

Collaborative Multilingual Dictionaries

Christian M. Meyer



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Lexicografía multilingüe en rede (MULTILEX).
October 20–21, 2015. Santiago de Compostela, Spain.



Collaborative Multilingual Dictionaries

Typically, dictionaries
are the product of few
expert lexicographers

Collaborative Multilingual Dictionaries

Dictionaries
created by users

Most dictionaries
cover one or
two languages

Collaborative **Multilingual** Dictionaries

Dictionaries
created by users

Dictionaries for
> 2 languages

Collaborative Multilingual Dictionaries

Dictionaries
created by users

Dictionaries for
> 2 languages

- 1. User participation**
- 2. Quality of user contributions**
- 3. Multilingual Structures**

User Participation

Involving users in the dictionary-making process is not new

- Oxford English Dictionary started already in the 19th century (cf. Thier, 2014)

But with the advent of the World Wide Web, user participation became

- possible at a large scale
- easier
- faster
- more diverse

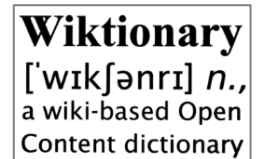


1. Direct User Contributions

Creation, modification, and deletion of dictionary articles by users

• Contributions to open-collaborative dictionaries

- Completely built by dictionary users
- No editorial control



• Contributions to collaborative-institutional dictionaries

- Often no modification of existing articles
- Reuse in other publisher-owned products



• Contributions to semi-collaborative dictionaries

- Checked by editorial staff
- Typically no extensive revision



(Abel/Meyer, 2013)

2. Indirect User Contributions

Feedback from users to dictionary makers

- **Explicit Feedback**

- Suggestions for articles or the dictionary as a whole
- Additions (e.g., quotations, usage examples)
- Corrections or quality flaws



- **Implicit Feedback**

- Users are often not aware of this type of feedback
- Log file analysis documenting dictionary use
- Integrating external content (e.g., from Flickr)



(Abel/Meyer, 2013)

3. Accessory User Contributions

Exchange beyond the dictionary content

- **Exchange between dictionary makers and users**

- Blogs, newsletter, social networks, etc.
- Dictionary games and usage guides
- Language consultation services

MACMILLAN
DICTIONARY

ANW | Algemeen
Nederlands
Woordenboek

canoonet

- **Exchange between dictionary users themselves**

- Forums
- Discussion pages
- User comments

Wiktionary
[ˈwɪkʃənɹɪ] n.,
a wiki-based Open
Content dictionary

 LEO

wordnik

(Abel/Meyer, 2013)

Types of User Contributions

1. Direct user contributions

Contributions to

- open-collaborative dictionaries
- collaborative-institutional dictionaries
- semi-collaborative dictionaries

2. Indirect user contributions

- Explicit feedback
- Implicit feedback

3. Accessory user contributions

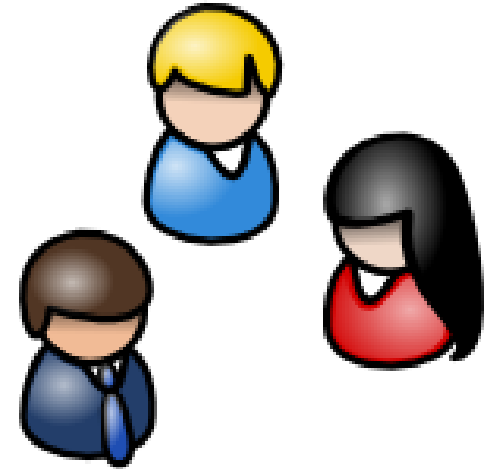
- Exchange between dictionary makers and dictionary users
- Exchange among dictionary users

(Abel/Meyer, 2013)

Multilingual Dictionaries

Users can...

- add/suggest translations
- revise (modify/delete) translations
- assume responsibility for a certain language pair, language variety, or domain
- check the quality of entries (collaborative quality assessment)
- comment the dictionary as a whole, give implicit and explicit feedback
- answer language-related questions beyond the dictionary (e.g., in a forum)
- ...



