

---

## REFLEJO DE LA POLIVALENCIA SEMÁNTICA EN EL CORPUS INTERCORP Y EN LOS DICCIONARIOS BILINGÜES ESPAÑOL-CHECO\*

DANIEL PETRÍK

Universidad Carolina, Praga

### REFLECTION OF THE SEMANTIC POLYVALENCE IN THE PARALLEL CORPUS INTERCORP AND CZECH-SPANISH DICTIONARIES

Lexeme translation between different languages can be viewed as finding equivalences between expressions. Since in these relations isomorphism is normally less frequent, it is important to take into account semantic polyvalence (polysemy or homonymy). In the case of ambiguous expressions the key of the translation is to perform properly the disambiguation, which can be done by monitoring both contexts in the corpus or lexeme translation dictionaries. As to the representation of polysemic words with a large number of meanings in the dictionary, it is necessary to select the meanings to present and in some way to organize their definition. In practice, it is desirable that such arrangement respects both the frequency of occurrences of meanings in the corpus, as well as the rate of equivalence, so that the less numerous meanings and those with a narrow equivalence rate are listed last. The article compares whether selected expressions in two large Spanish-Czech dictionaries reflect these criteria based on the frequency obtained from the parallel corpus InterCorp, or rather point out the shortcomings of this corpus.

**Key words:** polysemy – homonymy – equivalence – disambiguation – frequency of meanings

**Palabras claves:** polisemia – homonimia – equivalencia – desambiguación – frecuencia de significados

---

### 1. Introducción

Es bien sabido que a la hora de traducir un texto de una lengua a otra (llamémoslas lengua origen y lengua término respectivamente) uno de los mayores problemas surge del hecho de que entre las palabras de dos lenguas diferentes no siempre existen relaciones mutuamente unívocas. Con frecuencia ocurre que un término español se puede traducir mediante más de una palabra checa dependiendo del contexto en el que aparezca; incluso en un diccionario español monolingüe tal término podría adoptar varias acepciones, o si

---

\* Este artículo forma parte del proyecto “Program rozvoje vědních oblastí na Univerzitě Karlově č. P10 *Lingvistika*”, subprograma “*Románské jazyky ve světle jazykových korpusů*”.

queremos, significados. La situación que acabamos de describir la abarca el término de la polivalencia semántica. Dentro de la polivalencia encontramos dos fenómenos básicos: la polisemia y la homonimia entre los cuales en la práctica, a veces, resulta difícil hacer diferencias de manera infalible. La teoría de conjuntos aplicada al campo semántico (Pottier 1970) afirma que en la polisemia dos significados diferentes de un término comparten al menos un sema, mientras que en la homonimia estos significados no tienen ningún sema en común. Como el análisis de semas nunca es objetivo del todo, es posible en algunos términos vacilar si se trata de polisemos u homónimos. Sin embargo, en este texto la diferencia entre ambos fenómenos no va a jugar ningún papel importante y siempre nos conformamos con que se trata de palabras polivalentes.

En nuestra época de creación de traductores automáticos una de las tareas claves relacionadas con la polivalencia semántica es la desambiguación, o sea cómo diferenciar dos significados distintos de una palabra utilizada en la comunicación. Los informáticos solucionan este problema observando los contextos en los que aparece el término en un gran número de textos o discursos y a cierto contexto se le atribuye el significado correspondiente. Asimismo un corpus paralelo checo-español suficientemente amplio y compuesto de textos variados se puede convertir en un instrumento muy útil para la desambiguación. Añadamos que antes de la era computacional la única vía para los traductores en caso de términos polivalentes era, después de consultar un diccionario, hacer su propio juicio acerca del significado pertinente de todos los posibles, partiendo del contexto. No obstante, en la desambiguación a través de un corpus o un diccionario bilingüe completado por el juicio del traductor podemos enfrentarnos con algunas complicaciones. Se trata de insuficiencias tanto en la parte del corpus como en la parte del diccionario y sus creadores. En este texto presentaremos algunas de estas insuficiencias a base del estudio de ciertas palabras polivalentes elegidas según nuestro juicio con la intención de presentar ejemplos de distintas clases de palabras y con posible violación de algunos principios mencionados en este texto.

