. Aunque esto nuevamente refleja las prioridades intelectuales de sus autores, tiene consecuencias posiblemente inconvenientes para que lo usen los propios hablantes de tzotzil.

Otra variante de la lista de palabras etnolingüística, esta vez proveniente de Australia, ilustra otro aspecto del dilema del lexicógrafo de campo. Muchos lingüistas han documentado las lenguas aborígenes de Australia con el muy escaso número de hablantes que aún queda y que a menudo no las hablan fluidamente. Mi propio trabajo en el ahora extinto

³Está en prensa una versión tzotzil-español (2005), que publicará el Centro de Investigaciones y Estudios Superiores en Antropología Social (CIESAS) de México. A medida que aumentan los hablantes de tzotzil que cruzan la frontera de México rumbo a los Estados Unidos, el número de hablantes bilingües del tzotzil y del inglés no hará más que crecer.

idioma de Barrow Point (véase Haviland 1998) es un ejemplo menor. En estos casos, las listas de palabras reflejan una oportunidad afortunada, más que una planeación sistemática, y la cobertura es dispareja, basada en circunstancias fortuitas y en la suerte. Sin embargo, aun las listas de palabras reunidas azarosamente pueden ser importantes cuando los procesos políticos—por ejemplo, el movimiento de los aborígenes por recuperar sus territorios originales—se basan en pruebas lingüísticas para establecer vínculos entre la tierra y la sociedad y cultura aborígenes (Henderson y Nash 2002). Todo, desde el nombre de un lugar hasta el nombre de una planta, puede tener una importancia insospechada. En estos casos, la "cobertura" no es una cuestión de la "exhaustividad" científica del trabajo, sino una cuestión ideológica con evidente valor político, otro punto al que regresaré brevemente al final del capítulo.

También hay una tradición *pedagógica* en la elaboración de diccionarios, que ha dado origen a los diccionarios más comunes: los que usan los estudiantes para buscar palabras poco conocidas o los turistas para traducir la carta en un restaurante. Aquí la cuestión de la dimensión es reveladora. Por ejemplo, los diccionarios del español de México (mencionemos el de Lara Ramos 1986) están clasificados por tamaño: una versión breve destinada a los niños en edad escolar con varios miles de palabras "básicas", una versión intermedia con un mayor número de palabras y así sucesivamente. Todos celebran el español mexicano, la variedad que más se habla de esta lengua, pero que está relegada a una categoría subsidiaria por la academia de la lengua de la madre patria. El vocabulario elegido y los hechos de uso están tomados de un enorme corpus de material textual de México, desde cartas hasta artículos de periódico y canciones populares. En Chiapas, el gobierno encargó una variedad similar de "diccionarios de bolsillo" para las lenguas indígenas del estado. Éstos, junto con una serie de esbozos gramaticales, tienen el propósito de servir como herramienta pedagógica y trofeo político, prueba de la preocupación del gobierno por los indígenas luego del levantamiento zapatista de 1994. Con un diseño similar, pero surgidos del impulso ideológico opuesto, se encuentran los silabarios escolares, o listas básicas de palabras, concebidos como auxiliares de la alfabetización en las escuelas de las comunidades zapatistas, que se resisten a cualquier tipo de ayuda gubernamental y a los materiales escolares estandarizados.

2. Indeterminación referencial y otros tropiezos del trabajo de campo

¿Qué tipo de criaturas son los "significados" de las palabras que queremos asentar en una base de datos léxicos? Es difícil escapar al peso de los siglos en que Occidente ha filosofado sobre la materia (aunque hay útiles antídotos en aquel ensayo primigenio de J. L. Austin, "The Meaning of a Word", en Austin 1961). Según Frege (1892), es usual partir de la idea de que los hablantes usan las palabras (típicamente, sustantivos) para elegir entidades en el mundo —los "referentes" de las palabras— en razón de su "sentido" o "denotación", independientemente de cualquier instancia de su uso para hacer referencia a o predicar sobre un estado específico de cosas. De acuerdo con este punto de vista, las palabras son una especie de instrucción del hablante hacia el oyente, basada en un entendimiento compartido de los "significados" de las expresiones y que en general se propone lograr una referencia común.

Aun en casos aparentemente sencillos, los enigmas de la referencia como una teoría del significado afloran de inmediato. Supongamos que alguien quiere referirse a mí mientras dicto una conferencia. Consideremos las siguientes expresiones que esa persona puede usar:

1. Expresiones que se refieren al mismo referente.
 - a. Esc hombre (señalándome con una indicación gestual).
 - b. El profesor de lingüística de Oregon.
 - c. El hombre alto de bigote negro que está al frente de la sala.
 - d. El mexicano de bigote negro que está al frente de la sala.

Las "instrucciones" del hablante, suponiendo que logren su objetivo —es decir, que induzcan al interlocutor a elegirme como la persona a quien él quiso referirse— se basan en tipos muy diferentes de relación con los "significados" de las palabras que utiliza. La primera expresión se basa en algún tipo de entendimiento de categorías de aquello a lo que podemos referirnos como "hombre", combinado con dos recursos indexicales directos, el deíctico "ese" y la indicación gestual. En el extremo opuesto, (b) elige un individuo supuestamente identificable a partir de la intersección de series de denotaciones generadas composicionalmente con las palabras constituyentes (tal vez junto con otros presupuestos de existencia y singularidad que se construyen con el artículo definido "el"). La expresión (c) combina dicha estrategia composicional con alguna deixis implícita (en la que se calcula *qué* sala es y *cuál* es el frente) y, paradójicamente, (d) puede alcanzar el mismo objetivo que (c), a pesar del hecho de que, con

todo y que vivo y doy clases en México y probablemente hasta parezca mexicano, *no soy* mexicano a final de cuentas: es decir, la suma de los "significados" de las palabras constituyentes no necesariamente da por resultado una denotación verdadera.

Así, aunque la referencia es nuestro punto de partida en la lingüística de campo, no puede ser nuestro punto de llegada. El famoso ejemplo de Quine de la palabra *gavagai* (Quine 1960) subraya la profunda indeterminación referencial del comportamiento lingüístico: un lingüista imaginario que realiza un trabajo de campo oye la palabra *gavagai* en presencia de unos conejos, pero no puede determinar si la palabra significa "conejo" o "parte del conejo" o "esencia del conejo", etc. Tal vez sea más reveladora la analogía de Zgusta (1971: 25-26) cuando intenta descubrir los significados de las señales de tránsito (en un sistema como el europeo), pero basándose únicamente en las regularidades del comportamiento de los conductores. Zgusta especula que, con el tiempo, uno podría descifrar mediante la observación directa los significados de las luces roja, amarilla y verde en un semáforo, pero sería mucho más difícil adivinar el significado de "una gran H mayúscula en un letrero rectangular (lo que en muchos países significa que hay un hospital cercano)", pues estas señales se colocan en lugares muy diferentes y "difícilmente produce un efecto uniforme y observable en el comportamiento de los conductores".

El siguiente es un ejemplo mucho menos fantástico, tomado de "la vida real" de la lexicografía de campo. En 1770, el capitán James Cook y su tripulación reunieron listas de palabras de la lengua guugu yimithirr, que se hablaba cerca de lo que ahora se llama Cooktown, en el noreste de Australia (por cierto, una de las palabras registradas era *gangurru*, una especie particular de lo que ahora llamamos "canguro"). Si se cotejan las entradas léxicas registradas por diferentes observadores de esa expedición, puede verse precisamente cómo la indeterminación referencial de la variedad *gavagai* asolaba a estos primeros lexicógrafos. Por ejemplo, bajo la glosa "rama (con brotes o tallo)", el ilustrador del barco, Parkinson, pone *maiye*, mientras que el botánico Banks escribe *mave butai* (y añade la anotación "con hojas") o *mayi bambier*. A partir del idioma moderno, puedo suponer que estas expresiones se basan en la palabra *mayi* "planta comestible" —así que no se trata de cualquier rama— y, más específicamente, *mayi bambier* "el fruto (comestible) de la especie de mangle llamado *bambier*". El otro "nombre" que Banks registra es simplemente la expresión *mayi buday*, que en realidad es una oración completa que significa "la parte comestible fue

comida" o "alguien se comió el fruto".⁴ Una entrada en el diario de Cook muestra que era dolorosamente consciente de los problemas señalados por Quine para extraer la información léxica:

... la lista de palabras que he dado no pudo obtenerse sino preguntándoles por medio de señas qué significaba tal cosa en su idioma, un método propenso a muchos errores. Por ejemplo un hombre sostiene una piedra en la mano y pregunta su nombre: el indio puede darle por respuesta el verdadero nombre de la piedra, alguna de sus características, como su dureza, aspereza, suavidad, etc., alguno de sus usos o el nombre peculiar de alguna especie de piedra en particular, que el interrogador se apresura a anotar como si ese nombre quisiera decir "piedra" (Diario de Cook, véase Cook 1955).

Es claro que parte del problema reside en un modelo primitivo tanto de referencia como de demostración ostensible: lo que puedes elegir señalándolo con el dedo o lo que puedes mostrar "al indio".