Quality of User Contributions



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Strengths

- Rapid growth
- Language varieties
- Many language (pairs)
- Subjective language intuitions
- Open licenses
- Quality assessment

Weaknesses

- Inconsistency
- Incorrect, unspecific, old-fashioned descriptions
- Plagiarism
- “Complicated” articles missing
- Corpus evidence



(Meyer, 2013; Meyer/Gurevych, 2014)

Quality of User Contributions



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Strengths

- Rapid growth
- Language varieties
- Many language (pairs)
- Subjective language intuitions
- Open licenses
- Quality assessment

Weaknesses

- Inconsistency
- Incorrectly transcribed, old-fashioned descriptions
- Plagiarism
- “Connected” and missing
- Corpus evidence

***How to steer user contributions
to get the best ones only?***

(Meyer, 2013; Meyer/Gurevych, 2014)

Collaborative Quality Assessment



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Is this translation correct and useful?	
English	loodering [Scot.] [Irish]
German	heftige Prügelei {f}
Class	noun
Comment	• http://www.oed.com/view/Entry/... " 'loodering n. a severe beating; a hiding. 1866 W. Gregor Dial. Banffshire (Philol. Soc.) 227 Llouthera severe beating. 1912 J. Campbell Judgment i. 9 She'll get over it. It's not the 1985 L. Shannon in Stories (Boston) No. 12 14 Whenever s biggest loodering she's ever had in her whole life. "
	YES (100% correct) NO / MAYBE ...

(Meyer/Gurevych, 2014)

Collaborative Multilingual Dictionaries

Dictionaries
created by users

Dictionaries for
> 2 languages

1. User participation
2. Quality of user contributions
- 3. Multilingual Structures**

LEO Dictionary Portal: Example



TECHNISCHE
UNIVERSITÄT
DARMSTADT

DE
▼
EN

▶ defender <i>also</i> [law]	☐ ▶ der Verteidiger die Verteidigerin <i>Pl.: die Verteidiger, die Verteidigerinnen</i>
▶ apologist	☐ ▶ der Verteidiger die Verteidigerin <i>Pl.: die Verteidiger, die Verteidigerinnen</i>
▶ pleader	☐ ▶ der Verteidiger <i>Pl.: die Verteidiger</i>
▶ solicitor (<i>Brit.</i>)	☐ ▶ der Verteidiger <i>Pl.: die Verteidiger</i>
▶ vindicator	☐ ▶ der Verteidiger <i>Pl.: die Verteidiger</i>
▶ back [sport.]	☐ ▶ der Verteidiger <i>Pl.: die Verteidiger</i> [<i>Fußball</i>]
defence ^{BE} counsel [law]	☐ ▶ der Verteidiger <i>Pl.: die Verteidiger</i> - <i>im Strafprozess</i>

DE
▼
ES

▶ el defensor la defensora	☐ ▶ der Verteidiger die Verteidigerin <i>Pl.: die Verteidiger, die Verteidigerinnen</i>
▶ el apologista la apologista	☐ ▶ der Verteidiger die Verteidigerin <i>Pl.: die Verteidiger, die Verteidigerinnen</i>
▶ abogado defensor [jur.]	☐ ▶ der Verteidiger die Verteidigerin <i>Pl.: die Verteidiger, die Verteidigerinnen</i>
▶ el líbero [dep.]	☐ ▶ der Libero <i>Pl.: die Liberos</i> <i>italienisch</i> [<i>Fußball</i>] - <i>freier Verteidiger</i>

LEO , <http://dict.leo.org>

Bilingual Translation Tables



TECHNISCHE
UNIVERSITÄT
DARMSTADT

DE
▼
EN

Verteidiger | Verteidigerin

defender [law]

Verteidiger | Verteidigerin

apologist

Verteidiger [Fußball]

back [sport.]

- Sense distinctions and usage comments?
- Translation shifts?
- Multilingual queries difficult (e.g., EN ► ES)

?