## 2. Perspectiva bilingüe en la polivalencia semántica

Analizando la equivalencia entre lexemas en dos lenguas diferentes llegamos a los siguientes casos:

- a) isomorfismo o equivalencia unívoca (cuando un lexema en la lengua origen corresponde de manera unívoca a un lexema en la lengua término),
- b) anisomorfismo o equivalencia parcial (cuando uno o más lexemas en la lengua origen está(n) asociado(s) con más lexemas en la lengua término o viceversa),
- c) lagunas o ausencia de equivalencia (cuando en la lengua término no encontramos ningún lexema que corresponda al lexema en la lengua origen)

El caso del isomorfismo no es muy frecuente y ocurre, sobre todo, en términos internacionalmente definidos (por ejemplo, 'la gripe' (*chřipka*)), mientras que el anisomorfismo es la equivalencia léxica predominante. Dado al concepto del relativismo lingüístico, desde la perspectiva checo-española en el marco del anisomorfismo existen términos polivalentes exclusivamente en español (por ejemplo, 'el gato' cuyos dos significados básicos son 1º: animal felino (*kočka*), 2º: herramienta para levantar el coche (*zvedák*),

*hever*), términos polivalentes únicamente en checo (por ejemplo, *strana* cuyos significados pueden corresponder a términos ‘lado’, ‘página’, ‘parte’, ‘partido’, etc.) y finalmente los términos polivalentes de modo similar en las dos lenguas (por ejemplo, ‘pluma’ (*pero*) con los significados de 1º: instrumento para escribir, 2º: una parte de la cubierta del cuerpo de aves). En este texto siempre partiremos de realizar la desambiguación de términos polivalentes por medio de la traducción entre el checo y el español, ya que el corpus paralelo se basa en la idea de comparar traducciones del mismo texto en dos o más lenguas. Obviamente si la polivalencia se da solamente en una de las dos lenguas, podemos recurrir al uso del corpus paralelo que nos facilita la tarea antes de consultar un diccionario, ya que los textos han sido traducidos previamente. Aún más complicado es el caso de la polivalencia que aparece de modo similar en las dos lenguas como en el caso del término ‘pluma’. En el corpus paralelo inmediatamente se observan los significados correspondientes de la palabra, pero utilizando un diccionario el traductor debe confiar en su juicio cuestionando el contexto de la palabra y el sentido de los sintagmas en los que la palabra aparece.

### 3. Algunos aspectos de la representación del léxico en diccionarios bilingües

Un diccionario bilingüe es, simplemente dicho, una lista de entradas léxicas. Todas las entradas se organizan según un esquema jerárquico, denominado microestructura, distinto en cada diccionario. La definición de la entrada (o de cada una de sus acepciones) suele ser la información central de la microestructura, que además puede incluir información morfológica, lemas, variantes (de las formas flexivas), categoría gramatical, informaciones fonéticas, etimología, partición silábica, marcas de uso, textos de ejemplificación, remisiones semánticas, derivados, compuestos, etc. En este artículo nos interesamos, ante todo, por las informaciones de carácter semántico, o sea cómo está definido el significado de cada acepción de un lexema. Excluimos de nuestro estudio las definiciones por paráfrasis o comprensión, puesto que nos centramos en la equivalencia léxica entre dos lenguas. En los párrafos siguientes trataremos las preguntas del tipo: ¿Según qué criterios ordenar varias acepciones de una entrada?, ¿cómo jerarquizar las acepciones?, ¿es conveniente agrupar acepciones similares en una entrada o dejarlas completamente separadas?, ¿cómo representar los términos polifuncionales?