Un modelo muy diferente de "ejemplificación" es el que defiende J. L. Austin en "A Plea for Excuses" (en Austin 1961). Al enfrentarse a un par de expresiones aparentemente similares (en el caso de Austin son las famosas *by mistake* "por error" vs. *by accident* "por accidente"), se dilucida la diferencia en sus significados construyendo un cuidadoso ejemplo de cuándo se usaría la primera expresión y no la segunda, y viceversa. Con semejante método, uno no señala cosas, sino contextos de uso. Los propios contextos pueden ser claves para acceder al conocimiento léxico. Por ejemplo, al tratar de recuperar palabras del idioma de Barrow Point con el ya fallecido Roger Hart, ambos trabajamos mucho a partir del guugu yimithirr, que para ambos era una segunda lengua (véase Haviland 1998). Con frecuencia buscábamos—a veces de manera bastante ingenua—el equivalente en el idioma de Barrow Point de alguna palabra en guugu yimithirr. Sin embargo, aun cuando buscábamos los nombres de especies de plantas o animales, a menudo nos metíamos en embrollos, en parte porque la flora y la fauna de Barrow Point eran muy diferentes de las de Cape Bedford, que está más de cien kilómetros al sur, pero en parte también porque en general el ambiente era sencillamente el equivocado. Roger había hablado su idioma tribal antes de que lo apartaran de su familia, lo cual

⁴Véase Haviland (1974). Las observaciones de Nick Evans (2002) sobre los malentendidos de las expresiones aborígenes, inclusive en inglés, en las audiencias ante el Australian Land Tribunal (Tribunal Australiano de Tierras) muestra hasta qué punto estos malentendidos pueden tener graves consecuencias legales.

ocurrió cuando tenía unos seis años de edad. No obstante, la primera vez que lo oí hablar el idioma sin ningún titubeo fue sesenta años después. Luego de recorrer por tierra un largo camino de regreso, fuimos a dar a la playa donde había nacido. La tierra que no había visto durante sesenta años, sus árboles, piedras y animales, parecían hablarle en su idioma natal, y sólo ahí pudo responder fluidamente.

A pesar de ser el elemento principal de casi toda la semántica formal moderna, la referencia—o más precisamente, aquellos aspectos de las expresiones lingüísticas que las vuelven útiles para lograr la referencia—es, desde luego, una base inadecuada para entender el significado en un sentido ordinario. La idea tradicional de "connotación", por ejemplo, se basa en la intuición de que hay palabras diferentes que en algún sentido pueden "referirse a la misma cosa", sin por ello "tener el mismo significado". Esto no es lo mismo que la distinción clásica de Frege entre sentido (lo que significa una expresión) y referencia (justamente a lo que en ese caso se refiere, como una función de lo que significa), en la cual dos expresiones diferentes, con sentidos diferentes, pueden en dado caso referirse al mismo individuo. El ejemplo bastante peculiar de Zgusta es la triada lingüística *decease, die, peg out*, "fallecer, morir, estirar la pata" (en mi propio dialecto del inglés la última sería más bien *check out* o tal vez *go belly up*, que en el español mexicano sería algo así como "colgar los tenis" o "petatearse"). Zgusta (1971: 39-40) cita el armenio como un idioma que tiene contrapartes exactas (*vačxanvel, mernel, satkel*) de estas palabras en inglés, y el chino, con una serie bastante más elaborada que abarca el mismo territorio referencial. (Claro que podríamos añadir más expresiones en inglés, con lo que cambiarían las dimensiones evocadas de "connotación": *pass [away], go [to a better place] o [to meet his/ her maker], croak*, etc.; y lo mismo en español: "pasar a mejor vida", "dejar este mundo", "quebrarse", etc.) La forma de aprehender la diferencia entre todos estos términos no es a partir de las condiciones de verdad del estado de cosas que describen (consideradas idénticas), sino por medio de las condiciones apropiadas⁵ de las circunstancias indexicales de su uso: quién emplea cuál expresión, ante quién, hablando sobre qué tipo de entidades muertas y en qué tipo de situaciones, entre otras cosas.

Zgusta equipara el problema del lexicógrafo ante la connotación con otros relacionados con rangos de significado, restricciones de selección y especificidad colocacional. Uno de los ejemplos de Quine era *ackled* "huero", que se dice sólo de los huevos y el cerebro (véase McIntosh 1961).

⁵Véase la noción de "reglas de uso" (*rules of use*) en Silverstein (1976).

Zgusta cita a Černý para dos palabras georgianas que significan "tener": *makvs* (usado para cosas) vs. *mqas* (usado para personas y animales), "pero los automóviles no son considerados cosas, sino animales, porque uno dice *mankana mqas*, "tengo un automóvil" (Zgusta 1971: 44)." El estudio de Berlin (1967) sobre el tseltal¹ ejemplifica un fenómeno paralelo, pues en este idioma existen seis verbos que significan "comer" y cada uno se aplica a distintos tipos de comida. Aquí nos encontramos con enredos conceptuales que no podemos esclarecer en este capítulo por falta de espacio: si bien las palabras con diferentes connotaciones parecen apropiadas en diferentes contextos de uso o para diferentes actitudes del hablante, ¿podemos distinguir entre restricciones de selección y limitaciones de denotación? Tal vez *makvs* denota un estado de cosas diferente de *mqas*, no sólo "el mismo concepto" aplicado a diferentes tipos de objeto. Tal vez el *we'* del tseltal "comer (tortillas, por ejemplo)" es realmente un acto diferente de *k'ux* "comer (cosas crujientes, por ejemplo)". Un ejemplo más cercano puede ser el verbo en inglés *to be*, que en español equivale, según el caso, a "ser" o "estar", dos actos muy diferentes, no intercambiables y aplicables en circunstancias distintas. Cualquiera que sea nuestra teoría semiótica, estas distinciones sistemáticas de significado evidentemente pertenecen al lexicón documental, de modo que registrarlas es parte del "deber" del lexicógrafo y una tarea a la que debe dirigirse atención metodológica.

Aquí, el problema de la evidencia negativa (o, más bien, la falta de ella en el habla tal y como ocurre naturalmente) es de importancia crucial en la compilación de una base de datos léxicos para una lengua amenazada. Las pruebas sobre los límites en el rango de significado de una palabra o frase, o sobre las restricciones de su uso o sus condiciones de felicidad en diferentes contextos culturales e intertextuales pueden ser simplemente inexistentes en un corpus textual y la extracción sistemática de intuiciones lexicográficas específicas puede ser imposible. En el diccionario de tzotzil de la Colonia aparece la glosa "pesar el negocio con cordura o diligencia" (*treat a matter prudently or diligently* en Laughlin 1988), para la cual los frailes dan una versión conjugada de la expresión tzotzil: *-a'i ta -olonton*, que es literalmente "oír (o sentir, o entender) con el corazón". La frase en tzotzil precisa ser completada morfosintácticamente, porque el verbo transitivo *-a'i* requiere un sujeto sintáctico (quien supuestamente "pesa" el negocio) y un objeto (el "negocio" pesado). Además, la palabra *olonton*,

¹De acuerdo con Jost Gippert, "los hablantes nativos georgianos confirman que *mqas* se aplica a cualquier cosa móvil, sean autos, bicicletas, aviones o similares"

²En las obras de Berlin se usa la vieja ortografía "tseltal"

"corazón", también requiere obligadamente un poseedor, que a juzgar por el idioma moderno debe ser correferencial al sujeto del verbo, de modo que "x oye con su *propio* corazón", no con el de alguien más. Estas restricciones morfosintácticas no son evidentes a partir del uso original. Tampoco es claro que la expresión se limite al tipo de contexto referencial sugerido por las paráfrasis en inglés o en español; al contrario, sencillamente parece sugerir una consideración cuidadosa de algo, ya sea el "negocio" o algo menos específico o concreto. Si no se tiene acceso a hablantes nativos con plena fluidez, es imposible proporcionar más detalles léxicos. Más problemático—y quizás más relevante para la documentación de una lengua amenazada—es el caso de una palabra arcaica o una que se usa poco en la comunidad de habla. Una vez más, el tzotzil colonial ofrece un ejemplo instructivo. El lenguaje ritual del tzotzil moderno usa la expresión *tza-uk*, que obviamente se forma con una raíz nominal (hipotética) *tza* más un sufijo subjuntivo o de irrealis *-uk*. Laughlin (1975) sugiere como un significado para *tzauk* "take heed" (prestar o poner atención, hacer caso), una traducción sugerida por hablantes modernos bien informados. Sin embargo, de manera al parecer arbitraria, en el diccionario moderno pone la palabra bajo la raíz *tzak* "atrapar, coger". Sólo el descubrimiento del diccionario colonial (Laughlin 1988) reveló la raíz arcaica *tza*, que había caído completamente en desuso en el tzotzil zinacanteco, salvo por el uso ritual sobreviviente. Los lexicógrafos de la Colonia la registraron con los significados de "habilidad, conocimiento, destreza, intuición, laboriosidad, inteligencia, opinión, prudencia, pericia, especulación, talento, pensamiento", pero en el uso moderno no se encuentran muestras de esto.

Tal vez el cuento más viejo de la lingüística antropológica sea la diversidad de denotaciones en las cartografías léxicas de la "realidad", captada en el refrán de que "las palabras diferentes" implican "mundos diferentes". Un dominio clásico es la etnoanatomía, la división léxica (¿y, por ende, quizás conceptual?) que separa el cuerpo en unidades discretas. Los hablantes de inglés o español distinguen "manos" de "brazos", pero no los hablantes de ruso o de tzotzil. El tzotzil tiene la raíz única *k'Ab*,² que puede querer decir "mano" o "brazo". Es más, también puede significar "rama", "manga", "travesaño (de una cruz)", "pata delantera (de un gato)", etc. En tzotzil, *ni'* "nariz", denota no sólo narices, sino cualquier protuberancia relativamente puntiaguda o el extremo delgado de casi cualquier tipo de objeto, no solamente de un rostro o una cabeza. Así que, ¿por qué privilegiar la acepción de "parte del cuerpo", como "mano" o

²El símbolo *A* denota una vocal hipotética que alterna entre *a* y *o* en raíces derivadas.

