

**El análisis de corpus es una
ceremonia superflua y un
total malgasto de tiempo y de
dinero del gobierno**

SVEN TARP

CENTRO DE
LEXICOGRAFÍA

UNIVERSIDAD DE
AARHUS

DINAMARCA

LUCHA DE IDEAS

- **“El análisis de corpus es una ceremonia superflua y un total malgasto de tiempo y de dinero del gobierno”**
- **El título de esta ponencia no representa mi opinión personal y tampoco la de los colaboradores del Centro de Lexicografía de Aarhus.**
- **El título es una combinación de dos citas de Lees (1962) y Itkonen (1976).**
- **La lucha de ideas entre estas posiciones y las de los defensores de los corpus fue un tema importante en la segunda parte de los años 70 del siglo pasado.**

TOTAL MALGASTO DE TIEMPO Y DINERO

- Para defender la idea de que el análisis de corpus es un total malgasto de tiempo y de dinero del gobierno, Lees dijo, entre otras cosas:
- “Usted es un hablante nativo del inglés, y en 10 minutos podrá producir más ilustraciones de cualquier punto de la gramática inglesa que lo que encontrará en muchos millones de palabras de texto aleatorio.”

(Robert Lees 1962)

UN GRANO DE VERDAD

- Hoy es evidente que los corpus son de gran utilidad en las investigaciones relacionadas con la lingüística, y también, con la lexicografía.
- No obstante, defenderé la idea de que hay un **grano de verdad** en las muy categóricas opiniones de Lees y Itkonen...
- ... por lo menos en lo que se refiere a la **lexicografía especializada**.

LIMITACIONES DEL USO DE CORPUS

- Dentro de la lexicografía especializada ya no se trata de luchar contra los molinos de viento sino **determinar las limitaciones del uso de corpus** para la confección de diccionarios especializados.
- Mi tesis es que la llamada “**revolución de corpus**” dentro de la lexicografía – es decir, el uso de corpus para la confección de diccionarios – ha llegado **demasiado lejos** y hasta **resulta contraproducente** en la realización de varios tipos de tareas.

EXTRAÑAS FORMAS DE LEXICOGRAFÍA

- Nuevos términos introducidos en las últimas décadas, quizás sin reflexionar suficiente sobre su significado real:
- “lexicografía de corpus”
- “lexicografía dirigida por corpus”
- “lexicografía basada en corpus”

LEXICOGRAFÍA DE CORPUS

- El término “**lexicografía de corpus**” no tiene sentido
- De esta manera, una determinada base empírica se convierte en una forma específica de lexicografía:
- No existe, p.ej.:
 - “lexicografía de **prueba**”,
 - “lexicografía de **archivo**”,
 - “lexicografía de **cuestionario**”
 - “lexicografía de **log files**”
 - (...)

LEXICOGRAFÍA DIRIGIDA POR CORPUS

- El término “**lexicografía dirigida por corpus**” no tiene sentido
- El corpus es un ente pasivo y no puede, por sí solo, determinar lo que se incluye en un diccionario. Debe haber criterios lexicográficos de selección y presentación.
- Ejemplos:
 - Descripción, prescripción, y proscripción (recomendación)
 - **Martes** en malgache
 - El número **67**

LEXICOGRAFÍA DIRIGIDA POR CORPUS

	GWDS ¹⁺²⁺³			google
-beinig	+	+	+	
einbeinig	+	+	+	2.115
zweibeinig	+	+	+	2.039
dreibeinig	+	+	+	1.158
vierbeinig	+	+	+	6.277
fünfbeinig	-	-	-	32
sechsheinig	+	+	+	578
siebenbeinig	-	-	-	12
achtbeinig	-	-	-	421
neunbeinig	-	-	-	1
zehnbeinig	-	-	-	10
elfbeinig	-	-	-	0
zwölfbeinig	-	-	-	9

LEXICOGRAFÍA BASADA EN CORPUS

- El término “**lexicografía basada en corpus**” también es problemático
- Es evidente que en la producción de un moderno diccionario digital hay tareas que – con gran beneficio para el lexicógrafo – pueden realizarse usando los corpus ...
- ... pero también hay muchas tareas que no tienen nada que ver con un corpus ...
- ... y hay otras que solo se basan parcialmente en un corpus.

PESCA DE TÉRMINOS Y DEFINICIONES

- A continuación, discutiré el conocido método terminológico de “**pesca**” de términos y definiciones en varios tipos de corpus.
- Mostraré como este método muy a menudo conduce a soluciones lexicográficas **insuficientes, incorrectas** y hasta **peligrosas**, por no hablar del **malgasto de tiempo y dinero**.
- Sobre esta base, y sin negar el uso de corpus para otras tareas, propondré **otros métodos** para la selección de términos y definiciones en diccionarios especializados.