En el caso de palabras polivalentes, si en la lengua origen a un lexema se le atribuyen más acepciones de la lengua término, es frecuente que dichas acepciones no sean independientes del todo y mantengan ciertas correlaciones (por ejemplo, un significado puede subordinarse a otro, etc.). En consecuencia, ante la polivalencia el lexicógrafo debe tomar algunas decisiones sobre cómo ordenará las acepciones y de qué modo establecerá jerarquías entre los significados. *Manuál lexikografie* (Čermák 1995) presenta dos puntos de vista básicos a la hora de tratar los posibles criterios de la ordenación de acepciones: la frecuencia de su aparición en el corpus y el grado de equivalencia. Acorde con estos aspectos, las acepciones menos frecuentes y de una equivalencia restringida o especializada deberían aparecer en el último lugar de la ordenación. Otros criterios relevantes para la ordenación se pueden basar en la etimología, en el testimonio histórico, en el

encadenamiento lógico de significados y en el uso. No obstante, anotemos que desde el punto de vista puramente práctico es deseable que las acepciones se ordenen en función de su frecuencia en el corpus, o sea que los equivalentes más corrientes (o frecuentes) en las traducciones en una lengua aparezcan al principio de la ordenación para que el usuario del diccionario se oriente rápido y sin dificultades.

Si se consulta un diccionario amplio, lo que más apreciamos es la claridad y fácil orientación, la cual se complica en proporción al número de acepciones (o significados) de las que conste la entrada. Por lo consiguiente, en el caso de las llamadas “grandes entradas”<sup>1</sup> es racional tanto agrupar significados similares en una sola acepción sin mencionarlos explícitamente, como omitir algunos significados de baja frecuencia en el corpus.

La última cuestión de representación de términos polivalentes que vamos a discutir en este artículo es cómo los diccionarios diferencian entre la homonimia, polisemia y polifunción de las entradas. Desde el punto de vista funcional es posible distinguir los homónimos lexicales<sup>2</sup>, gramaticales<sup>3</sup> y léxico-gramaticales<sup>4</sup>. En la clasificación previa podemos incluir también la polisemia bajo la homonimia lexical o léxico-gramatical. La polifunción se asociaría con la homonimia léxico-gramatical. A continuación observaremos que los diccionarios analizados muestran un grado diferente de agrupación de los diferentes significados en acepciones en caso de homonimia o polisemia léxica. Los homónimos gramaticales o léxico-gramaticales se mencionan a través de un exponente específico en dos entradas separadas.

#### **4. Estudio de la representación de términos concretos: ‘banco’, ‘cuarto’ y ‘saber’**

La parte práctica de este artículo se orienta a comparar la representación de algunos términos polivalentes en dos grandes diccionarios español-checos de la actualidad con sus equivalentes<sup>5</sup> correspondientes registrados en el corpus paralelo español-checo InterCorp. Para cada uno de los términos vamos a mencionar en forma de tabla los equivalentes de los dos diccionarios en orden de aparición y también los lemas procedentes del análisis de equivalencias del corpus organizándolos del lema más frecuente al menos frecuente. En el corpus siempre examinamos todas las acepciones de ambos diccionarios, sin embargo, no mencionamos en la columna correspondiente a InterCorp las acepciones cuya frecuencia de aparición es cero. También excluimos de las tablas algunas definiciones parafrásicas, ejemplificaciones o algunos compuestos registrados por los diccionarios, al igual que algunos lemas de traducción ausentes en el corpus.

---

<sup>1</sup> Términos cuya definición exhaustiva podría contener decenas o cientos de significados como, por ejemplo, el verbo ‘pasar’.

<sup>2</sup> Expresiones homonímicas que son de la misma categoría gramatical (‘banco’: *banka* o *lavička*).

<sup>3</sup> Expresiones homonímicas que no son de la misma categoría gramatical (‘vino’: *víno* o *příšel*).

<sup>4</sup> Expresiones homonímicas en las que se ha registrado cierta transformación gramatical, como la sustantivación del infinitivo, etc. (‘saber’ o ‘el saber’: *vědět* o *vzdělání*, *vědomost*).

<sup>5</sup> Bajo los términos ‘equivalente’ o ‘lema’ entendemos aquí no sólo la forma básica de un término (o sea el infinitivo, el singular o el masculino dependiendo de qué clase de palabras estamos tratando), sino todas las variantes gramaticales de éste (por ejemplo, en caso de verbos, cualquier tiempo, modo, número, persona, etc.).