"nariz"? Tal vez estas partinomias involucren algún modelo que no es anatómico. Otra posibilidad es que un "significado básico" se extienda de varias maneras hacia una cadena o un continuum de significados derivados carente de puntos terminales bien definidos. Cruse (1986) argumenta que términos como "boca" en inglés participan en un "espectro de sentidos" (*sense spectrum*) en el que cada significado "derivado" o "metafórico" lleva a otro.

2. "Espectro de sentidos" (Cruse 1986: 71 y ss.)

John keeps opening and shutting his *mouth* like a fish.

[John abre y cierra la *boca* como un pez.]

This parasite attaches itself to the *mouths* of fishes, sea-squirts, etc.

[Este parásito se adhiere a la *boca* de los peces, ascidias, etc.]

The *mouth* of the sea-squirt resembles that of a bottle.

[La *boca* de las ascidias se parece a la de las botellas.]

The *mouth* of the enormous cave was also that of the underground river.

[La *boca* de la enorme cueva también era la del río subterráneo.]

Los tipos de elementos significativos que uno escoge para una base de datos léxicos también se encuentran inextricablemente ligados a la totalidad del análisis categorial que uno hace de la lengua, qué "partes del habla" se postulan y qué tipo de perfiles semánticos se asocian con ellas. El clásico punto de partida semántico formal, según el cual los sustantivos corresponden a las cosas (es decir, a conjuntos), los adjetivos a las "propiedades" (es decir, a subconjuntos) y los verbos a acontecimientos o estados de cosas (predicados sobre múltiples entidades), se desintegra rápidamente ante los diversos tipos de *fusión (conflation)* semántica (Talmy 1985) que se observan constantemente en los artículos léxicos. Un ejemplo común es el inglés *climb* "escalar, ascender, trepar", cuya definición en el sitio del proyecto FrameNet⁹ es *to move vertically usually upwards, usually with effort* ("moverse verticalmente, por lo general hacia arriba y con esfuerzo"). Esto es, el verbo sugiere, por default, un movimiento vertical ascendente acompañado por el tipo de esfuerzo que Fillmore llama *clambering* "trepar a gatas, encaramarse, gatear". Cualquiera de estos elementos fusionados—movimiento ascendente o esfuerzo—puede suspenderse sin causar una anomalía semántica, pero no ambos:

⁹Véase <http://framenet.icsi.berkeley.edu/index.php> y el apartado 3 páginas adelante.

3. Fusión en *climb* (Fillmore 1982)

The snake climbed (up) the tree. [movimiento ascendente + esfuerzo]

The monkey climbed (up/down) the tree. [movimiento ascendente + esfuerzo/ sólo esfuerzo]

¿?The snake climbed down the tree. [ninguno de los dos]

Otro presupuesto común de la lingüística antropológica es que los idiomas combinan dominios semánticos en formas inesperadas, quizás de manera más característica en los verbos. Por ejemplo, los siguientes predicados posicionales del tzotzil podrían todos recibir la misma glosa, ya sea *stuck* en inglés o "atorado" en español.

4. Palabras en tzotzil para "atorado"

<i>Kakal</i>	"atorado (entre dos superficies)"
<i>Ch'ikil</i>	"atorado (en una grieta estrecha o apretada)"
<i>Katz'al</i>	"atorado (en un orificio en forma de mandíbula)"
<i>Xojol</i>	"atorado (en un agujero cerrado)"
<i>Tz'apal</i>	"atorado (un objeto puntiagudo anclado en una superficie)"

Sin embargo, como muestran las acepciones detalladas, cada palabra especifica diferentes configuraciones, distintas maneras de estar atorado y diferentes formas tanto para el objeto atorado como para el suelo o lugar donde está atorado.¹⁰ La fusión exacta implica, me parece, factores como los siguientes, si tomamos la raíz *tz'ap* como ejemplo:

5. Fusión en *tz'ap*

- a. el "extremo" del objeto está "dentro" del suelo;

¹⁰El inglés tiene diferencias interesantes en sus elaboraciones, como puede verse en las entradas que arroja FrameNet para *being attached* ("estar unido"): *affixed, anchored, attached, bolted, bound, chained, fastened, fused, glued, handcuffed, lashed, manacled, moored, nailed, pasted, pinned, plastered, riveted, sewn, shackled, stapled, stuck, taped, tethered, tied, welded* (fijado, anclado, adherido o anejado, acerrojado, atado, encadenado, asegurado, fusionado, pegado, esposado, atado con cuerdas, maniatado, atado con cables, clavado, engomado, sujeto con alfileres, emplastado o revoado, remachado, cosido, sujeto con grilletes, engrapado, atorado, pegado con cinta adhesiva, amarrado con una correa como animal, amarrado, soldado). En inglés, la variable central parece ser el tipo de material que genera la unión.

- b. el suelo no necesita tener un *y-ni* "interior" (o quizás no necesita tener esa estructura, quizás se conciba más bien como mera superficie);
- c. el objeto tiene un "extremo" "puntiagudo" (en tzotzil, *s-ni* "nariz");
- d. normalmente, el objeto está "atorado" en el Suelo con el extremo puntiagudo, es decir, está sujeto y se sostiene solo, y
- e. normalmente está orientado verticalmente.

Los lingüistas han planteado diversas clasificaciones de tipos semánticos, en diferentes clases de raíces, y el lexicógrafo de campo debería aprovechar sin pruritos estas tipologías, desde marcos hasta tipos de verbos (Dixon 1972) o clases de verbos basados en patrones de diátesis (Levin 1993).

La multiplicidad de los "juegos del lenguaje"—algo que ya no puede quedar oculto para un lingüista de campo profesional—complica aún más el tradicional enfoque referencial del significado léxico. Usamos las palabras para referir; pero también para muchas otras cosas. Consideremos parte de las que enlista Wittgenstein:

Dar órdenes y obedecerlas - Describir la apariencia de un objeto o dar sus medidas - Construir un objeto a partir de una descripción (un dibujo) - Informar sobre un acontecimiento—Especular sobre un acontecimiento Formar y probar una hipótesis—Presentar los resultados de un experimento en tablas y diagramas—Hacer un cuento y leerlo - Representar-actuar—Cantar cantinelas Resolver acertijos—Hacer un chiste y contarlos—Resolver un problema de aritmética práctica - Traducir de un idioma a otro- Preguntar, agradecer, maldecir, saludar, rezar (Wittgenstein 1958: sección 23).

Cruse (1980: 270 y ss.) nos recuerda las diferencias entre sus "modos semánticos" (*semantic modes*), como en el contraste entre las dos siguientes enunciaciones:

6. "Modos semánticos"

De repente sentí un dolor agudo.
¡Ay!

Si la semántica fuera sólo referencia y predicación, sería difícil aprehender el significado de "¡Ay!" semánticamente, porque la palabra no implica ni referencia ni predicación. En cambio, sería importante entender fenómenos

como las interjecciones (véase Kockelman 2003) en términos de modalidades semióticas muy diferentes: como indicios de la posición del hablante, la relación del interlocutor con el hablante, los supuestos estados corporales y afectivos, las respuestas esperadas y demás. El hecho de que palabras como "ay" sean difíciles de modelar en términos de lo denotado no libera al lexicógrafo de la responsabilidad de registrarlas y explicar cómo funcionan, problema al que regresaré más adelante.

Una concepción más amplia y apropiada del significado proviene de una de las famosas tricotomías de los modos en que los signos pueden significar o "representar" otras cosas, debida a C. S. Peirce (1932). Los tres modos semióticos se basan en principios muy diferentes, aunque generalmente se entremezclan en la mayoría de los signos, sean lingüísticos o de otro tipo. Peirce señala que algunos signos representan ciertas cosas debido a un parecido entre el vehículo o forma física del signo y la cosa significada (de esta manera, la fotografía de una persona puede representar a esa persona, por ejemplo, en un directorio o un catálogo). El signo comporta un parecido "icónico" con lo que significa, aunque el carácter del "parecido" puede variar enormemente (consideremos los diagramas, los dibujos, las siluetas o las gráficas, por ejemplo, o palabras ya estandarizadas, pero que no dejan de ser onomatopeyas cuyos sonidos sugieren sus significados: "mugir", "maullar" o "piar"). También puede haber una relación "indexical" entre el signo y lo significado, tales como las relaciones físicas, espaciales o de causa directa entre el vehículo del signo y lo que significa. Por ejemplo, una buella quizás no se "parezca" a la persona que la hizo (aunque naturalmente puede "parecerse" a su pie), pero se presenta como un "indicio" de la persona en virtud del hecho de que fue preciso que el pie de la persona dejara la marea (con lo que se indica, por ejemplo, que esa persona estuvo en cierto lugar). En la lengua, "¡ay!" representa (en realidad, exhibe) un dolor súbito precisamente porque imaginamos que el dolor mismo de alguna manera (¿involuntariamente?) provoca la enunciación. Es parecida la forma en que sabemos a qué persona se refiere "yo" o "tú" si observamos la relación contextual entre el signo—la palabra—y la persona que lo enuncia o a quien se lo enuncia. Esas palabras, pues, se apoyan en una relación indexical (en un contexto) para transmitir sus significados. Finalmente, hay signos cuyo significado carece esencialmente de motivaciones por semejanza o contexto: éstos son los "símbolos" peirceanos, que se basan en una relación convencional entre significante y significado, en la "arbitrariedad del signo lingüístico" de Saussure, según la cual "gato" sólo significa gato porque eso es lo que una tradición lingüística en particular ha normado.

La Figura 1 presenta un signo que combina claramente las tres modalidades semióticas de Peirce: la semejanza icónica entre el dibujo y un (estilizado) cigarrillo humeante; el significado convencional (al menos en gran parte del mundo occidental) del círculo sombreado tachado con una diagonal como una "prohibición" y, finalmente, la ubicación del signo mismo, cuya posición física señala indexicalmente con exactitud *en dónde* está prohibido fumar.

