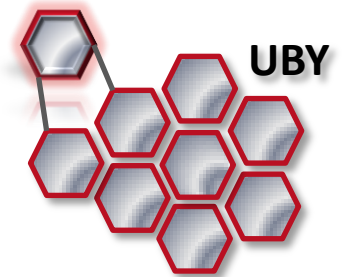


Automatische Vernetzung von Verben mit kontrastiven Valenzstrukturen in UBY



TECHNISCHE
UNIVERSITÄT
DARMSTADT

Judith Eckle-Kohler und Christian M. Meyer



3. Sektionentagung der Gesellschaft für Angewandte Linguistik (GAL),
Sektion Lexikographie „Valenz und Kollokation im (digitalen) Wörterbuch“,
Frankfurt (Oder), 23.–25. September 2015.

Englischlerner mit Deutsch als Muttersprache

Deutsch: *genießen* mit zu-Infinitiv:
Sie genießen es, unabhängig *zu sein*.

Englisch: *enjoy* mit *to*-Infinitiv oder *ing*-Form?
They enjoy *to be* independent. ?
They enjoy *being* independent. ?

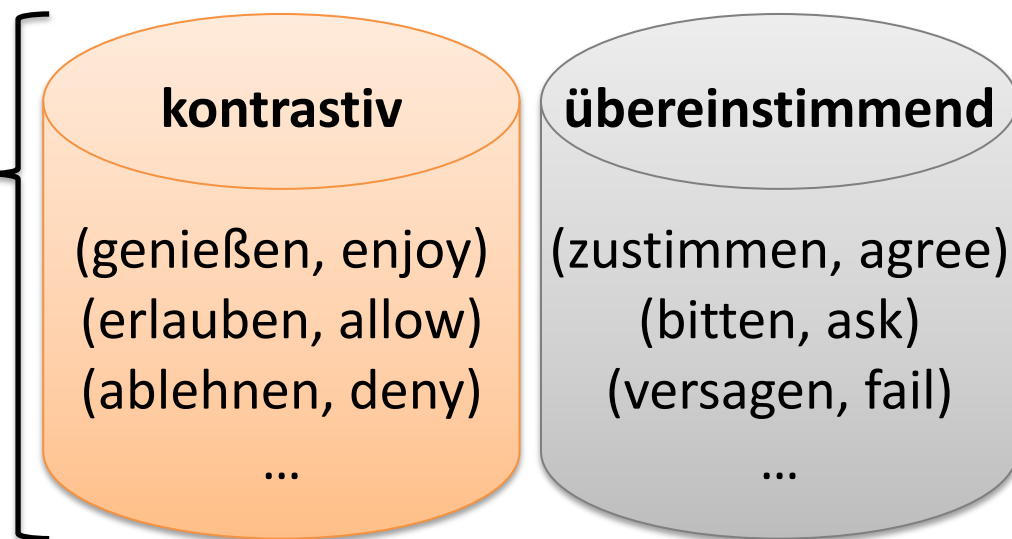


Listen von deutschen und englischen Verben mit...

- kontrastiven Valenzstrukturen
- übereinstimmenden Valenzstrukturen

Beispiele für Kontraste:

- Nicht-finite Verbergänzung
- reflexive Verben
- *da/es*-Korrelat im Hauptsatz
- Präposition von P-Objekten
- ...