**Tabla 1.** Las acepciones del término ‘banco’

**i) banco**

	DUBSKÝ, REJZEK	CHALUPA	INTERCORP Nº DE APARICIONES		FRECUENCIA RELATIVA
1.	<i>lavice, sedátko, stůl, ponk, stolice</i>	<i>lavice, sedátko</i>	<i>banka</i>	1488	66,43%
2.	<i>hoblíce</i>	<i>stůl, ponk, stolice</i>	<i>lavička</i>	244	10,89%
3.	<i>útes</i>	<i>útes, mělčina</i>	<i>lavice</i>	214	9,55%
4.	<i>kra</i>	<i>kra, iceberg</i>	<i>stolička</i>	12	0,54%
5.	<i>ložisko, hejno</i>	<i>ložisko, hejno</i>	<i>hejno</i>	12	0,54%
6.	<i>vrstva</i>	<i>vrstva</i>	<i>píščina</i>	4	0,18%
7.	<i>píščina, mělčina</i>	<i>banka (krevní banka, databanka, spermabanka, etc.)</i>	<i>mělčina</i>	4	0,18%
8.	<i>banka (zemědělská, pověřená, ..., spořitelna, databanka, krevní banka, etc.)</i>		<i>ložisko</i>	4	0,18%
9.			<i>sedátko</i>	4	0,18%
10.			<i>databanka</i>	3	0,13%
11.			<i>stůl</i>	3	0,13%
12.			<i>sedadlo</i>	1	0,04%
13.			<i>ponk</i>	1	0,04%
14.			<i>trojnožka</i>	1	0,04%

(Nº total de apariciones: 2240)

Si reflexionamos sobre la etimología de la palabra ‘banco’<sup>6</sup>, podemos observar que desde la Edad Media hasta la actualidad el término ha pasado por un proceso de una intensiva divergencia semántica, ya que los bancos actuales como instituciones financieras no tienen nada en común con los bancos que sirven para sentarse o trabajar. Además de estos dos significados básicos se registran otros más. Es curioso que ninguno de los dos diccionarios coloque en primer lugar el significado (*banka* y *databanka*) que mayormente (con frecuencia de 66,5%) aparece en el corpus. A lo mejor los lexicógrafos en este caso han preferido la sucesión etimológica de los significados al grado del uso actual en la lengua. El diccionario de Dubský y Rejzek agrupa los significados *lavice, sedátko, stůl, ponk, stolice* bajo una acepción, mientras que el diccionario de Chalupa distingue entre los tipos de asiento (*lavice, sedátko*) y los muebles sobre los que se trabaja y normalmente no sirven para sentarse (*stůl, ponk, stolice*). Además el primer diccionario menciona un equivalente técnico (*hoblíce*) que no se registra en el otro diccionario, ni en el corpus. En

<sup>6</sup> Los primeros bancos italianos eran mesas o bancos (*stolice*) que se instalaban en las plazas de las ciudades, allí los primeros banqueros recaudaban el dinero de los ciudadanos a cambio de un boleto o billete en el cual el banquero reconocía su deuda.

el resto de las acepciones no notamos mayores diferencias entre los dos diccionarios. El corpus ofrece un equivalente bastante especializado (*trojnožka*) que no está registrado en los diccionarios. Las acepciones *lavička*, *lavice*, *stolička*, *trojnožka*, *sedátka* y *sedadlo* juntas representan el segundo significado más frecuente en el corpus (21,25%) y las demás acepciones muestran una frecuencia insignificante (no superior a 1%) en el corpus.