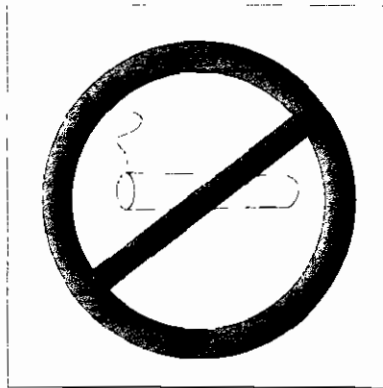


Figura 1. Un signo semióticamente tricotómico

Una descripción adecuada del significado de elementos lingüísticos debe capturar los tres modos de significación, aunque las principales tradiciones lexicográficas se limitan en buena medida al significado "convencional" o simbólico, casi exclusivamente en términos referenciales.

3. Los metalenguajes para significados y unidades de conocimiento léxico

Un segundo conjunto muy importante de cuestiones para las bases de datos léxicas es cómo representar los significados de los objetos léxicos y cómo delimitar esos objetos por principio de cuentas. Las equivalencias bilingües de las definiciones a menudo son evidentemente inadecuadas, por las razones que siempre han preocupado a los traductores: el desajuste en la clase gramatical, la inexactitud o falta de equivalencia entre los términos del idioma de origen y el de destino, los rangos de significado incompatibles,

los círculos viciosos o las regresiones infinitas, etc. Mucho depende de los metalenguajes disponibles. Mi colega Matt Pearson, con el propósito de ilustrar la interdependencia de las diferentes modalidades expresivas del lenguaje, confronta a los estudiantes que empiezan a estudiar lingüística con la siguiente pregunta: "¿pueden definir 'espiral' sin usar las manos?" (usted mismo puede intentarlo antes de seguir leyendo).

Repito: todo depende de los metalenguajes disponibles. Hasta un principiante en matemáticas puede responder presentando una fórmula para una gráfica tridimensional, es decir, definiendo una serie de valores para los ejes (x, y, z). Aquí hay algunas fórmulas ilustrativas:

7. Espiral

($\coseno(t)$, $\seno(t)$, t)

[para una espiral en forma de resorte]

($c*t*\coseno(t)$, $c*t*\seno(t)$, $c*t$) (donde c es una constante, y $*$ es la multiplicación) [para una espiral en forma de cono]

Con el único fin de ver cómo funcionan estas fórmulas, las Figuras 2 y 3 presentan dos gráficas de los resultados, trazados por mi colega Albyn Jones, especialista en estadística.

La belleza del metalenguaje matemático empleado es su precisión, parsimonia y supuesta universalidad.¹¹ La desventaja son los arcanos de su potencial incomprensión.¹² Además, aunque las fórmulas pueden describir con bastante precisión una clase de formas geométricas y quizás hasta ayudarían a definir "espiral", es posible que nos sigamos enfrentando a la necesidad de echar mano de otros recursos no matemáticos (que tal vez sean igualmente generales) para aprehender el significado de la palabra en expresiones como "los precios suben en una espiral que se sale de control", o bien "tenemos que controlar la enloquecida espiral de la proliferación nuclear".

¹¹ Hay propuestas desde la lingüística misma de un "metalenguaje semántico natural" con el cual las definiciones de nociones complejas podrían enmarcarse en términos de primitivos semánticos más simples y "universales" (por ende "naturales"). En <http://www.unc.edu/arts/LCL/disciplines/linguistics/nmpage.htm> aparece una bibliografía de las muchas publicaciones de Anna Wierzbicka.

¹² Ante el reto de Pearson, Chris Hauk, un estudiante de últimos años del Reed College, no tardó en salir con "ah, a lo que usted se refiere es a un hilo enrollado alrededor de un cilindro, en comparación con un hilo enrollado alrededor de un cono" (Albyn Jones, comunicación personal, 1º de marzo de 2005), lo que demuestra que los matemáticos también pueden ser lexicógrafos.

Una dificultad al suponer que existe un metalenguaje semántico independiente de la lengua (además de imponer un prejuicio ante la semiosis de las palabras y limitarla a información referencial, preocupación de la sección anterior) es que puede violentar la organización conceptual de lenguas particulares. Aquí nos encontramos con la dicotomía émica-ética de la lingüística antropológica clásica: ¿a qué damos prioridad, a la organización de formas y significados específica de la lengua o a las categorías descriptivas derivadas de concepciones externas? Una de las primeras demostraciones del dilema es la forma en que Conklin trabajó con los pronombres del hanunoo.

8. Los pronombres del hanunoo (Conklin 1962)

<i>kuh</i>	"yo" 1s
<i>muh</i>	"tú" 2s
<i>yah</i>	"él/ella" 3s
<i>tah</i>	"nosotros dos" 1du
<i>tam</i>	"todos nosotros" 1pl INCL
<i>yuh</i>	"todos ustedes" 2pl
<i>dah</i>	"ellos" 3pl
<i>mih</i>	"nosotros (pero tú no)" 1plEXCL

Si adoptamos el metalenguaje pronominal estándar, *kuh* sería parafraseado como "primera persona del singular" o *tam* como "primera persona del plural inclusivo". El metalenguaje implica, pues, un componente de persona (con los valores posibles 1, 2 o 3), un componente de "número" (con los valores posibles para el hanunoo del singular, dual o plural) y un componente de "inclusividad" (con los posibles valores "inclusivo" y "exclusivo", y quizás un valor no marcado), que es defectivo en el sentido de que por definición sólo puede aplicarse a pronombres de la primera persona que no son del singular. A partir de estos componentes de significado, debiera ser posible distinguir entre 11 y 13 distintas formas pronominales (tres personas diferentes, con tres números diferentes y una distinción inclusiva/exclusiva en todas las formas de la primera persona que no son del singular). Sin embargo, el paradigma sólo tiene ocho pronombres. Peor aún, los términos básicos del metalenguaje descriptivo (las categorías de número y persona, más los términos "inclusivo" y "exclusivo") dan por sí solos un total de ocho, lo que sugiere que no hay mucha razón para recomendar este metalenguaje en particular en vez de usar los propios términos del hanunoo como elementos "primitivos" o "indefinibles".

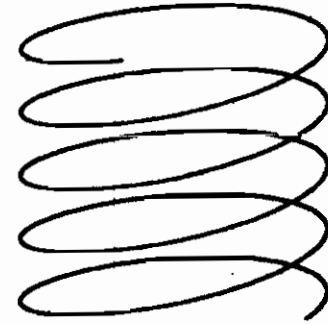


Figura 2. (coseno(6t), seno(6t),t) para t en (0, pi)

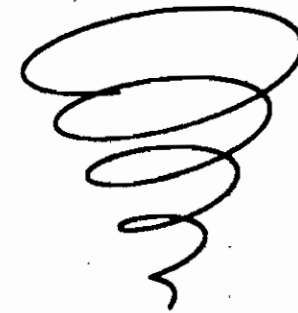


Figura 3. (t coseno(t),t seno(t),t) para t en la misma ubicación

Conklin observa que es posible hacer un mejor análisis si se toma como medida de evaluación la eficiencia (de modo que tres distinciones binarias deberían permitir distinguir $2^3 = 8$ términos) y la "lealtad" a la lógica hanunoo original. Los tres rasgos binarios que propone son \pm Hablante (H), \pm Oyente (O) y \pm Mínimo (M), lo que da lugar a la Tabla 1, cuya simetría estética infunde la esperanza de que uno está descubriendo y no imponiendo el sistema sustentador

Tabla 1. Los pronombres del hanunoo

	H	O	M
<i>kuh</i> "yo" 1s	+	-	+
<i>muh</i> "tú" 2s	-	+	+
<i>yah</i> "él/ella" 3s	-	-	+
<i>tah</i> "nosotros dos" 1du	+	+	+
<i>tam</i> "todos nosotros" 1pl INCL	+	+	-
<i>yuh</i> "todos ustedes" 2pl	-	+	-
<i>clah</i> "ellos" 3pl	-	-	-
<i>mih</i> "nosotros (pero tú no)" 1plEXCL	+	-	-

Otro paradigma descriptivo de utilidad que se aplica ampliamente en la práctica lexicográfica (y que de hecho es impulsado por ella) es la "semántica de marcos" (*frame-semantics*), asociada con Charles Fillmore (véase, por ejemplo, Fillmore y Atkins 1992). Según este enfoque, las palabras individuales proyectan "marcos" más amplios y estructurados, configuraciones de elementos y acciones, algunos de los cuales reciben realización gramatical explícita, mientras que otros quedan implícitos en el marco. Así, las familias de palabras comparten marcos. Por ejemplo, la descripción en FrameNet del marco *commerce-buy* ("comercio-comprar")—que puede realizarse con verbos como *buy*, *lease* o *rent* (comprar, alquilar o rentar)—es la siguiente:

Palabras que describen una transacción comercial básica en la que hay un comprador y un vendedor que intercambian dinero y mercancías, desde la perspectiva del comprador. Las palabras varían individualmente en los patrones de realización de los elementos del marco que permiten. Por ejemplo, el patrón típico para el verbo *BUY* "comprar": *BUYER buys GOODS from SELLER for MONEY*. [COMPRADOR compra MERCANCÍAS a VENDEDOR por DINERO.] *Abby bought a car from Robin for \$5,000*. [Abby le compró un auto a Robin por \$5 000.]