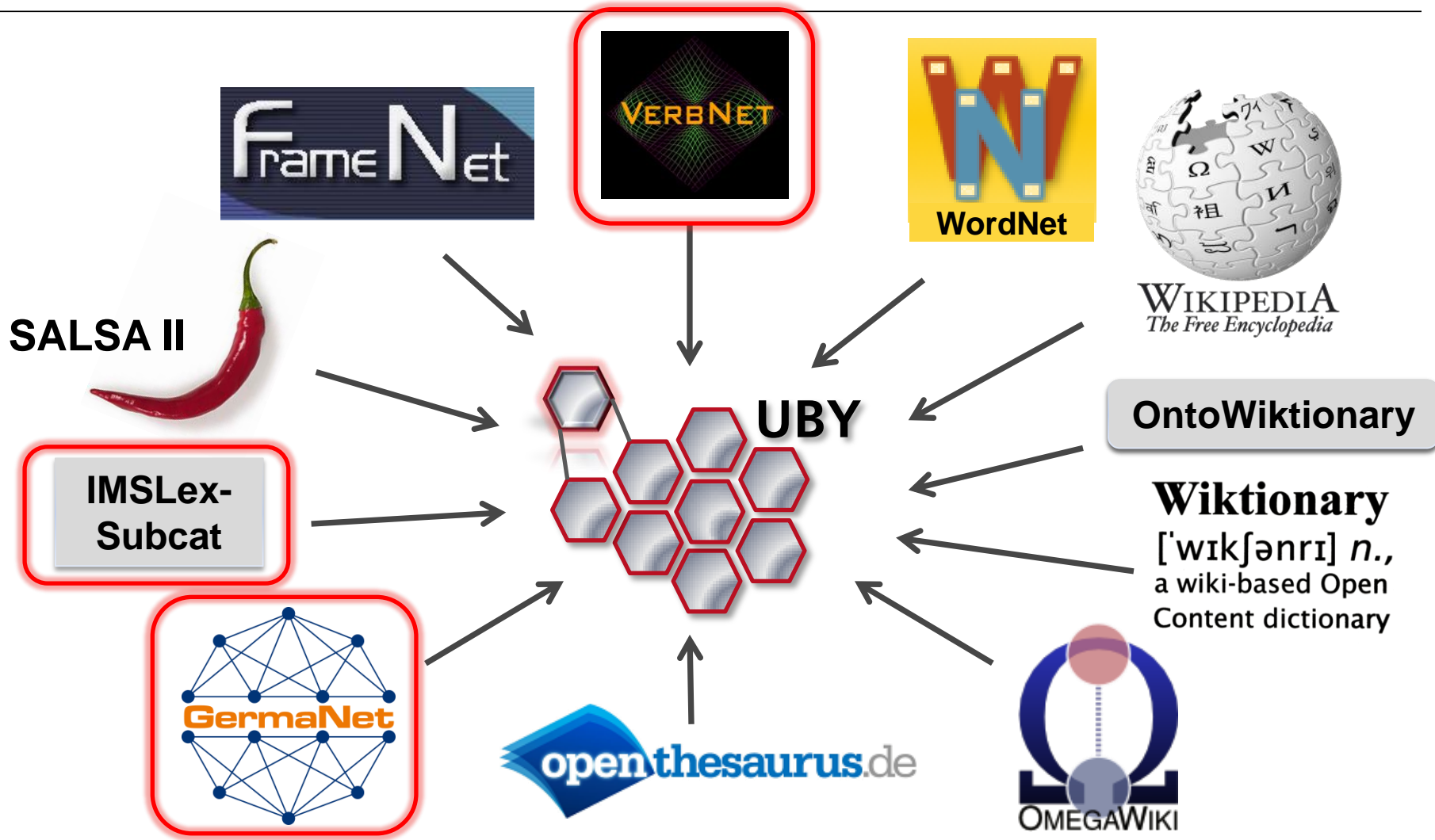
Bestehende Arbeiten:

- Sprachspezifische Valenzwörterbücher
- Manuelle Analyse von Kontrasten, z.B. Fischer (1997)

Unser Ansatz:

- Standardisierte Repräsentation für Valenzstrukturen
- Erkennen von ähnlichen Valenzstrukturen über Sprach- und Lexikongrenzen hinweg
- Automatisches Framework für nötige „Breite“

UBY – <https://www.ukp.tu-darmstadt.de/uby/>



Umfangreiches Valenzlexikon für Verben, Nomen und Adjektive

- ca. 10.500 Verben mit 350 Subkatrahmen
- automatisch aus Zeitungskorpora extrahiert und manuell korrigiert
- Teil des *IMSLex German Lexicon*
- siehe Eckle-Kohler (1999)

genießen (subj(NP_nom))