**Tabla 2.** Las acepciones del término ‘cuarto’  
ii) **cuarto**

	DUBSKÝ, REJZEK	CHALÚPA	INTERCORP Nº DE APARICIONES		FRECUENCIA RELATIVA
1.	<i>čtvrtý</i>	I) <i>čtvrtý</i>	<i>pokoj, pokojík</i>	1097	32,75%
2.	<i>čtvrtina</i>	<i>čtvrté poschodí</i>	<i>čtvrtý</i>	590	17,61%
3.	<i>čtvrthodina, čtvrt, čtvrt’ (Měsíce)</i>	II) <i>čtvrt, čtvrtka, čtvrtina, čtvrthodina</i>	<i>místnost</i>	297	8,87%
4.	<i>kvartový formát, kvart</i>	<i>pokoj, místnost (koupelna, obývací pokoj, komora, ...)</i>	<i>čtvrt</i>	261	7,79%
5.	<i>pokoj, místnost, přízemí, koupelna, obývací pokoj, herna, velín, formovna, kormidelna, komora</i>	<i>čtvrťfinále</i>	<i>koupelna</i>	206	6,15%
6.	<i>předek o zadek zvířete</i>	<i>peníze, (šesták)</i>	<i>čtvrtina</i>	193	5,76%
7.	<i>čtyrák, vindru</i>		<i>byt, byteček</i>	41	1,22%
8.			<i>ložnice</i>	39	1,16%
9.			<i>záchod</i>	35	1,04%
10.			<i>komora</i> <sup>7</sup>	29	0,87%
11.			<i>čtvrthodina</i>	17	0,51%
12.			<i>čtvrťák, čtvrtáci</i>	16	0,48%
13.			<i>čtvrtka</i>	10	0,30%
14.			<i>čtvrťák, čtvrtáky</i>	7	0,21%
15.			<i>peníze</i>	7	0,21%
16.			<i>čtvrť</i>	7	0,21%
17.			<i>hala</i>	7	0,21%
18.			<i>čtyři</i>	4	0,12%
19.			<i>cela</i> <sup>8</sup>	4	0,12%
20.			<i>obývánk</i>	4	0,12%

<sup>7</sup> ‘Cuarto de los armarios’, etc.

<sup>8</sup> ‘Cuarto blindado’.

	DUBSKÝ, REJZEK	CHALUPA	INTERCORP Nº DE APARICIONES		FRECUENCIA RELATIVA
21.			<i>šesták</i>	4	0,12%
22.			<i>kotelna</i> <sup>9</sup>	3	0,09%
23.			<i>rameno, ramínko</i>	3	0,09%
24.			<i>garsonka</i>	2	0,06%
25.			<i>pracovna</i>	2	0,06%
26.			<i>čtvrťfinále</i>	2	0,06%
27.			<i>kormidelna</i>	2	0,06%
28.			<i>salón</i>	1	0,03%
29.			<i>sál</i>	1	0,03%
30.			<i>kuchyně</i>	1	0,03%
31.			<i>vindru</i> <sup>10</sup>	1	0,03%

(Nº total de apariciones: 3350)

El diccionario de Dubský y Rejzek recoge todas las acepciones del lexema ‘cuarto’ en una sola entrada mientras que el diccionario de Chalupa las divide en dos entradas diferentes<sup>11</sup> en función de las clases de palabras. Ambos diccionarios mencionan los significados asociados con numerales ordinales o medidas (*čtvrtý, čtvrt, čvrtina, čvrtka*, etc.) antes que el significado relacionado con la descripción de vivienda (*pokoj, místnost, koupelna, obývací pokoj, komora*, etc.) a pesar de que el corpus InterCorp comprueba que las segundas acepciones son más frecuentes. Es decir que el 53,87% de las apariciones del término ‘cuarto’ se asocia con la descripción de vivienda (41,61% se relaciona con denominaciones generales como *pokoj, pokojík* o *místnost*; 9,97% corresponden a nombres de habitaciones concretas como *koupelna, obývací pokoj, záchod, komora, pracovna, kotelna* que se realizan por medio de compuestos, por ejemplo ‘el cuarto de baño’ o mediante deixis y 1,28% describe una casa entera<sup>12</sup> con equivalentes *byt* o *byteček*) y el 31,85% equivale al numeral ordinal *čtvrtý* o a fracciones o medidas del tipo *čtvrt, čvrtka, čvrtina*, etc. De estas medidas tanto los diccionarios como el corpus mencionan algunos derivados compuestos como *čtvrthodina* (‘cuarto de hora’) o *čtvrťfinále* (‘cuarto de final’) cuya frecuencia de 0,57% no es muy significativa. A parte de los dos significados mayoritarios en los dos diccionarios encontramos también algunos equivalentes asociados con dinero como *peníze, čtvrták, čtyrák, šesták* o *vindru* (cuya frecuencia total en el corpus es de 0,54%) o expresiones que describen partes del cuerpo de animales como *rameno* o *ramínko (předek)* con frecuencia de 0,09%. Mencionemos a modo de curiosidad que dentro de las posibles acepciones del término se puede estudiar un caso de relación homóni-

<sup>9</sup> ‘Cuarto de la(s) caldera(s)’.