Claro está que los mismos marcos pueden estar relacionados entre sí. Compárese la descripción para el marco de *giving* ("dar, donar"), que "hereda" el marco *commerce* anterior:

Un Donante transfiere un Tema de un Donante a un Receptor.¹¹ Este marco incluye solamente acciones que son iniciadas por el Donante (el que al principio es el dueño del Tema). Las oraciones (aun las metafóricas) deben cumplir con las siguientes implicaciones: el Donante es el primero en tener posesión del Tema. Una vez hecha la transferencia, el Donante ya no tiene el Tema; ahora lo tiene el Receptor.

Existe un enfoque, desarrollado más explícitamente en Levin (1993), que utiliza diversos diagnósticos sintácticos—como patrones de diátesis—para dividir conjuntos léxicos en familias o clases y que guarda en cierta forma relación con lo anterior como un dispositivo metasemántico. Al usar como diagnóstico la ocurrencia de ciertos comportamientos sintácticos en verbos específicos, se dividen los verbos en clases que, de acuerdo con esta lógica, pueden exhibir aspectos de significado comunes. Por ejemplo, Levin propone las siguientes construcciones como pruebas relevantes para descubrir clases semánticas entre verbos transitivos.

9. Diagnóstico de diátesis

MEDIO: The bread cuts easily. (El pan se corta fácilmente.)

CONATIVO: Carla hit at the door. (Carla trató de tocar la puerta.)

ASCENCIÓN DE POSEEDOR DE PARTE CORPORAL: Terry touched Bill on the shoulder. (Terry tocó a Bill en el hombro.)

Si se aplican a verbos específicos (cada uno de los cuales puede tener una variedad de hipónimos, y formar así familias de significado), dichas pruebas revelan diferentes clases sintácticas correspondientes a supuestas familias de significado. Estas familias de significado pueden, a su vez, ser usadas para agrupar objetos léxicos individuales y los grupos se justifican así no sólo en términos conceptuales, sino también sintácticos.

¹¹Obsérvese que *Donor* "Donante" es aquí una entidad individual, definida en FrameNet como "the person that begins in possession of the Theme and causes it to be in the possession of the Recipient" (la persona inicialmente en posesión del Tema y que provoca que pase a estar en posesión del Receptor).

10. Diagnóstico de diátesis aplicado a diferentes verbos (basado en Levin 1993: 6)

	touch	hit	cut	break
CONATIVO:	No	Sí	Sí	No
ASC. DE POSEEDOR DE PARTE CORPORAL:				
	Sí	Sí	Sí	No
MEDIO:	No	No	Sí	Sí

4. La extracción sistemática de bases de datos léxicas

Después de que se han documentado las estructuras básicas de una gramática y reunido un amplio corpus de textos, ¿cómo se suplementan los ejemplos elicitados y las muestras de uso señaladas en los textos para lograr una compilación sistemática del conocimiento léxico? La glosa interlineal de un corpus grande puede usarse mecánicamente para generar una lista estructurada de palabras, cuya perspicacia analítica está en proporción directa con el cuidado y la consistencia del compilador en el etiquetado morfológico y semántico durante el procedimiento de glosa. Diversas herramientas computacionales contribuyen a la extracción léxica de los corpora de textos, no solamente herramientas diseñadas especialmente para bases de datos lingüísticas, como Shoebox/Toolbox del Instituto Lingüístico de Verano, sino también herramientas tanto generales como especializadas de concordancia (escritas, por ejemplo, como *shell scripts* de unix o con lenguajes de programación como PERL o ICON).¹⁴

Otras técnicas computacionales también pueden ayudar a elicitar lexemas de un idioma, aprovechando los patrones fonológicos regulares. Un ejemplo bien conocido es el método de Terry Kaufman para generar listas exhaustivas de "raíces potenciales" en lenguas mayenses a partir del canon etimológico maya *CVC* o alguna variante sencilla derivada de ese canon. La Tabla 2 muestra un pequeño programa de Icon que comienza con todas las consonantes y las vocales del idioma mayense tseltal y produce una lista completa de todas las permutaciones de la forma *CV(:)(j)C*.¹⁵ El programa arroja 8 820 raíces potenciales (en la Tabla 3 aparecen las primeras de las que comienzan con *b*). Cada una de ellas puede cotejarse exhaustivamente (y extenuantemente) con los hablantes nativos para ver qué formas

¹⁴Visite <http://www.cs.arizona.edu/icon/>.

¹⁵El programa simboliza las consonantes glotalizadas o ejectives y las vocales largas con letras mayúsculas, y se usa una # para señalar la ausencia de la *j* medial.

realmente arrojan ítems léxicos reconocibles (muchos hablantes de lenguas mayenses y otras con fonotácticas igualmente directas han sido sometidos en el transcurso de los años a esta abrumadora tarea mental).

Tabla 2. Ensalada de raíces tseltales, en el lenguaje programador de Icon

```

procedure ma:ra:
  l := "bcDkFlmnpBjJ BwvYz"
  / := "aA-BiIoOuU"
  M := "0j"
  every (l) do {
    every (v) := 'V' do {
      every (ml) := 'M' do {
        every (c) := 'C' do {
          root := clivl (ml, c)
          write root
        }
      }
    }
  }
end

```

Tabla 3. Las primeras raíces tseltales posibles que comienzan con *b*

```

ba' bab bach bach' baj bak bak' bal bam ban bap bap' bat bas bar bat' baw
bax bay bats bats' baj bajb bajc baj,h' bajj bajk bajk' bajl bajm bajn
bajp bajp' bajr bas bajs bajt bajt' bajw bajx bajy bajz bajz' baa' baab
baach ... etc.

```

Las listas de palabras generadas de manera mecánica inevitablemente revelarán áreas que requieren mayor trabajo lexicográfico—por ejemplo, frases lexicalizadas, paradigmas definidos sintagnáticamente, elementos "funcionales" vs. "léxicos" o partículas—y generalmente también hacen evidentes ciertos dominios léxicos especialmente complejos que vale la pena explorar más a fondo. Por otra parte, estos dominios pueden no emerger de lagunas obvias o alguna hipertrofia en los conjuntos léxicos generados a partir de las colecciones de textos o listas de palabras elicitadas, sino que pueden aparecer cifrados en las prácticas comunicativas de una comunidad de habla: por un lado, están los juicios estéticos sobre el discurso "bello" o "elocuente"—o bien "feo" o "torpe"—los tipos de habla especialmente marcados y evaluados o bien géneros o ejecuciones verbales especializados, entre otros, y, por otro lado, las "preocupaciones" culturales por la expresión léxica asociada, que se manifiestan en los vocabularios especializados para diferentes profesiones, actividades u otro tipo de intereses, o bien la insistencia en "encontrar la palabra adecuada" o en la expresión "correcta" y "precisa".

La mayoría de los métodos de elicitación léxica son, para bien o para mal, "extensionales" y "referenciales", es decir, básicamente consisten en presentar ejemplares de cosas o situaciones en el mundo ante los hablantes nativos y preguntar cuáles son las expresiones lingüísticas

adecuadas que pueden usarse para referirse a ellas o caracterizarlas. Este método tal vez sea incluídible para la documentación léxica en un primer nivel, pero deja sin respuesta preguntas muy difíciles sobre la intensión de las palabras: lo que en realidad significan, qué distinciones de significado codifican, qué tipo de relaciones de significado establecen con otras palabras y expresiones, y no solamente a qué estado de cosas pueden referirse verdaderamente cuando se emplean. Estas técnicas de elicitación a menudo son igualmente inútiles para captar aspectos no referenciales del significado, como los registros de cortesía, los usos y contextos especializados y otros similares. Estas cuestiones pueden—y quizás deben dejarse de lado en las primeras etapas de construcción de bases de datos léxicos en la documentación lingüística, pero no pueden pasarse eternamente por alto.

Presento un solo ejemplo de mi propio trabajo de campo con el guugu yimithirr. Muy pronto aprendí que en esta lengua la palabra coloquial *nambal* significa "piedra", pero que también se extiende para significar "dinero". Mi principal maestro (y padre social) de la comunidad, que en algunas ocasiones recibía de mí dinero prestado, a menudo usaba (o me susurraba) otra palabra cuando quería referirse al dinero: *wambugan*. Sin embargo, *wambugan* es en realidad un equivalente cortés de la palabra común *nambal* en el vocabulario respetuoso, obligatorio al hablar con los parientes políticos y que en la bibliografía publicada se llama "lenguaje del cuñado" (*brother-in-law language*; Haviland 1979). El rango denotativo de *wambugan* es de hecho un tanto más amplio que el de *nambal*: incluye piedras (entre ellas, las piedras de moler, que tienen nombres especiales, el cuarzo y otras que normalmente no se llaman *nambal*) y dinero. Crucialmente es una palabra exageradamente cortés, que ya no se usa en el Hopevale moderno con los parientes políticos y que de hecho tampoco es ampliamente conocida; sólo la usan unos cuantos ancianos, entre quienes sigue comportando un tono eufemístico de respeto. Ambos factores se combinan para hacer de *wambugan* una palabra en código perfecta para un asunto embarazoso, como pedirle un préstamo a tu hijo adoptivo y discípulo.

Por ahora dejaremos de lado estas dificultades para considerar algunas técnicas que suplementan la información léxica azarosamente reunida mediante el revertimiento mecánico de los corpora textuales. Obviamente, el truco es la elicitación sistemática pero controlada, presentando o simulando ante los hablantes nativos aspectos de la realidad "externa", para estimularlos a usar palabras o expresiones que representan estados de cosas que aún no han ocurrido en nuestra presencia. De modo un

tanto artificial, dividí los métodos de muestreo según los aspectos de la "realidad" que pretenden simular: hechos "naturales", instituciones socioculturales y, en las últimas secciones, hechos de (inter)acción y construcción ideológica sobre el idioma y la sociedad.