DICCIONARIO DE CÍTRICOS

- Experiencia propia en una conferencia de lexicografía en 1994:
- **Proyecto de diccionario de cítricos**
- Dos autores
- ... llevaban 9 años trabajando en el proyecto,
- ... tenían unos 800 términos seleccionados,
- ... pero todavía faltaban la mayoría de las definiciones

TUVO LUGAR LA SIGUIENTE CONVERSACIÓN:

- **Pregunta:** Ustedes, ¿qué formación tienen?
- **Respuesta:** Somos agrónomos con especialidad en los cítricos.
- **Pregunta:** ¿Conocen el significado de los términos?
- **Respuesta:** Sí.
- **Pregunta:** Entonces, ¿por qué no escriben ustedes mismos las definiciones?
- **Respuesta:** No está permitido escribirlas. Según las reglas tenemos que encontrarlas en un corpus como documentación.

EURODICAUTOM

- Experiencia propia en la década de los 90:
- **Subasta de términos para EuroDicAutom**
(hoy IATE – Terminología Interactiva para Europa)
- Pagaban 50 euros por término por lengua
(probablemente el proyecto lexicográfico más caro jamás)
- Nos adjudicaron unos 500 términos relacionados
con **la biología molecular**

EURODICAUTOM

- Las definiciones las escribió un **experto de biología molecular**, autor de varios diccionarios sobre el tema y con quien colaboramos en aquel entonces
- Este experto fue duramente criticado por no indicar sus fuentes
- **Solución:** Puso su propio nombre como referencia detrás de todas las definiciones
- Ya todo estaba en orden ...
- ... **pero nosotros nos quedamos con la boca abierta**

DICCIONARIOS DE CONTABILIDAD

Deemed cost

Deemed cost is an amount used instead of cost or depreciated cost at a specific date. Any following amortisation or depreciation is made on the assumption that the enterprise initially recognised the asset or liability at a cost equal to the deemed cost.

(Diccionarios de Contabilidad)

‘Deemed cost’ is a surrogate for cost at a given date. For example if a building is purchased at \$100000 this is cost and also the deemed cost at that given date, ...

(Definición obtenida por Kilgarriff 2012 mediante Google)

DICCIONARIOS DE CONTABILIDAD

- “Las entradas enciclopédicas precisas y extensivas existen muy a menudo de antemano y es muy fácil accederlas mediante Google, como en este caso. Se podría discutir el valor aportado por los lexicógrafos.”

(Kilgarriff 2012: 27)

- Para cualquiera persona **que no sea experto** de la contabilidad puede resultar muy difícil:
 - ver la **diferencia** entre las dos definiciones, la de los Diccionarios de Contabilidad y la del Internet,
 - y juzgar cuál de las dos es **la mejor**.

DICCIONARIOS DE CONTABILIDAD

- La definición encontrada en el Internet mediante Google puede ser correcta en un contexto específico,
- pero no sirve como definición del término tratado en los Diccionarios de Contabilidad, en los que ha sido definido de acuerdo con las **Normas Internacionales de Contabilidad (IFRS)** vigentes, y en beneficio del usuario específico previsto.
- La definición obtenida por Kilgarriff es **demasiado amplia y cubre mucho más** que las normas IFRS, relevantes para los Diccionarios de Contabilidad.

OXFORD DICTIONARY OF ECONOMICS

A **corner solution** is a special solution to an agent's maximization problem in which the quantity of one of the arguments in the maximized function is zero. The more usual solution will lie in the non-zero interior at the point of tangency between the objective function and the constraint. ... ([www.wordvia.com/dictionary/corner solution](http://www.wordvia.com/dictionary/corner%20solution))

- Esta definición de **corner solution**, encontrada en el Internet, es similar a la de la 2^{da} edición del **Oxford Dictionary of Economics**:

corner solution

A solution to a system of equations where some variables are zero...

OXFORD DICTIONARY OF ECONOMICS

- Comentario de Hashimzade, Myles & Myles (2014), **economistas y autores** de la 4^{ta} edición del **Oxford Dictionary of Economics**:
- “Como definición general de **corner solution**, ambas definiciones son incorrectas. Podrían ser correctas si estuvieran en un contexto apropiado, pero sin tal contexto, las definiciones son **incorrectas y engañosas.**”

OXFORD DICTIONARY OF ECONOMICS

corner solution

In the context of a constrained optimization problem this is a solution that does not change in at least one direction in response to any arbitrarily small perturbation to the gradient of the objective function at the optimum.