<sup>10</sup> Se trata del acusativo del sustantivo *vindra*, que aparece en la locución *nemám ani vindru*.

<sup>11</sup> En la tabla las señalizamos mediante los números romanos.

<sup>12</sup> Anotemos que el significado de *byt* o *byteček* no está registrado en ninguno de los diccionarios estudiados.

mica que se produce de igual modo en las dos lenguas<sup>13</sup>, o sea ‘cuarto’ (*čtvrťák*) para denominar una moneda y ‘cuarto’ (*čtvrťák*) para hablar del cuarto curso o de los que lo cursan. El corpus InterCorp sólo es capaz de evaluar el número total de ocurrencias de la equivalencia ‘cuarto’ ↔ *čtvrťák*, no obstante, ha sido necesario realizar la diferenciación propia de las dos acepciones de otra forma (en nuestro caso simplemente analizamos los contextos y así determinamos si se trata de uno u otro significado).

**Tabla 3.** Las acepciones del término ‘saber’  
iii) **saber**

	DUBSKÝ, REJZEK		CHALUPA		INTERCORP Nº DE APARICIONES		FRECUENCIA RELATIVA
1.	I)	<i>vědět, znát, umět, dovést, rozumět, dozvědět se, nevědomky, bůhví, bůhví, kdoví</i>	I)	<i>vědět</i>	<i>vědět</i>	13533	61,45%
2.		<i>páchnout, zapáchat, vonět, chutnat, smrdět, nelíbit se</i>		<i>znát, mít znalost</i>	<i>dozvědět se</i>	1046	4,75%
3.	II)	<i>vědění, vzdělání, vědomost, znalost, vědomí</i>		<i>mít zprávy, informace</i>	<i>umět</i>	947	4,30%
4.				<i>umět, dokázat</i>	<i>znát</i>	877	3,98%
5.				<i>chutnat, mít chuť</i>	<i>dokázat</i>	693	3,15%
6.				<i>být chytrý, mít za ušima</i>	<i>zjistit</i>	411	1,87%
7.				<i>být (ne)přijemný</i>	<i>tušit</i>	329	1,49%
8.				<i>kdoví</i>	<i>pochopit</i>	306	1,39%
9.				<i>totiž, tj., a sice</i>	<i>dovést</i>	248	1,13%
10.				<i>být úplně vedle, nemít ani páru</i>	<i>chápat</i>	243	1,10%
11.			II)	<i>vědění, vzdělání, vědomost, znalost, vědomí</i>	<i>dovédět se</i>	235	1,07%
12.					<i>zajímat</i>	172	0,78%
13.					<i>rozumět</i>	154	0,70%
14.					<i>vědění</i>	139	0,63%
15.					<i>být si jistý</i>	111	0,50%
16.					<i>uvědomovat si</i>	93	0,42%
17.					<i>bůhví</i>	79	0,36%

<sup>13</sup> Como ocurre con el término ‘pluma’ discutido anteriormente.

	DUBSKÝ, REJZEK	CHALUPA	INTERCORP Nº DE APARICIONES	FRECUENCIA RELATIVA	
18.			<i>vědomí</i>	67	0,30%
19.			<i>uvědomit si</i>	65	0,30%
20.			<i>chutnat</i>	51	0,23%
21.			<i>znalost</i>	41	0,19%
22.			<i>rozeznat</i>	38	0,17%
23.			<i>uhodnout</i>	19	0,09%
24.			<i>vědomost</i>	14	0,06%
25.			<i>kdoví</i>	11	0,05%
26.			<i>napadnout</i>	7	0,03%
27.			<i>(ne)mít páru</i>	6	0,03%
28.			<i>nevědomky</i>	5	0,02%
29.			<i>vonět</i>	3	0,01%
30.			<i>a sice, totiž</i>	3	0,01%
31.			<i>páchnout</i>	2	0,01%
32.			<i>připomenout</i>	1	0,00%