4.1. "La Naturaleza"

La tradición en la lingüística antropológica, llamada indistintamente "semántica etnográfica" o "etnociencia", pretende desplegar el conocimiento culturalmente específico sobre el mundo natural detallando la semántica de los dominios léxicos relacionados con los fenómenos naturales correspondientes: las plantas medicinales en banunoo, las categorías tseltales para la leña, la etnobotánica o etnozooología, las partes de una casa o del cuerpo, las taxonomías de las enfermedades, la tecnología local, etc. Un ejemplo clásico del género es el exhaustivo estudio de Berlin (1968) sobre los clasificadores numerales del tseltal, un detallado compendio de los varios cientos de clasificadores que alguna vez fueron obligatorios en las expresiones numerales del tseltal de Tenejapa. Los clasificadores numerales especifican las unidades contables de diferentes tipos de sustancias, a menudo basándose en la forma. Para nuestros propósitos, la característica notable del estudio de Berlin es el uso de fotografías cuidadosamente elaboradas y usadas a la vez como estímulos (es decir, para elicitación en los hablantes las expresiones numerales del tseltal) y como vehículo de representación metasemántica, pues acompañan e ilustran la caracterización verbal de las formas tseltales elicitadas. (Berlin también usó el procedimiento mecánico de Kaufman para generar raíces potenciales de clasificadores numerales, como ya se describió páginas atrás). Para dar una idea de la especificidad semántica de las formas tseltales y del carácter de los estímulos fotográficos, presentamos dos fotos del estudio de Berlin (obsérvese que en la Figura 5, que ilustra el clasificador *hiht*, la imagen sugiere que el hablante también mencionó un uso adecuado para *behé* en la misma fotografía de estímulo, un buen ejemplo de las afortunadas consecuencias que puede traer el uso de tales estímulos).



Figura 4. Tseltal *b'ehē'*: "enrollados individuales de objetos flexibles y delgados en una espiral secuencial en torno a algunos objetos largos y rígidos, como un trozo de madera". En la foto aparece *lahunb'ehē'* 'lazo en el estado de diez envolturas secuenciales en torno a un objeto largo y rígido'" (Berlin 1968: 39, lámina I).



Figura 5. Tseltal *hiht'*: "enrollados individuales de objetos largos y delgados en lazadas envolventes alrededor de dos piezas de objetos largos y rígidos, formando ángulos de 90° uno con respecto al otro, como en la construcción de ceras". En la foto aparece: *hoʔhiht'laso'* 'lazo en cinco lazadas envolventes alrededor de dos piezas de objetos largos y rígidos' [del lado izquierdo de la foto se señaló el lazo en estado de *osb'ehē'* 'tres enrollados continuos']" (Berlin 1968: 39, lámina II)

Recientemente se han explorado otros campos semánticos con estructuras cognitivas un tanto más abstractas, también con la ayuda de diversos estímulos artificiales. A partir de la deconstrucción tipológica de

Talmy de los verbos de movimiento (Talmy 1985) y utilizando una variedad de "dispositivos de elicitación", como fotografías, dibujos, videos y caricaturas,¹⁶ los investigadores de campo han explorado en detalle los sistemas lingüísticos de adposiciones espaciales,¹⁷ los direccionales, los verbos de movimiento y otros auxiliares, así como lo que se ha llamado "marcos de referencia" espaciales (Levinson 2003).

El siguiente es un tipo de ejemplo ligeramente diferente. Así como los hablantes de tzotzil usan un conjunto sumamente elaborado de raíces posicionales semánticamente específicas, en la práctica resulta claro que algunas "familias" de verbos agrupados según categorías gruesas de significado nocional (Dixon 1991) incorporan distinciones por lo común desconocidas para los hablantes de otros idiomas y que requieren una delimitación lexicográfica cuidadosa.

Zgusta (1971: 89 y ss.) ofrece una rica discusión de estas familias de verbos, lo que llama "cadenas" de "cuasi sinónimos", y pone como ejemplo las múltiples palabras chinas que equivalen a "cargar". En tzotzil hay muchos verbos transitivos monolexemáticos que en su mayoría bien podrían traducirse al español como "cargar", aunque no parece haber ningún criterio que justifique agruparlas, salvo su traducción a español o inglés. Así, por ejemplo:

- | | |
|-------------|--|
| <i>kuch</i> | "cargar (algo grande) en la espalda, generalmente con ayuda de un mecapal" |
| <i>pet</i> | "cargar o sostener en los brazos frente al cuerpo (un bebé, por ejemplo)" |
| <i>lik</i> | "cargar por medio de una agarradera de la que cuelga lo cargado (como una cubeta)" |
| <i>kach</i> | "cargar apretando entre dos superficies, generalmente entre las mandíbulas (como un perro que lleva un hueso)" |
| <i>jop</i> | "cargar en el cuenco de la mano o en alguna otra superficie cóncava (como en un mandil)" |
| <i>tom</i> | "cargar con la mano, generalmente un objeto largo, empuñado, pero de modo que el objeto se extiende encima o más allá de ella (como una antorcha, un rifle)" |
| <i>mich</i> | "cargar algo apretándolo, generalmente entre los dedos o el puño", etc. |

¹⁶Véanse las descripciones de diversos paquetes de estímulos desarrollados por el "Language and Cognition Group" del Instituto Max Planck de Psicolingüística en http://www.mpl.nl/world_data/fieldmanuals/.

¹⁷Véase Levinson *et al.* (2003) para un enfoque francamente extensional, comparativo.

Por cierto, no conozco ningún verbo en tzotzil tan general como "cargar" que pudiera usarse para abarcar todos estos casos.

Otra familia semejante de verbos en tzotzil es la de "insertar" o "meter" (Haviland 1994), en la que— como en el caso de los verbos para "cargar"— los criterios distintivos implican las formas del objeto insertado y el lugar donde se mete, los tipos de contacto o contención a que se hace referencia, la estrechez con que encaja, las orientaciones del contenedor y el objeto insertado, etc. Tanto para elicitación esta información como para ilustrar estas distinciones, grabé breves películas de diferentes formas de "inserción" realizadas con objetos conocidos, que los hablantes pueden ver y discutir: ¿cuál es la mejor manera de describir lo que ven?, ¿hay otras formas de describirlo?, etcétera.¹²

Es difícil saber de antemano qué áreas del vocabulario tendrán hipertrofia léxica en una lengua que no se ha documentado. La ventaja de las herramientas de elicitación desarrolladas por el Instituto Max Planck de Psicolingüística en Nimega y otros más es que pueden usarse para invitar a los hablantes a explotar plenamente su repertorio de recursos expresivos describiendo estímulos estandarizados. Las caricaturas para niños, como las series *Maus* de la televisión alemana,¹³ son entretenidas y útiles a la vez para investigar los dominios de movimiento, por ejemplo. Naturalmente, es una cuestión problemática que los hablantes de lenguas diferentes, con diferentes tipos de bagaje cultural y experiencias de vida, vean tales estímulos con "el mismo" sentido, y de hecho es un tema central que debe investigarse en el trabajo de campo.

4.2. La realidad sociocultural

De interés evidente para la documentación de una lengua son los dominios léxicos que engloban aspectos centrales de la sociedad. Una vez más, la antropología lingüística proporciona un ejemplo clásico, las terminologías de parentesco, que fueron parte central de la etnografía comparada y son un área de gran preocupación personal y conceptual para los hablantes de muchas lenguas en peligro de extinción (véase también el capítulo 8). En las sociedades en que las categorías sociales centrales están definidas por las relaciones familiares, ya sea que estén construidas genealógicamente o de

¹²En el sitio de Internet del libro se encuentra disponible un breve video utilizado para extraer descripciones de las acciones para "insertar" en tzotzil.

¹³En <http://wdrmaus.de/lachgeschichte/mausspots> se encuentran disponibles muestras del tipo de caricaturas que me resultaron útiles para estas tareas en formato de video *streaming*.

otra manera, la terminología que denota tales categorías es esencial para cualquier caracterización de la vida social. La asimetría en la terminología del tzotzil para el parentesco entre hermanos, por ejemplo, es reveladora de las relaciones familiares. Para un ego masculino, el tzotzil distingue a los hermanos mayores y menores (*bankil*, *itz'in*) de las hermanas mayores y menores (*vix*, *ixlel*). Sin embargo, para un ego femenino, se neutraliza la distinción de género entre los hermanos y las hermanas menores. Así, una hablante distingue al hermano mayor de la hermana mayor (*vibuel*, *vix*), pero identifica a los hermanos menores de ambos géneros (*muk*). Obsérvese además que la distinción del género del ego se neutraliza precisamente en el caso del término usado para la hermana mayor, que es *vix* para los hablantes de ambos sexos (véase la Figura 6). Estas asimetrías sugieren que la relación entre la hermana mayor y sus hermanos menores de cualquier género está especialmente marcada terminológica y conceptualmente. Una explicación plausible es que en muchas comunidades en las que se habla tzotzil se espera que la hermana mayor asuma responsabilidades similares a las de la madre en el cuidado de sus *muk* o hermanos menores, sin importar su sexo. Este cuidado especial se corresponde con la proyección terminológica recíproca de los hermanos menores, quienes consideran a su *vix* o hermana mayor como una especie de madre sustituta.