DE
▼
ES

Verteidiger | Verteidigerin

el defensor | la defensora

Verteidiger | Verteidigerin

el apologista | la apologista

Libero [Fußball]

el líbero [dep.]

LEO , <http://dict.leo.org>

Refugee Phrasebook

- Open-collaborative dictionary by the Open Knowledge Foundation
- Targeted at refugees and voluntary helpers



- Phrases from everyday life, medicine&healthcare, and law
- Available as public domain (Creative Commons CC0)
- > 500 phrases in > 30 languages since September 2015
- <http://www.refugeephrasebook.de/project/>

Refugee Phrasebook: Example



TECHNISCHE
UNIVERSITÄT
DARMSTADT

German	English	Syrian/Arabic alphabet	Spanish	Bangla	Urdu
Richter	judge	القاضي	El juez / La jueza	বিচারক	اب کو جج کی ضرورت ہے؟
Rückübernahmeabkommen	re-admission agreement (between two countries)	اتفاقية إعادة القبول (بين دولتين)	Acuerdo de readmisión	পুনরায় দেশে ঢোকার চুক্তি (দুই দেশের মধ্যে)	دوباره داخلے معاہدہ
subsidiärer Schutz nach EU Recht	subsidiary protection under EU law	الحماية الثانوية في قانون الاتحاد الأوروبي (أي الحماية التي تطال رعايا دول أخرى)	protección subsidiaria bajo legislación europea	ইউরোপিয়ান ইউনিয়নের সুরক্ষা আইন	
Übersetzung	Translation	الترجمة	La traducción	অনুবাদ	ترجمہ
Verhaftung	arrest/detention	التوقيف / الاعتقال	La detención / el arresto	গ্রেফতার / আটক	گرفتار
Visum	visa	تأشيرة دخول	Visa / visado	ভিসা	ویزا
Visaablehnung	refusal of visa application	رفض تأشيرة الدخول	Denegación de visado	ভিসার আবেদন নাকচ	ویزا کی درخواست سے انکار
Wiedererwägung	reconsideration	إعادة نظر	La reconsideración	পুনরায় আবেদনপত্র যাচাই।	
Pflichtverteidiger	assigned counsel/duty counsel	محامي بالمجان	Abogado de oficio	নির্ধারিত/ডিউটিতে থাকা পরামর্শক	
Verteidiger	counsel for defence	محامي الدفاع	Abogado defensor	বিবাদীর উকিল	
Freispruch	acquittal	الحكم بالإبراء	Verdicto absolutorio	দোষ থেকে মুক্তি	بری

Multilingual Translation Table

German	English	Syrian/Arabic alphabet	Spanish	Bangla	Urdu
Richter	Richter	judge	القاضي	El juez / La jueza	اب کو جج
Rückübernahmeabkommen	re-admission agreement (between two countries)	اتفاقية إعادة القبول (بين دولتين)	Acuerdo de readmisión	পুনরায় দেশে ঢোকার চুক্তি (দুই দেশের মধ্যে)	دوباره داخلے معاہدہ
subsidiärer Schutz nach EU Recht	subsidiary protection under EU law	الحماية الثانوية في قانون الاتحاد الأوروبي (أي الحماية التي تطال زعماء دول أخرى)	protección subsidiaria	ইউরোপিয়ান ইউনিয়নের সুরক্ষা আইন	
Übersetzung	Visum	visa	تأشيرة دخول	Visa / visado	ترجمہ
Verhaftung	arrest/detention	التوقيف / الاعتقال	La detención / el arresto	গ্রেফতার / আটক	گرفتار
Visum	Verteidiger	counsel for defence	محامي الدفاع	Abogado defensor	ویزا
Visaablehnung	refusal of visa application	رفض تأشيرة	Denegación de visado	ভিসার আবেদন নাকচ	ویزا کی درخواست سے انکار
Wiedererwägung	reconsideration	إعادة النظر	reconsideración	পুনরায় বিবেচনা	
Pflichtverteidiger					
Verteidiger					
Freispruch					بری

- Multiple translations or variants (e.g., gender)?
- Translation shifts?
- Sense distinctions and usage comments?