(Nº total de apariciones: 22024)

El término ‘saber’ en el diccionario de Dubský y Rejzek está recogido en dos entradas léxicas señalizadas en la tabla a través de números romanos. La primera corresponde al verbo mismo y la segunda tiene valor de sustantivo que se ha derivado a través de la sustantivación del infinitivo. En la primera entrada se incluyen sólo dos acepciones diferentes, una relacionada con actividades mentales humanas (presentando equivalentes como *vědět*, *znát*, *umět*, *dověst*, *rozumět*, *dozvědět se*, etc.) y la otra con las percepciones sensoriales (con equivalentes como *páchnout*, *zapáchat*, *vonět*, *chutnat*, *smrdět*, *nelíbit se*). Asimismo, el diccionario de Chalupa representa el término ‘saber’ por dos entradas diferentes en función de las clases de palabras, aunque a diferencia del primer diccionario los significados de la primera entrada asociados con las actividades mentales están distribuidos entre más acepciones (en la tabla les corresponden las líneas 1, 2, 3, 4, 6, 8, 9 y 10). Los dos diccionarios registran el término como parte de varias frases hechas, modismos o marcadores pragmáticos a los que no prestaremos mucha atención<sup>14</sup>. Si analizamos las ocurrencias de cada equivalente en el corpus Intercorp, comprobaremos que en este caso los dos diccionarios satisfacen mejor (que en los casos anteriores) el criterio del ordenamiento según la frecuencia. El equivalente (*vědět*) más frecuente (61,45%) está mencionado en primer lugar en los dos diccionarios. Si agrupamos todos los equivalentes similares a *vědět*<sup>15</sup> en una acepción

<sup>14</sup> Los modismos y marcadores que se estudian son, por ejemplo, ‘no sabe Dios dónde’ (*bůhvíkde*), ‘sepa Dios’ (*bůhví*), ‘vete a saber’ (*kdoví*), ‘a saber’ (*kdoví*, *a sice*, *totiž*), ‘no saber por dónde se anda’ (*nemít ani páru*), etc.

<sup>15</sup> *vědět*, *dozvědět se*, *dovědět se*, *znát*, *zjistit*, *pochopit*, *chápat*, *zajímat*, *rozumět*, *uvědomovat si*, *uvědomit si*, *rozeznat*, *uhodnout*, *napadnout*, *připomenout*, *být si jistý*, *umět*, *dokázat*, *dověst*, *nemít ani páru*, *nevědomky*, *bůhví*, *kdoví*, *a sice*, *totiž*.

como en el primer diccionario, llegaremos incluso a la frecuencia del 87,64% de todas las ocurrencias del término. Partiendo del segundo diccionario observamos que la segunda acepción (*znát*) es realmente el tercer equivalente más frecuente en el corpus (3,98%) y en segundo lugar están los significados *umět, dovést* o *dokázat* (8,57%) que el diccionario de Chalupa menciona en cuarto lugar (o en tercer lugar si incluimos las definiciones parafrásicas como *mít zprávu, informace* en el equivalente *vědět*). Los equivalentes del tipo *chutnat, vonět* o *páchnout* son bastante insignificantes en cuanto a su frecuencia en el corpus (0,25%). Para la función sustantiva (*vědění, vzdělání, vědomost, znalost, vědomí*) el corpus registra 1,19% de todas las ocurrencias, lo cual es más significativo que en el caso de los equivalentes del tipo *chutnat, vonět* o *páchnout*. Añadamos que en los diccionarios aparecen algunos equivalentes que no encontramos en el corpus: *smrdět, vzdělání*; y, al revés, en el corpus aparecen significados no registrados por los dos diccionarios: *zjistit, tušit, pochopit, chápat, zajímat, uvědomit si, uvědomovat si, rozeznat, uhadnout, napadnout, připomenout*; aunque conviene admitir que un diccionario no puede recoger todas las posibles variantes de traducción, además, si éstas mantienen ciertos rasgos comunes con otros significados (en este caso *vědět*).

