

Ansätze einer intralingualen kontrastiven Korpuslinguistik - aufgezeigt am Beispiel von Rechtstexten aus Südtirol, Österreich und Deutschland

Andrea Abel / Stefanie Anstein / Isabella Ties
Institut für Fachkommunikation und Mehrsprachigkeit
Europäische Akademie Bozen

LSP 2007

Kolloquium „Sprach- und Textmuster in der Fachkommunikation“
Hamburg, den 30. August 2007



Aufbau

1. Hintergrund
 - a) Die deutsche Sprache in Südtirol
 - b) Untersuchungen zur deutschen Standardsprache in Südtirol
 - c) Untersuchungen zur deutschen Rechtssprache in Südtirol
 - d) Forschungsdesiderat
2. Ziele des Beitrags
 - a) neue Methoden zur Untersuchung von Varietäten
 - b) Besonderheiten der deutschen Rechtssprache in Südtirol
3. Datengrundlage - Korpora
4. Vorstellung der Methode -
Beschreibung des Vorgehens und der Analysemuster
5. Aufzeigen von Beispielen
6. Fazit und Ausblick



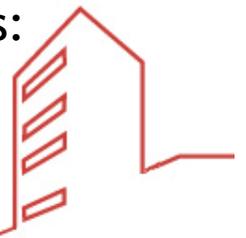
1a) Die deutsche Sprache in Südtirol

- Pariser Vertrag 1946 (Gruber-De Gasperi Abkommen): Gleichberechtigung der deutschen und italienischen Sprache, besonderer Schutz der deutschen Sprachgruppe, autonome Gesetzgebungs- und Vollzugsgewalt für die Provinzen Trient und Bozen. Unzulänglichkeiten =>
- Zweites Autonomiestatut 1972: u.a. wichtige Bestimmungen in Bezug auf die Sprachenrechte (Bereiche wie Sprachengleichstellung, Sprachgruppenproporz, Sprachgruppenzugehörigkeitserklärung, Zweisprachigkeitsprüfung)



1a) Die deutsche Sprache in Südtirol

- „Die deutsche Sprache ist in der Region der italienischen Sprache, die die amtliche Staatssprache ist, gleichgestellt [...]“ (Art. 99)
- „Die deutschsprachigen Bürger der Provinz Bozen haben das Recht, im Verkehr mit den Gerichtsämtern und mit den Organen und Ämtern der öffentlichen Verwaltung [...] ihre Sprache zu gebrauchen.“ (Art. 100)
- „In der Provinz Bozen wird der Unterricht in den Kindergärten, Grund- und Sekundarschulen in der Muttersprache der Schüler, das heißt in italienischer oder deutscher Sprache, von Lehrkräften erteilt, für welche die betreffende Sprache ebenfalls Muttersprache ist [...]“ (Art. 19) (aus: Das neue Autonomiestatut 2003)



1b) Untersuchungen zur geschriebenen deutschen nicht fachsprachlichen Standardsprache in Südtirol

- Untersuchungsschwerpunkte
 - kontaktsprachliche Erscheinungen/sprachliche Besonderheiten besonders lexikalischer, aber auch morphologisch-syntaktischer Art (z.B. Rizzo-Bauer 1962, Riedmann 1972, Pernstich 1984, Forer/Moser 1988, Lanthaler/Saxalber 1995, Ammon 2001, 2004, Abfalterer 2005)
 - kaum Untersuchungen zu Erscheinungen auf syntagmatischer (Kollokationen, idiomatische Wendungen) und textueller Ebene (z.B. Riehl 1997)
 - kaum Untersuchung kontaktsprachlicher Erscheinungen in übersetzten Texte (z.B. Putzer 1984 - Zweisprachigkeitsprüfungen - auf verschiedenen sprachlichen Ebenen)



1b) Untersuchungen zur geschriebenen deutschen nicht fachsprachlichen Standardsprache in Südtirol