Wiktionary: Example



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Verteidiger (Deutsch) [Bearbeiten]

Bedeutungen:

- [1] *Sport, Fußball*: defensiv eingestellter Spieler, dessen Aufgabe es ist, das gegnerische Spiel zu stören, um anschließend das eigene Spiel wieder aufbauen zu können
- [2] *Recht*: Rechtsanwalt, der vor Gericht die Interessen des Angeklagten vertritt
- [3] *Militär, auch übertragen in Spielen*: Gegenpart zum Angreifer oder Aggressor, in der Regel innerhalb der eigenen Grenzen
- [4] Person, die sich für etwas (zum Beispiel eine Idee) stark macht oder sich gegen etwas (zum Beispiel einen Angriff) wehrt

Übersetzungen [Bearbeiten]

- | | |
|---|--|
| <ul style="list-style-type: none">• Bulgarisch: [1] бранител (branitel) → <i>bg</i>• Englisch: [1] back → <i>en</i>; [2] defence → <i>en</i>• Französisch: [1] défenseur → <i>fr</i> <i>m</i>; [2] avocat de la défense → <i>fr</i> <i>m</i>• Isländisch: [1] verjandi → <i>is</i>• Italienisch: [1] difensore → <i>it</i> <i>m</i>• Katalanisch: [1] defensa → <i>ca</i>• Litauisch: [1] gynėjas → <i>lt</i> | <ul style="list-style-type: none">• Niederländisch: [1, 2] verdediger → <i>nl</i> <i>m, f</i>• Okzitanisch: [1] arrièr → <i>oc</i>• Polnisch: [1, 2] obrońca → <i>pl</i> <i>m</i>, [1] defensor → <i>pl</i> <i>m</i>, [2] adwokat → <i>pl</i> <i>m</i>• Russisch: [1, 2] защитник (zašitnik) → <i>ru</i> <i>m</i>• Schwedisch: [1] back → <i>sv</i> <i>u</i>; [2] försvarsadvokat → <i>sv</i> <i>u</i>• Spanisch: [1] defensa → <i>es</i> <i>m, f</i>, [2] abogado → <i>es</i> <i>m</i> |
|---|--|

<https://de.wiktionary.org/wiki/Verteidiger>

Fixed Source Language



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Verteidiger (Deutsch) [Bearbeiten]

Bedeutungen:

- [1] *Sport, Fußball*: defensiv eingestellter Spieler, dessen Aufgabe es ist, das gegnerische Spiel zu verhindern
- [2] *Recht*: Person, die vor Gericht die Interessen eines anderen vertritt
- [3] *Militär, auch übertragen in Spielen*: Gegenpart zum Angreifer, in der Regel mit Waffen ausgerüstet
- [4] *Person, die sich für etwas (zum Beispiel eine Idee) einsetzt*: Person, die gegen etwas (zum Beispiel eine Idee) wehrt

Verteidiger₁

Verteidiger₂

Verteidiger₃

Verteidiger₄

back

defensa

защитник

defence

abogado

Verteidigung

Wehr

?

Übersetzungen [Bearbeiten]

- Bulgarisch: [1] *branitel* → bg
- Englisch: [1] *back* → en; [2] *defence* → en
- Französisch: [1] *défense* → fr
- Isländisch: [1] *varnari* → is
- Italienisch: [1] *difensore* → it
- Katalanisch: [1] *defensor* → ca
- Litauisch: [1] *gynėjas* → lt

- Niederländisch: [1, 2] *verdediger* → nl m, f
- Okzitanisch: [1] *arrièr* → oc
- Spanisch: [1] *defensa* → es m, f, [2] *abogado* → es m

- Meaning of the translated word?
- Consistency?
- Effort?