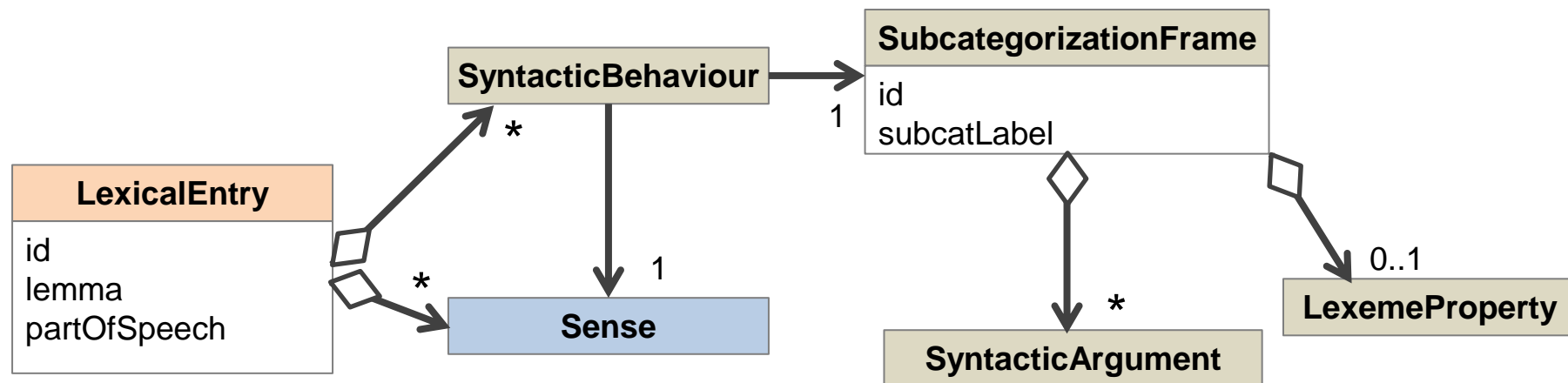
genießen (subj(NP_nom), corr-obj(PRON_es), obj(VP_zu-inf-pres))

genießen (subj(NP_nom), obj(NP_acc))

genießen (subj(NP_nom), s-comp(C_daß))

genießen (subj(NP_nom), v-comp(VP_zu-inf-pres))

Syntaktisches Modell in UBY-LMF



LexicalEntry: *genießen* (verb)