## 5. Conclusión

En este artículo hemos estudiado cómo se representan los homónimos o polisemos en dos de los diccionarios español-checos más extensos e importantes de la actualidad. Para los términos 'banco', 'cuarto' y 'saber' planteamos las preguntas de si la microestructura de los diccionarios respeta las frecuencias de los equivalentes en el corpus paralelo checo-español InterCorp y si los equivalentes registrados en el corpus también están recogidos en los diccionarios o viceversa. Existen divergencias en la microestructura de los dos diccionarios; las diferencias más notables son:

- 1) El diccionario de Dubský y Rejzek es más extenso que el diccionario de Chalupa, aunque en los términos estudiados el segundo que contiene menos entradas y, por lo consiguiente, menos acepciones en total presenta algunos equivalentes de los que el primero no habla (por ejemplo, en término 'cuarto' *čtvrťfinále* o la denominación elíptica *čtvrté patro*). Considerando la extensión de las dos obras no es sorprendente que el primero mencione más acepciones que el segundo (por ejemplo, en el término, 'banco', *píščina, hoblice*; en la palabra 'cuarto', *předeck* o *zadek zvířete* y en el término 'saber', *vonět, páchnout, smrdět*).
- 2) El primer diccionario tiende a agrupar los significados similares en una acepción más compleja, mientras que el segundo prefiere distribuirlos entre varias acepciones. La ventaja del primer enfoque reside en la presunta rentabilidad, es decir, que el diccionario produce la sensación de registrar entradas léxicas no tan extensas, mientras que la ventaja del segundo enfoque, el cual preferiríamos, consiste en una mayor objetividad, ya que agrupar significados en una sola acepción supone contar con cierta similitud entre ellos que normalmente viene definida por la subjetividad del lexicógrafo.

Igual que un diccionario bilingüe nunca puede ser una obra completa que contenga todo el vocabulario de una lengua, el corpus InterCorp tampoco lo es. Se dan casos cuan-

do una entrada recogida en los diccionarios no aparece en el corpus<sup>16</sup> y, al revés, (que es la situación predominante) hay muchos términos que salen del análisis de equivalencias en el corpus, pero no los recogen los diccionarios (basta comparar la tercera columna de las tablas anteriores con las primeras dos columnas). Por consiguiente, basándonos en el estudio de este artículo, podemos juzgar que el corpus paralelo InterCorp es una obra más completa que los diccionarios, aunque es posible reprocharle que los resultados del análisis de equivalencias dependen del estilo y de la calidad de las traducciones, que son dos factores que van a costa de la objetividad de cualquier trabajo realizado con este corpus paralelo.

---

### BIBLIOGRAFÍA

- Dubský, J. – Rejzek, V. (2000): *Španělsko-český slovník, diccionario español-checo*. Praha: Leda.
- Čermák, F. – Blatná, R. (eds.) (1995): *Manuál lexikografie*. Praha: H+H.
- Čermák, F. (2004): Dvoujazyčný slovník a korpus: některé poznámky. Ústav Českého národního korpusu. FF UK. [en línea] [citado 2011-29-09] Accesible de la WWW: <<http://ucnk.ff.cuni.cz/doc/korplex.rtf>>
- Český národní korpus – InterCorp. Ústav Českého národního korpusu FF UK, Praha. [en línea] [citado 2011-20-09]. Accesible de la WWW: <<http://www.korpus.cz/InterCorp>>
- Chalupa, J. (eds.) (2008): *Španělsko-český, česko-španělský slovník*. Praha: Fin Lexika.
- Petrík, D. (2011): Volba lexikálních prostředků ve španělštině ze sémanticko-stylistické perspektivy. Rigorózní práce, FF UK Praha.
- Pottier, B. (1970): Hacia una semántica moderna. *Lingüística moderna y Filología hispánica*. Madrid: Gredos.

*Daniel Petrík*  
*Instituto de Estudios Románicos*  
*Universidad Carolina, Facultad de Filosofía y Letras*  
*nám. Jana Palacha 2, 116 38 Praha 1*  
*dapet@email.cz*

---

<sup>16</sup> Véase la página anterior, en el término ‘saber’ los equivalentes *smdět* o *vzdělání* no aparecen en el corpus.