<https://de.wiktionary.org/wiki/Verteidiger>

Wikipedia: Example



TECHNISCHE
UNIVERSITÄT
DARMSTADT

WIKIPEDIA
Die freie Enzyklopädie

Hauptseite
Themenportale
Von A bis Z
Zufälliger Artikel

Mitmachen

Artikel verbessern
Neuen Artikel anlegen
Autorenportal
Hilfe
Letzte Änderungen
Kontakt
Spenden

In anderen Sprachen

Čeština
English
Lietuvių
Norsk bokmål
Русский
Türkçe
Українська

Strafverteidiger (Deutschland)

Ein **Strafverteidiger** ist ein dem **Beschuldigten** in einem **Strafverfahren** zur Seite stehender Beistand. In der deutschen **Strafprozessordnung** (StPO) wird er als *Verteidiger* bezeichnet.

Das Recht eines jeden Beschuldigten auf Hinzuziehung eines Verteidiger ist in **§ 137 StPO** (*Recht des Beschuldigten auf Hinzuziehung eines Verteidigers*) normiert.

Inhaltsverzeichnis [\[Verbergen\]](#)

- 1 Organ der Rechtspflege
- 2 Person des Verteidigers
- 3 Wahl- oder Pflichtverteidigung
- 4 Verteidigungsstrategien
- 5 Vergütung des als Verteidiger tätigen Rechtsanwalts
- 6 Siehe auch
- 7 Literatur
- 8 Weblinks
- 9 Einzelnachweise

Organ der Rechtspflege [\[Bearbeiten\]](#)

Der Verteidiger ist neben der **Staatsanwaltschaft** und dem **Gericht** ein unabhängiges, selbständiges Organ.

https://de.wikipedia.org/wiki/Strafverteidiger_(Deutschland)

Interwiki Links



TECHNISCHE
UNIVERSITÄT
DARMSTADT

WIKIPEDIA
Die freie Enzyklopädie

Strafverteidiger (Deutschland)

Strafverteidiger (Deutschland)₁

Criminal defense lawyer₁

Защитник в уголовном процессе₁

Müdafî₁

Abwehrspieler₁

Defender (association football)₁

مدافع (كرة قدم)₁

Defensa (fútbol)₁

Защитник (футбол)₁

Defans₁

- Translation shift?
- Multiple translations?

OmegaWiki: Example



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Verteidiger

Sprache: Deutsch

► **Verteidiger** : One who supports something.

▼ **Verteidiger** : Someone who defends with conviction one thesis, ideal or plan.

▼ Definition

Sprache	Text
Englisch	Someone who defends with conviction one thesis, ideal or plan.
Italienisch	Persona che difende con convinzione una tesi, un ideale, un progetto.
Niederländisch	Iemand die met overtuiging een stelling, ideaal of plan verdedigd.
Spanisch	Persona que defiende con convicción una tesis, un ideal o un proyecto.

▼ Synonyme und Übersetzungen

Ausdruck		
Sprache	Schreibweise	Annotation
Arabisch	مدعم	Anzeigen ▼
Bokmål	forsvarer	Anzeigen ▼
Bulgarisch	защитник	Anzeigen ▼
Dänisch	forsvarer	Anzeigen ▼
Deutsch	Verfechter	Anzeigen ▼
Englisch	defender	Anzeigen ▼

<http://www.omegawiki.org/Expression:Verteidiger>

Multilingual Synsets



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Verteidiger

Sprache: Deutsch

Sprache	Text
Deutsch	Verteidiger ₁ , Befürworter ₁ , Vertreter ₁ , partidaria ₁
Deutsch	advocate ₁ , proponent ₁ , partidario ₁ , proponente ₁
Englisch	Verteidiger ₂ , Verfechter ₂ , defensor ₁ , defensora ₁
Spanisch	defender ₁ , supporter ₁ , защитник ₁

- Translation shift?
- Missing lexicalizations in a language?