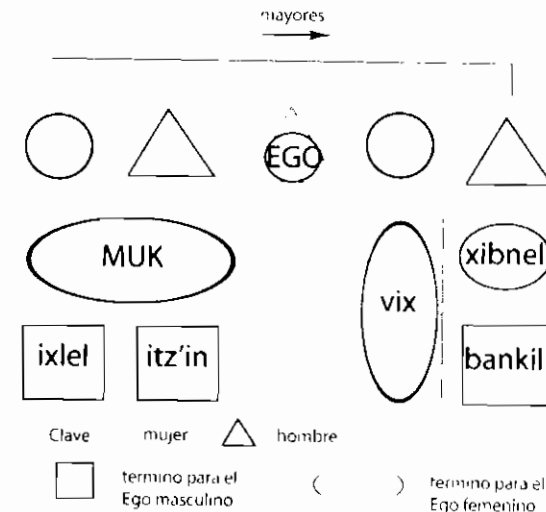


Figura 6. Términos del tzotzil para los hermanos

Sin embargo, como lo muestra el debate clásico, el "álgebra" y los diagramas de parentesco ocultan un problema central en la documentación del conocimiento léxico antes mencionado: la tensión entre los llamados metalenguajes "éticos" y las categorías "émicas". En cualquier lengua, es plausible poner en cuestión si los términos supuestamente universales para caracterizar una relación de parentesco en particular (en términos, digamos, de género, generación y linaje, o con términos de relación presuntamente primitivos, como F[ather], padre, M[other], madre, H[usband], esposo, W[ife], esposa, o con símbolos algebraicos como +, -, . . .) hacen justicia al significado de un término natural del idioma o de una relación específica entre dos individuos. Efectivamente, en sociedades que muestran una evidente obsesión con el parentesco y su terminología (por ejemplo, las comunidades aborígenes australianas en las que he trabajado), un área central de disputas y altercados conceptuales suele ser precisamente cómo asignar la etiqueta léxica adecuada a una relación o cómo explicar lo que implica una relación determinada que se designa con un término preciso. Mi principal maestro de guugu yimithirr, por ejemplo, acostumbraba señalarme a un pariente que iba pasando y decía: "a ése lo tienes que llamar X, porque su padre fue tu W, pero resulta que vino y se casó con tu Y, entonces, ¿en qué se convirtió?, ¿en tu Z?" No sólo la relación genealógica entre dos individuos determina cuál puede ser el término de parentesco pertinente, pues éste puede ser consecuencia de factores considerablemente más complejos sobre qué aspectos de la relación son los más sobresalientes. En el Zinacantán moderno, hay casos en que una relación ritual de compadrazgo (entre los padres y padrinos de un niño recién bautizado, por ejemplo) puede de hecho anteponerse a una relación genealógica inmediata: los hermanos que se vuelven compadres pueden dejar de interpelarse en términos de hermanos.

Para los fines de una documentación sistemática, este dominio ilustra de nueva cuenta la tensión entre un "corpus" de ejemplos y la elicitación sistemática de información. Ninguna red particular de relaciones genealógicas/sociales reales y sus distinciones terminológicas correspondientes puede aspirar a captar la sistematicidad del complejo terminológico-conceptual total. Tampoco existe un metalenguaje extensional (como la genealogía primitiva del álgebra de parentesco) que garantice que todas las variables socialmente significativas emerjan de la elicitación mecánica. Una base de datos léxicos adecuada debe combinar ambos tipos de información.

4.3. La realidad pragmática

En la bibliografía reciente se dificulta aún más encontrar métodos para enriquecer una base de datos léxicos de manera que incluya el uso de unidades lingüísticas indexicales inextricablemente unidas al contexto. Cualquier comportamiento lingüístico está, por supuesto, ligado al contexto y unido a la acción, pero algunos de los artículos léxicos más difíciles de tratar suelen tener vínculos *inherentes* a su entorno indexical: los pronombres y otros deícticos son los ejemplos más obvios, pues inclusive sus referentes (a quienes escogen) deben registrarse haciendo referencia a los contextos en que se usan. Los estudios sobre estos dominios léxicos sugieren que el único enfoque práctico para describir dichas partes del léxico es un tipo de trabajo de campo de observación exhaustiva. De este modo, Hanks (1990) ofrece un detallado análisis del sistema de demostrativos del maya yucateco basado en un extenso trabajo de campo, en el que registró detalladamente dónde ocurrían de manera espontánea los deícticos, induciendo del corpus y de las formas lingüísticas los componentes teóricos para presentar una explicación adecuada de la práctica deíctica.

Otro dominio ilustrativo es el de las exclamaciones e interjecciones. El amplio trabajo hecho por Kockelman con las interjecciones en q'eq'chi (Kockelman 2003) requirió una metodología de campo muy similar a la de Hanks. El autor registró sistemáticamente las circunstancias en que las enumeraciones categorizadas como interjecciones ocurrían en una comunidad de habla q'eq'chi en Guatemala. Basándose en este corpus elaboró una teoría de las interjecciones que va mucho más allá del modelo generalmente aceptado de que su carácter es "expresivo" (parte de una vieja tradición en el pensamiento lingüístico occidental que se remonta hasta los gramáticos latinos), para pasar a considerar las propiedades indexicales bidireccionales y múltiples de estas expresiones: manifiestan actitudes emotivas y afectivas, invitan explícitamente a los interlocutores a expresar a su vez las suyas, conminan a actuar, etc. Los estudios de esta índole sugieren que hay pocos atajos para producir una explicación adecuada de lo que significan estos elementos lingüísticos con carga pragmática, y que el trabajo de campo etnográfico extensivo es, por lo tanto, parte esencial de la lexicografía de campo.

Lo mismo puede decirse de un vocabulario más prosaico, desde las partes del cuerpo comunes y corrientes hasta los registros especialmente marcados como corteses y descorteses, como en el lenguaje para bromear y maldecir. Ya he mencionado las complejidades léxicas residuales generadas

en guugu yimithirr por el uso alternado de las formas respetuosas y el "vocabulario del cuñado", estas complejidades sólo se multiplican cuando los diversos registros del habla establecidos se encuentran en uso activo en una comunidad de habla. Las descripciones antropológicas clásicas de estos fenómenos dan fe de la sutileza y los matices transmitidos por la elección estratégica entre formas léxicas alternativas en sociedades de la Australia aborigen, de Samoa y también de Bali (Duranti 1992; Errington 1984; Geertz 1960), o entre términos para interpelar a alguien y los pronombres personales desde Europa hasta Japón (Brown y Gilman 1960). Laughlin (1975) propone una serie de etiquetas para distinguir en el tzotzil zinacanteco cosas como "habla ritual, habla para bromear, habla masculina y femenina, habla de los bebés, habla cortés, regaños, habla para denunciar, [palabras] arcaicas", etc. Más allá de si el lexicógrafo de campo puede o no proporcionar una explicación exhaustiva de estos hechos para una base de datos léxicos completa, es importante conocer los tipos de categorías metalingüísticas del habla que podrían ser relevantes en la propia comunidad de habla.

Por razones que saltan a la vista, la investigación sistemática de tales géneros —por ejemplo, del habla tabú— puede resultar difícil para personas sin experiencia en el trabajo de campo. Igualmente difíciles son los sistemas completos de tropos lingüísticos que a veces dominan partes de los recursos expresivos de un idioma. Una vez más, el único remedio parece ser un espectro de amplio alcance y una atención etnográfica sistemática. Los siguientes son dos ejemplos de mi propio trabajo de campo. Cuando estaba aprendiendo guugu yimithirr, advertí que muchas expresiones concernientes a inclinaciones humanas y "estados interiores" eran metáforas transparentes basadas en un pequeño conjunto de palabras que simplemente parecían nombrar partes del cuerpo. Independientemente de que estas expresiones representen —como en ciertas ocasiones han sugerido los antropólogos— ya sea una teoría implícita de la distribución anatómica de las emociones y las facultades mentales (lo que podría aducirse con expresiones en inglés o español como *hard-headed* "cabeza dura" o *hard-hearted* "corazón de piedra") o bien expresiones opacas convertidas en convenciones culturales (como podríamos aducir en el caso de *green-thumb* y *lily-livered* en inglés o "mano dura" y "sangre pesada" en español),²¹ no había duda de que el

²¹La expresión *green thumb*, literalmente "pulgar verde" que se usa para referirse a alguien que tiene "buena mano" para la jardinería, parece no ser exclusiva del inglés, pues en italiano se dice *pollice verde* (según Elena Collavin) y en alemán *grüne Daumen* (según Nikolaus Himmelmann). En cuanto a *lily-livered* "cobarde", literalmente "hígado de lirio", hay una expresión en italiano que sugiere el mismo significado: *corza leghato* "sin hígado".

guugu yimithirr tenía un sistema semiproductivo para generar diversas expresiones a partir de tropos con "partes del cuerpo". El ejemplo 11 está basado en la palabra *miil*, "ojo", en guugu yimithirr. La única forma que tenía para documentar el sistema era parar la oreja (por decirlo así) para atrapar expresiones relevantes en la conversación y tratar sistemáticamente de forzar nuevas combinaciones de partes del cuerpo con adjetivos y verbos, lo que generalmente producía más carcajadas que nuevos lexemas.

11. Expresiones en guugu yimithirr basadas en *miil* "ojo"

<i>miilgu</i>	= (lit., ojo + sufijo ENFÁTICO) despierto
<i>miil warnngu</i>	= (lit. "ojo dormir") somnoliento
<i>miil nhin-gal</i>	= (lit., "ojo sentar") vigilar, mantenerse atento a algo
<i>miil biyal</i>	= (lit., ojo tendón) que mira todo el tiempo
<i>miil ngamba</i>	= (lit. ojo descuidado) distraído, que eierra los ojos ante algo
<i>miil waarril</i>	= (lit., ojo volar) marearse, enloquecer, desmayarse, emborracharse ²¹
<i>miil bagal</i>	= (lit., ojo picar) engañar, embaucar, ponerse celoso
<i>miil bathibay</i>	= (lit., ojo hueso) de vista aguda, que siempre mira insistentemente
<i>miil biinii</i>	= (lit., ojo morir) quedarse ciego
<i>miil gulnggul</i>	= (lit., ojo pesado) somnoliento
<i>miilgu nhin-gal</i>	= (lit., ojo + sufijo ENFÁTICO sentar) mantenerse despierto...