- Interpretationsschwerpunkte zu kontaktsprachlichen Erscheinungen
 - 70er/80er Jahre: Kontaktphänomene als Beeinträchtigungen des Systems, Angst vor Sprachverfall (z.B. Riedmann 1972) (= Interferenzforschung vs. Beschreibung von Sonderwortschätzen aus variationslinguistischer Sicht)
 - neben nahe liegenden Übernahmen aus dem Italienischen besonders im Bereich der Amts-/Verwaltungs-/Rechtssprache weniger Auffälligkeiten als angenommen (z.B. Ammon 2001, Pernstich 1984)
 - Frage nach gutem/korrektem Deutsch - Unsicherheiten; Frage nach Norminstanzen; welches Deutsch in Südtirol? (z.B. Lanthaler/Saxalber 1995, Lanthaler 2001, Ammon 2001)



1b) Untersuchungen zur geschriebenen deutschen nicht fachsprachlichen Standardsprache in Südtirol

- Erhebungsmethoden
 - Belegsammlungen/Exzerpte (z.B. Riedmann 1972, Pernstich 1984, Riehl 1997, Ammon 2004, Abfalterer 2005)
 - Rückfragen an Experten/Gewährspersonen und Vergleich von einschlägigen Wörterbüchern und Sekundärliteratur (siehe dazu z.B. Abfalterer 2005)
 - Internet als Quelle für die Variationslinguistik (siehe dazu z.B. Bickel 2000, Ammon 2004)
 - Korpuslinguistik - C4-Verbund als viel versprechende Grundlage für vertiefende Untersuchungen zu den Varietäten des Standarddeutschen



1c) Untersuchungen zur geschriebenen deutschen Fachsprache in Südtirol

- Untersuchungsschwerpunkte
 - Form: Textstruktur und Inhalt (z.B. Urteile Arntz 1995)
 - Untersuchungen zum Einfluss der italienischen Sprache auf lexikalischer Ebene (Wölk 1996), z.B. Quästur, Ökonomat etc.
 - Untersuchungen zum Einfluss der in Österreich und Deutschland verwendeten Rechtsterminologie (vgl. Wölk 1996)
 - Konzeptorientierte Studien (Arntz 1995) / Studien zur vergleichenden Rechtsterminologie (Arntz/Mayer 1996 und Sandrini 1999);
 - Normierung der Rechtsterminologie > Paritätische Terminologiekommission (seit 1994; DPR 574/88, Art. 6, Abs. 1)
 - Rechtskollokationen: Projekt Zanichelli (seit 2006)



1c) Untersuchungen zur geschriebenen deutschen Fachsprache in Südtirol

- Interpretationsschwerpunkte zu kontaktsprachlichen Erscheinungen
 - Spiegelbildeffekt (Wölk 1995),
z.B. : Ministero per gli affari esteri = Ministerium für auswärtige
Angelegenheiten (nicht: Außenministerium)
 - Unverständlichkeit der Rückübersetzung (von L2 > L1)
 - Vielfalt von Synonymen (Coluccia 2000)



1c) Untersuchungen zur geschriebenen deutschen Fachsprache in Südtirol

- Erhebungsmethoden; Termini, Definitionen und Belegstellen aus
 - Handbüchern
 - Rechtstexten
 - Urteilen
 - ausgewählten Internetseiten (nur Rechtssprache)
 - fachspezifischen Korpora



1d) Forschungsdesiderat

- groß angelegte systematische Untersuchungen zu den Besonderheiten der deutschen allgemein- und fachsprachlichen Standardsprache in Südtirol mit Hilfe neuer Methoden (Korpuslinguistik)
 - auf syntagmatischer Ebene
(Analyse rekurrenter Sprachmuster, Kollokationen)
 - auf textueller Ebene

im intralingualen Vergleich (Varietäten)



2. Ziele des Beitrags

- Vorstellen neuer **Methoden** zur Untersuchung von Varietäten (auf lexikalischer und syntagmatischer Ebene) - Beitrag zur Methodik in der Variationsforschung
- Ermittlung von **Besonderheiten** der deutschen Standardsprache Südtirols an Hand von Rechtstexten (auf lexikalischer und syntagmatischer Ebene)