(artículo del Oxford Dictionary of Economics, 4^{ta} edición)

- ¡Solo un **experto** de la disciplina – en este caso un economista – puede determinar si una definición de este tipo es correcta o incorrecta!

TÉRMINOS EN DICCIONARIOS GENERALES

- El problema no solo se produce en los diccionarios altamente especializados, sino también en los diccionarios generales.
- P.ej. tanto el **Oxford Advanced Learner's Dictionary of Current English** como el **Collins COBUILD Advanced Dictionary of English** contienen términos jurídicos centrales ...
- ... y ambos usan en su publicidad el argumento de que se basan en grandes corpus autoritativos como el **British National Corpus**, el **Oxford English Corpus**, y el **Bank of English**.

TÉRMINOS EN DICCIONARIOS GENERALES

- No obstante, ninguno de estos dos diccionarios contiene los términos de derecho civil introducidos en la **Ley británica** en 1999.
- Esta Ley, p.ej., introdujo los términos:
 - **claim form** en vez de **statement of case**,
 - **writ of summons** en vez de **pleading**.
- Ambos diccionarios solo contienen los antiguos términos más de una década después de su sustitución por otros términos nuevos.

DICCIONARIO DE MÚSICA

- Para el **Musikordbogen** (Diccionario de Música) no se utilizó ningún tipo de corpus para seleccionar las entradas (lemas) y escribir las definiciones.
- La autora – quien es experta de música – hizo la selección basándose en los lemas en diccionarios ya publicados además de índices en manuales y libros de texto relacionados con la música.
- Las definiciones tampoco fueron extraídas de un corpus sino escritas por esta misma experta de música.

DICCIONARIO DE INGENIERÍA GENÉTICA

- Para el **Diccionario Enciclopédico de Ingeniería Genética** no se utilizó un corpus para escribir la siguiente definición de **gen**:

gen

Un gen es una secuencia de ADN que codifica una proteína, el ARNt o el ARNr. En el caso de los eucariotas podemos ser más precisos y definir un gen como una secuencia de ADN transcrito más secuencias reguladoras asociadas. En los procariotas, a menudo se codifica dos o más proteínas en la misma unidad de transcripción y esa unidad de transcripción más sus secuencias reguladoras asociadas reciben el nombre de operón.

DICCIONARIOS EN LÍNEA DE DANÉS

- Sin embargo, en un diccionario general para legos se ofrecen los siguientes elementos de significado para el lema **gen**, de los que solo el segundo se basa en un análisis de corpus:
 1. propensión hereditaria relacionada con las cromosomas en las células sexuales
 2. tendencia o disposición a hacer algo determinado.

CONCLUSIONES

- Básicamente, hay **dos métodos totalmente diferentes** para la selección de términos y la preparación de definiciones en diccionarios especializados, y también en los generales:
 1. **"Pesca"** de términos y definiciones en los corpus
 2. Dejar el trabajo a un **experto** en los diccionarios especializados y consultar a uno en los generales.

CONCLUSIONES

- No es suficiente consultar un corpus para realizar estas tareas:
 - Solo un experto puede seleccionar los términos correctos y relevantes.
 - Las definiciones encontradas en los corpus pueden ser **anticuadas, incorrectas o irrelevantes** para la disciplina en cuestión.
- No basta componer un corpus muy especializado y actualizado, pues las definiciones encontradas a menudo son **contextuales** y, por lo tanto, insuficientes.

DESESTIMACIÓN DE LOS EXPERTOS

- Algunos terminólogos dicen que no hace falta incluir a expertos en sus proyectos alegando que ellos mismos pueden convertirse en expertos durante el proyecto.
- Es una **desestimación total** de los expertos y del esfuerzo que requiere obtener suficientes conocimientos para poder calificarse como experto.
- Claro: es bien sabido que algunos proyectos terminológicos demoran 10, 15 o más años en terminarse...

RESUMEN

- “El análisis de corpus es una ceremonia superflua y un total malgasto de tiempo y de dinero del gobierno”
- Es evidente que no es así, ya que muchos proyectos lexicográficos requieren un corpus como base empírica ...
- ... **pero hay un grano de verdad**, pues los resultados del análisis de corpus relacionados con algunos proyectos y tareas son engañosos y peligrosos.
- En estos casos, el corpus no puede sustituir los conocimientos especializados de un experto.