Considérese también el lenguaje del tzotzil ritual (Gossen 1974, 1985; Haviland 1987, 1996, 2000). En contextos que van desde el rezo y la canción hasta la denuncia formal, los hablantes de tzotzil dejan de lado el léxico y la gramática comunes y recurren a un estilo discursivo sumamente estructurado con líneas paralelas que difieren por una sola palabra o frase. Estas líneas paralelas se interpretan en términos de una imagen estándar "estereoscópica" (Fox 1977), invocada por los dobles expresivos. Así, para referirse al cuerpo pueden usarse diferentes dobles, según el contexto. Uno es sumamente literal y emplea *pat*, *xokon* "dorso, costado" como una metonimia para el todo. Otro es considerablemente más opaco y sugiere una imagen de humildad, como en el siguiente extracto de un rezo

²¹Aquí paso por alto cuestiones sintácticas básicas. Por ejemplo, en la expresión *miil waarril*, la palabra *miil* "ojo" es el sujeto sintáctico de *waarril* "volar", mientras que en *miil bagal*, *miil* es el objeto sintáctico de *bagal* "picar".

curativo, en el que el doblete *lumal, ach'elal* "tierra, barro" (ambos con un posesivo) se refiere al cuerpo o al yo del paciente.²²

12. De un rezo curativo zinacanteco

ja' mc ta jmalalalumal
espero tu tierra

ta jmalalavach'elale
espero tu barro

Otro ejemplo es el doblete del habla ritual zinacanteca para referirse al licor, *xi'obil, sk'exobil*, literalmente "causa de temor, causa de vergüenza". Estas expresiones comparten características con el eufemismo, que siempre es un fenómeno problemático en la lexicografía y exige un cuidadoso trabajo de campo etnográfico.

La elicitación sistemática revela poco sobre el sistema general de imaginación en el lenguaje ritual, aunque es parte esencial del poder expresivo del idioma. El diccionario de Laughlin del tzotzil zinacanteco moderno anota e ilustra palabras que forman parte de construcciones paralelas bajo el encabezado "habla ritual". En mi propio trabajo, me he apoyado en el registro exhaustivo y la transcripción de rezos y otros géneros que emplean el paralelismo para expandir la lista de los dobletes.

5. Conclusión

¿Cuándo termina la documentación del léxico? Si bien el lexicón es un depósito para lo excepcional y lo caótico del idioma, también es una mina de considerable regularidad y productividad. Sin embargo, los lexicógrafos de campo como Laughlin expresan sus dudas sobre qué tan estructurado o generalizado está el conocimiento léxico en cualquier comunidad de habla, y basa su escepticismo en sus experiencias de elicitación tanto con campesinos zinacantecos como con estudiantes universitarios de Washington D.C. Notablemente difícil aun para los idiomas bien estudiados es la distinción entre los usos "literales" y los usos "figurados" o trópicos de las palabras: los hablantes más viejos del tzotzil describen los aviones como *xulem k'ok'*, literalmente (como nosotros decimos) "fuego de buitre" o los

²² En el tzotzil del vecino Larráinzar, el doblete ritual equivalente es a la vez humilde y literal: *ach'elal, takopal* "el barro, el cuerpo".

teléfonos como *ch'ojon tak'in*, "alambre de metal", y soportan las risillas burlonas de los hablantes más jóvenes (que en cambio usan sencillamente un préstamo del español). Más difícil aún es distinguir una polisemia oscura de una simple (pero formalmente desabrida) homonimia. El diccionario tzotzil de Laughlin plantea dos raíces homónimas, *jav* (2)—una raíz posicional que significa "panza (o boca) arriba"—y *jav* (1)—una raíz de verbo transitivo que significa "cortar en dos"—porque ambos significados parecen lo bastante divergentes para avalar entradas independientes. Sin embargo, la etimología popular del zinacanteco conjura una imagen sucinta que relaciona los dos sentidos: cuando se parte en dos un árbol, por ejemplo (acción espresada por un verbo basado en *jav* (1)), las dos mitades caen "panza arriba" (*jav* (2)). Aquí tenemos entonces un caso de polisemia encubierta,²³ o quizás de monosemia subyacente de una sola raíz con diferentes disfraces gramaticales. Estos fenómenos pueden permanecer sin trabajarse a lo largo de todo un proyecto de documentación léxica.

Asimismo, ¿hasta dónde debe el lexicógrafo incluir lo que puede catalogarse como "uso erróneo", es decir, palabras mal empleadas, juegos de palabras o palabras inventadas para la ocasión? Zgusta (1971: 56-57) distingue los usos "sistémicos" de los "ocasionales". Un autor puede usar "atadura" en vez de "matrimonio" sin por ello cambiar el significado sistémico de ninguna de las dos palabras. En 1970, durante varias semanas de sesiones de charlas lujuriosas, los hombres de Zinacantán acuñaron lo que en ese tiempo resultaba un eufemismo sexual de lo más creativo en tzotzil usando el préstamo *inyeksyon* del español "inyección", en una época en que las inyecciones hipodérmicas todavía eran algo relativamente novedoso traído de fuera. Algunos de estos hombres siguen usando el término en broma, casi cuarenta años después. La palabra no está en el diccionario tzotzil de Laughlin, pero quizás tendría que estar.

Por último, las cuestiones ya mencionadas sobre objetivos y usuarios—¿para quiénes se produce una base de datos léxicos?, ¿a qué propósitos se prestará?—complican las decisiones sobre qué palabras deben documentarse y de qué modo. Los problemas son especialmente fastidiosos cuando una base de datos léxicos puede servir de base para la estandarización o estabilización lingüística, especialmente en forma de un diccionario publicado.²⁴ Cuando la gente puede usar un diccionario para buscar una palabra, para ver cómo se escribe o para leer una definición, se altera irrevocablemente la autoridad de la comunidad de habla sobre el uso

²³ Véase la discusión sobre polisemia en Zgusta (1971: 77 y ss.); véanse también Evans y Wilkins (2000, 2001), Evans (1992).

²⁴ Véase la discusión de Jane Hill sobre el proyecto de diccionario del hopi en el capítulo 5.

“correcto”. Cuánto debe incluirse en la base de datos léxicos de un proyecto de documentación de un idioma es una cuestión que no se limita a la “exhaustividad” ni tampoco es una cuestión de “cobertura”; también implica decisiones ideológicas que pueden tener efectos de largo alcance en el futuro de un idioma.

Construir una base de datos léxicos es una parte esperada de cualquier proyecto de documentación, y quizás la tarea analítica más exigente de todas, a final de cuentas. Puede apoyarse en técnicas mecánicas aplicadas a corpora textuales y en el conocimiento profundo de las grandes tradiciones lexicográficas, que ya han lidiado con la mayoría de los problemas a los que se enfrenta quien hace trabajo de campo: unidades léxicas, la naturaleza del significado, las vaguedades del uso y, finalmente, las ideologías del idioma y la vida social. El producto final es esencial, pero producirlo depende tanto de la inspiración como del trabajo etnográfico minucioso, y tanto de la elicitación como de los hallazgos afortunados. Invariablemente, (re)descubrimos que el trabajo hecho *nunca* es suficiente y que cuando ponemos un alto y declaramos cerrada la base de datos, sólo nos detenemos a descansar en un lugar cualquiera de un largo camino.

Agradecimientos

Este capítulo, que se basa en términos generales en la conferencia presentada en la escuela de verano del programa DoBeS (Frankfurt, septiembre de 2004), tiene una fuerte deuda con las experiencias de lexicografía compartidas con mis maestros de *tzotzil* y *guugu yimithir*, con los comentarios de Nikolaus Himmelmann y Jost Gippert y con la hospitalidad de Elena, Renato y Lisetta Collavin durante la redacción final.

Capítulo 7

La prosodia en la documentación lingüística

Nikolaus P. Himmelmann

Introducción

Los aspectos prosódicos de un mensaje lingüístico, como la entonación y acento léxico, son elementos esenciales de su estructura formal. A la fecha lo esencial del análisis de los rasgos prosódicos todavía no se ha convertido en parte integral de la capacitación en trabajo de campo lingüístico y, consiguientemente, una documentación y descripción razonablemente detallada amplia de los mismos todavía no forma parte de las prácticas estándares del trabajo de campo lingüístico. Este capítulo aborda específicamente *documentación* de rasgos prosódicos, esto es, la cuestión de qué clase de datos tiene que contener una documentación lingüística para que se pueda realizar un análisis prosódico lo más completo posible. Para aprovechar sugerencias analizadas en este capítulo, es necesario un entendimiento básico de las unidades y procedimientos usados en el análisis prosódico; se desea revisar una introducción más amplia al trabajo de campo prosódico centrado en cuestiones de análisis y descripción, véase Himmelmann y Ladd (en prensa).

Dado que una documentación lingüística incluye un gran corpus de grabaciones de eventos comunicativos de diferentes tipos, bien puede cuestionarse si existe alguna necesidad de prestar atención especial a la prosodia cuando se compila dicha documentación. Siempre que las grabaciones sean de calidad aceptable,¹ no hay duda de que un corpus semejante puede usarse para analizar la prosodia incluso cuando no se ha prestado atención particular a los rasgos prosódicos al compilar el corpus.

¹ Los rasgos que definen una buena grabación se enlistan en el apartado 2.1 del capítulo.

² Ejemplos de lo que puede hacerse—y de lo que no puede hacerse—en términos de análisis prosódico sólo a partir de un corpus de grabaciones se encuentran en las tesis sobre entonación del *dyirbal* y del *bininj gun-wok* de King (1994) y Bishop (2004).