Sense: „mit Freude auf sich wirken lassen“

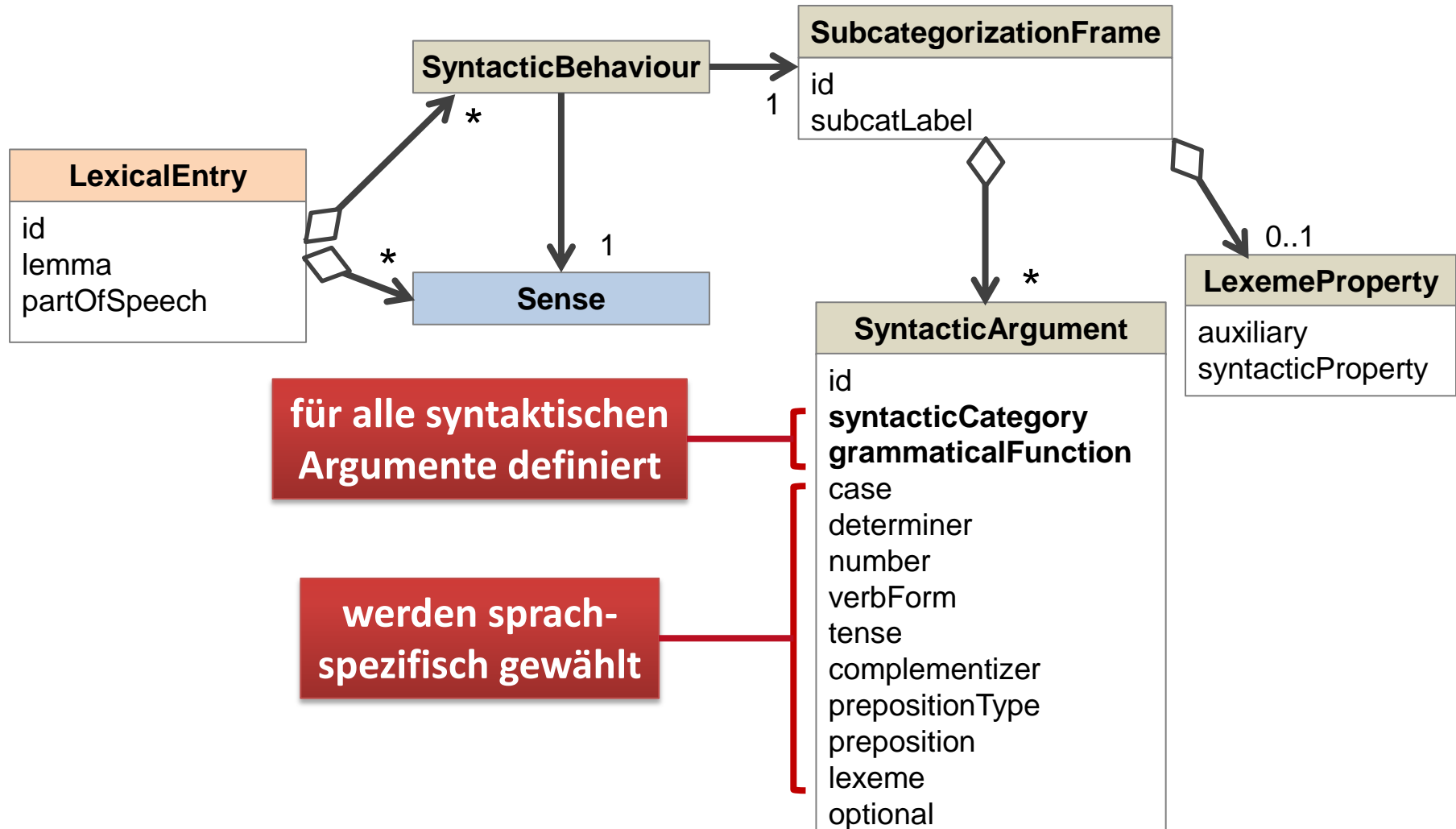
Subcategorization Frame:

SyntacticArgument 1: Subjekt (NP)

SyntacticArgument 2: Komplement (VP-zu)

LexemeProperty: Hilfsverb „haben“

Syntaktisches Modell in UBY-LMF



Standardisierte Valenzstrukturen in UBY

IMSLex-Subcat „genießen“ : (subj(NP_nom), v-comp(VP_zu-inf-pres))

Syntactic Argument

grammaticalFunction: subject
syntacticCategory: nounPhrase
case: nominative

Syntactic Argument

grammaticalFunction: complement
syntacticCategory: verbPhrase
verbForm: **toInfinitive**
tense: present

Syntactic Argument

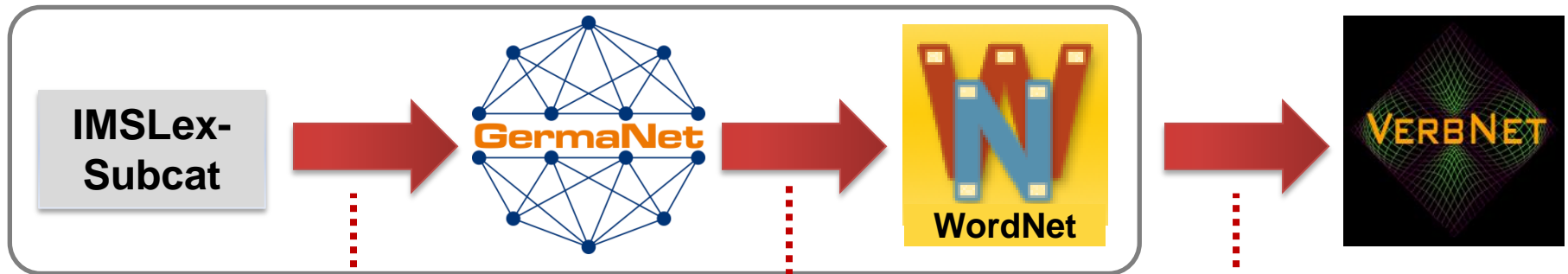
grammaticalFunction: subject
syntacticCategory: nounPhrase