3. Datengrundlage - Korpora

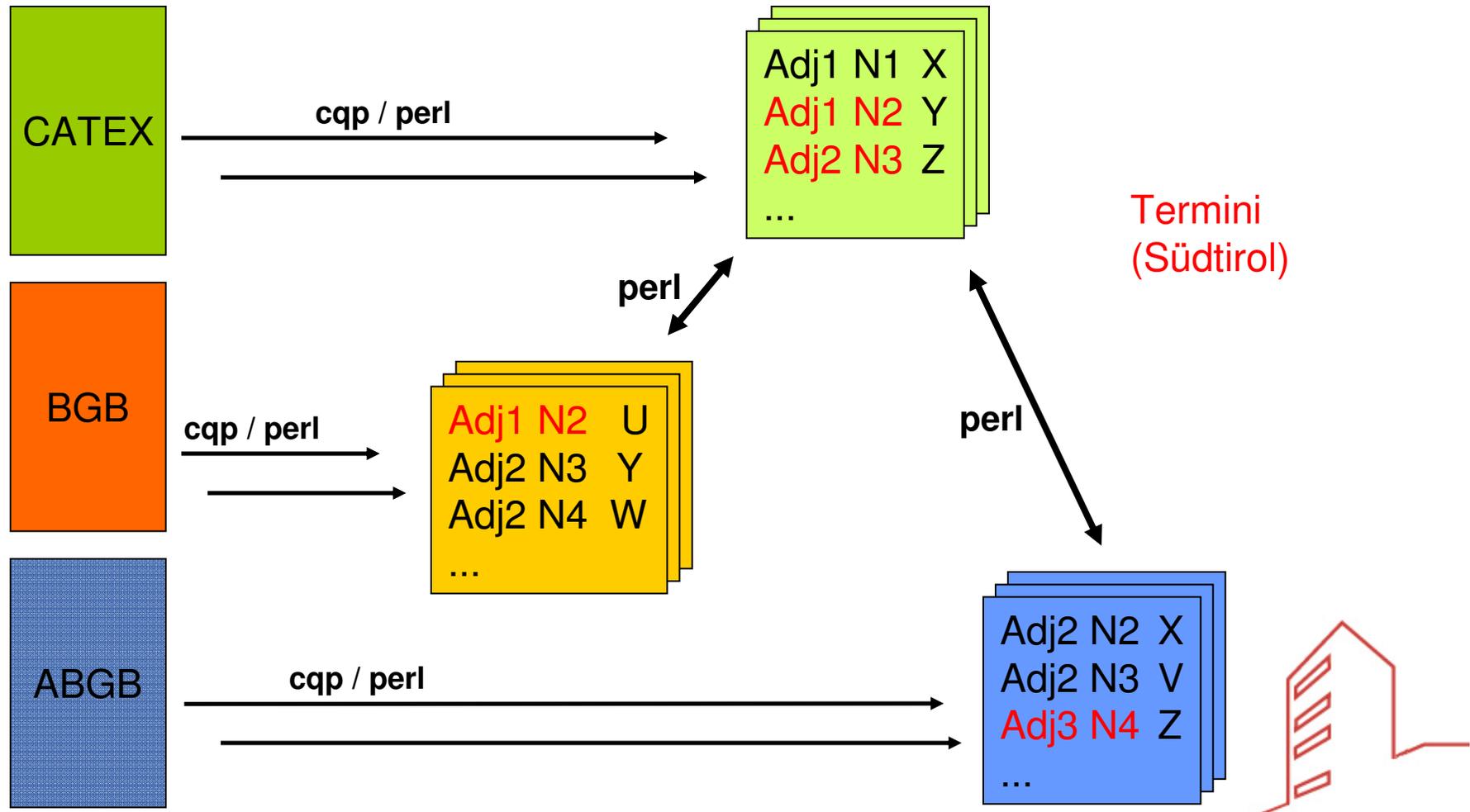
- CATEX
 - deutsche Übersetzung des 'Codice Civile' (Südtirol; Gamper 2000)
 - 302.080 Tokens
- BGB
 - Bürgerliches Gesetzbuch (Deutschland)
 - 243.513 Tokens
- ABGB
 - Allgemeines Bürgerliches Gesetzbuch (Österreich)
 - 208.614 Tokens
- wortartgetaggt und lemmatisiert mit dem TreeTagger (Schmid 1994)
- aufbereitet mit der Corpus WorkBench des IMS Stuttgart (Evert 2005)