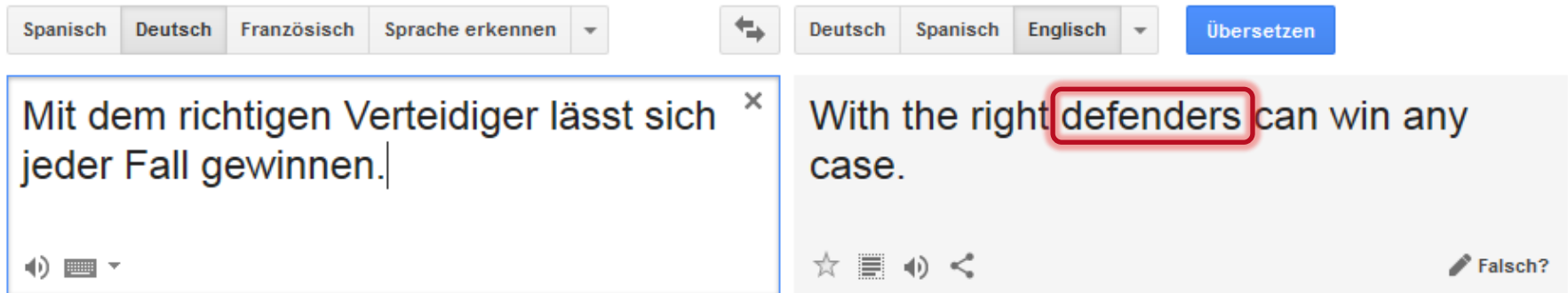
<http://www.omegawiki.org/Expression:Verteidiger>

Context-specific Organization?

- Organization types are usually static
- Either *defense lawyer* is a translation of *Verteidiger* or not
- **Can you imagine more flexible types of organization?**
- Users need to solve problems
- Ideally, a dictionary understand the problem and returns ONLY the relevant information

Context-specific Organization?

What some of my colleagues consider a “dictionary”:



Spanish Deutsch Französisch Sprache erkennen

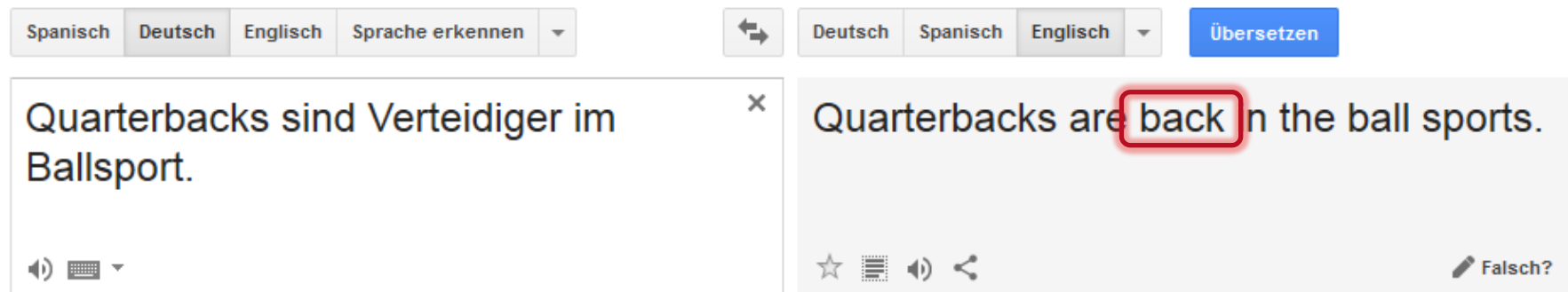
Deutsch Spanish Englisch Übersetzen

Mit dem richtigen Verteidiger lässt sich jeder Fall gewinnen.

With the right defenders can win any case.

☆ 📄 🔊 🔄 Falsch?

- At first sight: no descriptions; we stick with erroneous output...
- **But:** translations are chosen based on context!



Spanish Deutsch Englisch Sprache erkennen

Deutsch Spanish Englisch Übersetzen

Quarterbacks sind Verteidiger im Ballsport.

Quarterbacks are back in the ball sports.