Syntactic Argument

grammaticalFunction: complement
syntacticCategory: verbPhrase
verbForm: **ingForm**

VerbNet „enjoy“: NP V S_ING

Alignierung von Subkategorisierungsrahmen



SyntacticArgument
id
syntacticCategory
grammaticalFunction
case
determiner
number
verbForm
tense
complementizer
preposition
lexeme

Genauigkeit: 90%

Interlingual
Index (ILI)

SyntacticArgument
id
syntacticCategory
grammaticalFunction
case
determiner
number
verbForm
tense
complementizer
preposition
lexeme

Genauigkeit: 88%

Identifikation von Kontrasten

IMSLex-
Subcat

Alignierte Subkategorisierungsrahmen



SyntacticArgument

id
syntacticCategory
grammaticalFunction
case
determiner
number
verbForm
tense
complementizer
preposition
lexeme

Beispiel:
nicht-finite
Verbergänzung

Ergebnisse: nicht-finite Verbergänzung

44 Verbpaare mit Kontrast „zu-Infinitiv vs. *ing*-Form“:

confess	zugeben
consider	überlegen
convert	überzeugen
dare	wagen
desire	verlangen
dread	fürchten
encourage	ermuntern
fear	befürchten
fear	fürchten
...	...

Beispiel:

*Er überlegte, den Vortrag **zu beenden**.*

** He considered **to finish** the talk.*

*He considered **finishing** the talk.*

Ergebnisse: reflexive Verben

71 Verbpaare mit Kontrast „reflexive vs. nicht-reflexive Nutzung“:

close	schließen
complain	beklagen
consider	überlegen
constitute	bilden
continue	halten
copy	abschreiben
cry	weinen
curl	wellen
fear	fürchten
...	...

Beispiel:

*Sie schrieb **sich** die Literaturverweise ab.*

** She copied **herself** the references.
She copied the references.*

22 Verbpaare mit Kontrast „*da/es*-Korrelat im Hauptsatz“:

accept	akzeptieren
accomplish	erreichen
consider	bedenken
discover	erfahren
hope	hoffen
know	wissen
plan	vorsehen
realize	einsehen
think	denken
...	...

Beispiel:

*Wir hoffen **darauf**, eine interessante
Tagung zu haben.*

*We hope to have an interesting
conference.*

- Kontrastive Linguistik
- Zweitspracherwerbsforschung
- Exemplarische Anwendung:
*Erzeugung von kontrastivem
Übungsmaterial*
- Verfügbare Daten:
 - kontrastive Verblisten
 - UBY: <http://www.ukp.tu-darmstadt.de/uby>



Vielen Dank für die Aufmerksamkeit!

Judith Eckle-Kohler (1999): *Linguistisches Wissen zur automatischen Lexikon-Akquisition aus deutschen Textcorpora*, Dissertation, Berlin: Logos.

Judith Eckle-Kohler und Iryna Gurevych (2012): Subcat-LMF: Fleshing out a standardized format for subcategorization frame interoperability. In: *Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, 550–560, Avignon, France.

Klaus Fischer (1997): *German-Englisch verb valency: a contrastive analysis*. Tübingen: Narr (= Tübinger Beiträge zur Linguistik 422).


Iryna Gurevych, Judith Eckle-Kohler, Silvana Hartmann, Michael Matuschek, Christian M. Meyer, und Christian Wirth (2012): 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)*, 580–590. Avignon, Frankreich.


Kontakt / Contact


Judith Eckle-Kohler und 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

 {eckle-kohler, 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.