4. Methode

- Extraktion von direkten Kookkurrenzen (z.B. als Termkandidaten)
(Auswahl der Wortartmuster:
Terminologische Datenbank EURAC & Heid 2001)
 - Adjektiv-Nomen
 - Nomen-Artikel-Nomen
 - Nomen-Präposition-Nomen
 - Formulierung von Extraktionsanfragen für komplexere
Relationen (Topologisches Feldermodell: Höhle 1986)
 - Subjekt-Verb
 - Verb-Objekt
- CATEX
BGB
ABGB
-
- Abgleich: Südtiroler Termini ↔ direkte Kookkurrenzen
 - Überprüfung der Liste von nicht lemmatisierten Wörtern
- CATEX


A. Datenvergleich Varietäten



Beispielanfrage V2 Verb-Objekt

(/pp[] | [pos="ADV"]){1,3}

evtl. Präpositionalphrasen/Adverbien

[pos="VVFIN"]

finites Verb

/subj_np[]

Subjekt-Nominalphrase

[word="sich"]?

evtl. Reflexivpronomen

(/pp[] | [pos="ADV"]){0,3}

/obj_acc_np_t[]

Akkusativobjekt-Nominalphrase

(/pp[] | [pos="ADV"]){0,3}

[pos = "PTKVZ"]?

evtl. Verbpartikel

[word="\. |\, |\; |\: |und |oder"]

within s;

innerhalb eines Satzes

➔ <In diesem Fall **bestimmt** das Gericht die **Bedingungen** ,> die ...



5. Ergebnisse: Beispielliste Subjekt-Verb

Bestimmung finden	53	(CATEX)
Grund vorliegen	12	
Frist laufen	10	
Gesellschaft beteiligen	6	
Gericht treffen	5	
Fall stehen	5	
Verbindlichkeit erlöschen	5	
Vorschriften finden	4	
Sachverhalt bestehen	4	
Gesellschaft bestehen	4	
Landesgericht entscheiden	4	
Anwachsung erfolgen	4	
Beauftragte haften	3	
Trennmauer befinden	3	



Adj+N: Catex vs. BGB / ABGB

	CATEX	BGB	ABGB
vorhergehend Artikel	136	0	0
elterlich Gewalt	77	0	0
erst Absatz	69	0	0
beschränkt Haftung	41	0	10
nichtöffentlich Sitzung	27	0	0



N+Art+N: Catex vs. BGB / ABGB

	CATEX	BGB	ABGB
Absatz des Artikels	187	0	0
Präsident der Republik	127	0	0
Dekret des Präsidenten	124	0	0
Bestimmung des Artikels	94	0	0
Präsident des Landesgerichts	35	0	0
Vorbehalt der Inventarerrichtung	30	0	0
Bestimmung der Sondergesetze	29	0	0



N+Präp+N: Catex vs. BGB / ABGB

	CATEX	BGB	ABGB
Gesetz über die Familienrechtsreform	240	0	0
Fußnote zu Artikel	97	0	0
Fußnote vor Artikel	29	0	0
Absatz laut Artikel	28	0	0
Gesetz über die Ehescheidung	17	0	0
Gesetz über die Bonifizierung	16	0	0



Subjekt+Verb: Catex vs. BGB / ABGB

	CATEX	BGB	ABGB
Grund vorliegen	12	39	5
Gericht entscheiden	2	1	1

Verb+Objekt: Catex vs. BGB / ABGB

	CATEX	BGB	ABGB
Anwendung finden	68	45	7
Anordnung treffen	1	5	1



Subjekt+Verb: Catex vs. BGB / ABGB

CATEX	BGB	ABGB
Gericht bestimmt, bewilligt, entscheidet, genehmigt, ordnet an, spricht aus	Gericht bestimmt, entscheidet, stellt fest, trifft (Maßnahmen, eine Feststellung)	Gericht entscheidet, erteilt (Aufträge), ordnet an, schränkt ein

Verb+Objekt: Catex vs. BGB / ABGB

CATEX	BGB	ABGB
Anordnungen, Maßnahmen, Verfügungen, Vorsorge treffen	Anordnung, Fürsorge, Maßnahmen, Vereinbarung, treffen	Anordnungen, Übereinkunft, Wahl treffen
Maßnahmen treffen, ergreifen	Maßnahmen treffen	Maßnahmen treffen