☆ 📄 🔊 🔄 Falsch?

Context-specific Organization?



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Spanish Deutsch Französisch Sprache erkennen

Deutsch Spanish Englisch Übersetzen

Mit dem richtigen Verteidiger lässt sich jeder Fall gewinnen.

With the right defenders can win any case.

Falsch?

**Deliver dictionary knowledge
which exactly fits this (translation) context**

Take-Home Message



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Diverse forms of user participation

- Direct, indirect, accessory contributions
- Strengths and weaknesses

A variety of multilingual structures

- Bilingual and multilingual tables
- Fixed source language
- Interwiki links and multilingual synsets
- Context-specific organization
- **More flexible structures?**
- **Lexicographers–users cooperation?**



Further Readings



TECHNISCHE
UNIVERSITÄT
DARMSTADT

- Ch.M. Meyer/A. Abel: **User Contribution in the Era of the Internet**, chapter 27 in P.A. Fuertes Olivera (Ed.): *The Routledge Handbook of Lexicography*, London: Routledge, (in preparation).
- Ch.M. Meyer, I. Gurevych: **Methoden bei kollaborativen Wörterbüchern**, *Lexicographica* 30(1): 187–212, 2014.
- Ch.M. Meyer: **Wiktionary: The Metalexicographic and the Natural Language Processing Perspective** (= tprints 3654). Dissertation, Darmstadt: Technische Universität Darmstadt 2013.
<http://tprints.ulb.tu-darmstadt.de/3654/>
- A. Abel, Ch.M. Meyer: **The dynamics outside the paper: user contributions to online dictionaries**, in: *Proceedings of the 3rd eLex conference 'Electronic lexicography in the 21st century: thinking outside the paper'*, pp. 179–194, 2013. Tallinn, Estland.
- M. Matuschek, Ch.M. Meyer, I. Gurevych: **Multilingual Knowledge in Aligned Wiktionary and OmegaWiki for Translation Applications**, *Translation: Computation, Corpora, Cognition – Special Issue "Language Technology for a Multilingual Europe"* 3(1):87–118, 2013.
- I. Gurevych, J. Eckle-Kohler, S. Hartmann, M. Matuschek, Ch.M. Meyer, C. Wirth: **UBY – A Large-Scale Unified Lexical-Semantic Resource Based on LMF**, in: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 580–590, April 2012. Avignon, France.
- Ch.M. Meyer, I. Gurevych: **Wiktionary: a new rival for expert-built lexicons? Exploring the possibilities of collaborative lexicography**. chapter 13 in S. Granger, M. Paquot (Hrsg.): *Electronic Lexicography*, pp. 259–291, Oxford: Oxford University Press, November 2012.



Kontakt / Contact

Christian M. Meyer

Technische Universität Darmstadt

Ubiquitous Knowledge Processing Lab

 Hochschulstr. 10, 64289 Darmstadt, Germany

 +49 (0)6151 16–5386

 +49 (0)6151 16–5455

 meyer (at) ukp.informatik.tu-darmstadt.de

Rechtliche Hinweise

Die Folien sind für den persönlichen Gebrauch der Vortragsteilnehmer gedacht. Im Vortrag verwendete Photographien, Illustrationen, Wort- und Bildmarken sind Eigentum der jeweiligen Rechteinhaber oder Lizenzgeber. Um Missverständnisse zu vermeiden, wäre eine kurze Kontaktaufnahme vor Weitergabe oder -nutzung der Vortragsmaterialien empfehlenswert. Sofern Sie Ihre Rechte verletzt sehen, bitte ich ebenfalls um Kontaktaufnahme zur Klärung der Sachlage.

Legal Issues

The slides are intended for personal use by the audience of the talk. Photographies, illustrations, trademarks, or logos are property of the holder of rights. To avoid any misconceptions, I would strongly recommend to get in touch before reusing or redistributing the slides or any additional material of the talk. The same applies if you consider your rights infringed – please let me know to initiate further clarification.