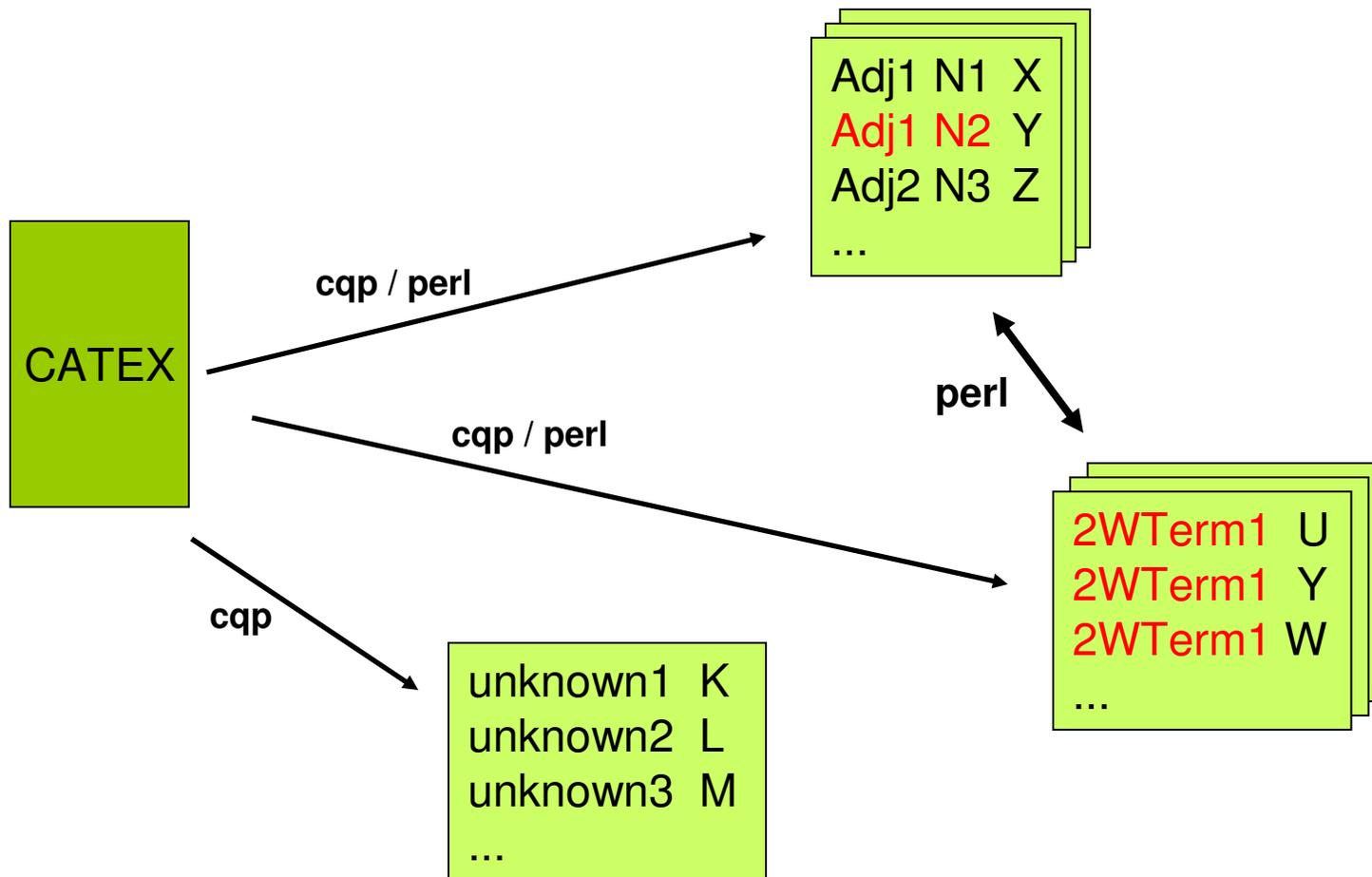


Südtirol-Termini: Catex vs. BGB / ABGB

	CATEX	BGB	ABGB
Dekret	194	0	0
Domizil	57	0	0
beschränkt Entmündigung	24	0	0
voll Entmündigung	15	0	0
Antrag auf Legitimation	5	0	0
Verschmelzung von Gesellschaften	4	0	0
Klagsanspruch auf Nichtigerklärung	4	0	0



B. Datenvergleich Südtirol



Ergebnisse Datenvergleich Südtirol

		Frequenz nach Wortartmuster	Vorkommen in Termlisten
Adj+N vs. 2-WT	gerichtlich Entmündigung	0	X
N+Art+N vs. 3-WT	Ablauf der Frist	17	-
	Herabsetzung des Kapitals	17	-
	Aufhebung des Vertrages	18	-
N+Präp+N vs. 3-WT	Antrag auf Legitimation	5	-
	Klagsanspruch auf Nichtigerklärung	4	-
	Verschmelzung von Gesellschaften	3	-



CATEX: ‘unknowns‘

Familienrechtsreform	242
Fruchtgenuß	70
Klagsanspruch	59
Wechselgesetz	41
Bestandnehmer	40
Ersitzung	39
Zwangsveräußerung	30
Bonifizierung	27
Indossament	24
Kontokorrentinhaber	8
Indossatar	8
...	



6. Fazit und Ausblick

Methode

- Bewältigung großer Datenmengen, aber
- Fehleranfälligkeit bei komplexen Anfragen

Ergebnisse

- Besonderheiten der Südtiroler Rechtssprache auf der Ebene der Ein- und Mehrworttermini bestätigt & Ermittlung von Termkandidaten möglich
- kaum Auffälligkeiten im Bereich der Subj/Obj-Verb-Verbindungen (Datenmenge für Varietätenvergleich ausreichend?)



6. Fazit und Ausblick

Ausblick

- Verbesserung der Lemma-Annotation der Korpora
- Verfeinerung der Extraktionsanfragen für Kookkurrenzen
- Kombination von Ansätzen bei der Vorverarbeitung der Korpora (Chunks, Phrasen, grammatische Funktionen) zur Reduzierung von Ambiguitäten
- manuelle Feinarbeit zur Datenanalyse
- Arbeit mit größeren Datenmengen
(→ Aussagekraft von Frequenz und Signifikanzmaßen)



Auswahlbibliographie

- Abfalterer, H. (2005): Der Südtiroler Sonderwortschatz aus plurizentrischer Sicht. Lexikalisch-
semantische Besonderheiten im Standarddeutsch Südtirols. Innsbruck (Dissertation)
- Ammon, U. (2001): Die Plurizentrität des Deutschen, oder: Wer sagt, was gutes Deutsch ist? In:
Egger, K. / Lanthaler, F. (Hrsgg.): Die deutsche Sprache in Südtirol: Einheitssprache oder
regionale Vielfalt. Wien - Bozen: Folio. 11-26.
- Ammon, U. (2004): Variantenwörterbuch des Deutschen: die Standardsprache in Österreich, der
Schweiz und Deutschland sowie in Liechtenstein, Luxemburg, Ostbelgien und Südtirol. Berlin: de
Gruyter.
- Arntz, R. (1995): Confrontare, valutare, trasporre: metodi e problemi della traduzione giuridica. In
La traduzione: nuovi approcci tra teoria e pratica. Arntz, Reiner [Hg.] Napoli: CUEN S. 137-162.
- Arntz, R. & Mayer, F. (1996): Vergleichende Rechtsterminologie und Sprachdatenverarbeitung: das
Beispiel Südtirol. In: Übersetzungswissenschaft im Umbruch. A. Lauer & Wilss, W. (Hrsg.)
Tübingen : Narr, S. 117-129.
- Bickel, H. (2000): Das Internet als Quelle für die Variationslinguistik. In: Häcki Buhofer, A. (Hrsg.):
Vom Umgang mit sprachlicher Variation: Soziolinguistik, Dialektologie, Methoden und
Wissenschaftsgeschichte. Tübingen - Basel (= Basler Studien zur deutschen Sprache und Literatur;
Bd. 80). 111-124.
- Coluccia, S. (2000): Il linguaggio giuridico in Alto Adige. In: Pasinato, A. (a cura di): Heimat.
Serie Saggi / Arti e lettere. Roma: Donzelli, S. 379-388.



- Lanthaler, F. (2001): Zwischenregister der deutschen Sprache in Südtirol. In: Egger, K. / Lanthaler, F. (Hrsgg.): Die deutsche Sprache in Südtirol: Einheitssprache oder regionale Vielfalt. Wien - Bozen: Folio. S.137-152.
- Evert, S. (2005): The CQP query language tutorial. Technical report, Institut für Maschinelle Sprachverarbeitung, Universität Stuttgart.
<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench>
- Forer, Rosa / Moser, Hans (1988): Beobachtungen zum westösterreichischen Sonderwortschatz. In: Wiesinger, Peter (Hrsg.): Das österreichische Deutsch. Wien - Köln - Graz: 189-209.
- Gamper, J. (2000): Aufbau eines zweisprachigen Textkorpus für Terminologie-Extraktion. In: Veronesi, D. (Hrsg.): *Linguistica giuridica italiana e tedesca*. Padova: Unipress. 307-319.
- Heid, U. (2001): Collocations in Sublanguage Texts: Extraction from Corpora. In: Wrigth, Sue Ellen / Budin, Gerhard (Hrsgg.): *Handbook of Terminology Management*. Amsterdam - Philadelphia: John Benjamins. 788-808.
- Höhle T. N. (1986): Der Begriff 'Mittelfeld'. Anmerkungen über die Theorie der topologischen Felder. In: Akten des VII. Internationalen Germanisten-Kongresses Göttingen 1985. Volume 3. pp. 329--340. Niemeyer, Tübingen.
- Lanthaler, F. / Saxalber, A. (1995): Die deutsche Standardsprache in Südtirol. In: Muhr, R. / Schrod, R. et al. (Hrsgg.): *Österreichisches Deutsch. Linguistische, sozialpsychologische und sprachliche Aspekte einer nationalen Variante des Deutschen*. 289-305.
- Mayer, F. (2000): Terminographie im Recht: Probleme und Grenzen der Bozner Methode. In: Veronesi, D. (Hrsg.): *Linguistica giuridica italiana e tedesca*. Padova: Unipress. 295-306.
- Pernstich, K. (1982): Deutsch-italienische Interferenzen in der Südtiroler Presse. In: Moser, Hans (Hrsg.): *Zur Situation des Deutschen in Südtirol. Sprachwissenschaftliche Beiträge zu den Fragen von Sprachnorm und Sprachkontakt (= Innsbrucker Beiträge zur Kulturwissenschaft, Germanistische Reihe, Bd. 13)*. Innsbruck. 91-182.



- Putzer, O. (1984): Interferenz in Übersetzungen: Aspekte der Übersetzungsleistungen bei der Zweisprachigkeitsprüfung (D.P.R. 752/76). Bozen: Assessorat für Schule und Kultur in italienischer Sprache.
- Riedmann, G. (1972): Die Besonderheiten der deutschen Sprache in Südtirol (= Duden Beiträge, Sonderreihe: Die Besonderheiten der deutschen Schriftsprache im Ausland 39). Mannheim.
- Riehl, C. M. (1997): Schreiben, Text und Mehrsprachigkeit: Beitrag zu einer Theorie der mehrsprachigen Gesellschaft am Beispiel der deutschsprachigen Minderheiten in Südtirol und Ostbelgien. Freiburg (Habilitation)
- Rizzo-Bauer, Hildegard (1962): Die Besonderheiten der deutschen Schriftsprache in Österreich und Südtirol. (= Duden Beiträge, Sonderreihe: Die Besonderheiten der deutschen Schriftsprache im Ausland 5). Mannheim.
- Sandrini, P. (1999). Übersetzen von Rechtstexten. Fachkommunikation im Spannungsfeld zwischen Rechtsordnung und Sprache. Forum für Fachsprachen-Forschung ; 54. Tübingen: Narr.
- Schmid, H. (1994): Probabilistic Part-of-Speech Tagging Using Decision Trees. Proceedings of International Conference on New Methods in Language Processing. <http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger>
- Woelk, Jens (1998): Anmerkungen zur deutschen Rechtssprache in Südtirol. In: Informator. Trento: ICA, S. 262-277.
- Zanichelli / DIGIT-Projekt:
http://www.eurac.edu/Org/LanguageLaw/Multilingualism/Projects/DIGIT_it.htm



Vielen Dank

für Ihre Aufmerksamkeit sowie
für Fragen, Anregungen, Kommentare ...

andrea.abel@eurac.edu

stefanie.anstein@eurac.edu

isabella.ties@eurac.edu

www.eurac.edu

