

Computerlinguistische Anwendungen zur Nutzung normierter terminologischer Daten

Verena Lyding & Isabella Ties

Sammlungen terminologischer Daten können neben der herkömmlichen Veröffentlichung in Druckform dem Benutzer auch in elektronischer Form zugänglich gemacht werden. Der Einsatz digitaler Medien ermöglicht dynamische Formen der Datenabfrage und -darstellung sowie die Einbindung der Daten in computerlinguistische Werkzeuge.

Im vorliegenden Artikel werden am Beispiel bistros – dem Informationssystem für Rechtsterminologie – drei Komponenten zur Unterstützung der Online-Terminologierecherche vorgestellt: die Termsuche, die Termerkennung und die Termextraktion. Die Funktionsweise jeder Komponente wird im Detail beschrieben. Im Anschluss wird diskutiert, welche Schwierigkeiten sich beim automatischen Zugriff auf sprachliche Daten stellen und wie mit diesen umgegangen werden kann.

1. Einleitung

1.1. Historischer Hintergrund

Die Einführung der deutschen Sprache als Amtssprache in Südtirol ist bereits im Sonderstatut für Trentino-Südtirol aus dem Jahre 1972 vorgesehen. Laut Artikel 99 ist die deutsche Sprache der italienischen Sprache gleichgestellt, und wie Artikel 2 besagt, werden jedem Bürger der Region unabhängig von seiner Sprachgruppe die gleichen Rechte zuerkannt. „Die deutschsprachigen Bürger der Provinz Bozen haben das Recht, im Verkehr mit den Gerichtsämtern und mit den Organen und Ämtern der öffentlichen Verwaltung, die ihren Sitz in der Provinz haben oder regionale Zuständigkeit besitzen sowie mit den Konzessionsunternehmen, die in der Provinz öffentliche Dienste versehen, ihre Sprache zu gebrauchen“ (Art. 100). Mit dem Dekret des Präsidenten der Republik vom 15. Juli 1988, Nr. 574, kommt es erstmals zur konkreten Umsetzung der Gleichstellung der deutschen

Sprache mit der italienischen Sprache. Die Gleichstellung bezieht sich auf den Verkehr der Bürger mit der öffentlichen Verwaltung und die Verwendung in Gerichtsverfahren. Mit dem Recht des Gebrauchs der deutschen Sprache wächst der Bedarf nach einer einheitlichen und eindeutigen deutschsprachigen Terminologie. Diese soll der öffentlichen Verwaltung entgegen kommen, die Rechtsakte und verwaltungsrechtliche Dokumente in italienischer und/oder deutscher Sprache verfasst. Weiters soll eine klare und verständliche Kommunikation der Verwaltung mit den Bürgern gewährleistet werden, die nach Bedarf des Bürgers sowohl in der einen als auch in der anderen Sprache stattfinden kann. In der oben genannten Bestimmung wird außerdem die Schaffung der „Paritätischen Terminologiekommision“ vorgesehen. Die Paritätische Terminologiekommision (TerKom) hat die Aufgabe, die Rechts- und Verwaltungsterminologie sowie sonstige Fachterminologie, die in den öffentlichen Organen und Ämtern verwendet wird, zu bestimmen und regelmäßig zu aktualisieren (DPR 574/88, Art. 6, Abs. 1). Seit 1994 kümmert sich die Paritätische Terminologiekommision zusammen mit dem Institut für Fachkommunikation und Mehrsprachigkeit der Europäischen Akademie Bozen um die Erarbeitung einer homogenen deutschen Rechtsterminologie für Südtirol.

1.2. Datenbasis

Seit Beginn der terminologischen Arbeiten am Institut wurden mehr als 50.000 Termini erarbeitet. Die Terminologie ist in mehrsprachigen terminologischen Einträgen organisiert. Die Einträge werden in enger Zusammenarbeit von Juristen und Linguisten erstellt. Jeder Terminus, der in die Datenbank eingefügt wird, enthält Angaben zu Grammatik, Sprachgebrauch (AT = Österreich, CH = Schweiz, DE = Deutschland, STAA = Südtirol, INT = Europäisches Recht), Wortform (Kürzel, Akronym, Lehnwort, Kollokation), Termstatus (abgelehnt, vorzulegen, Entscheidung der TerKom, Favorit, Vorzug) sowie Definition und Kontext in italienischer, deutscher und ladinischer Sprache. Bezüglich der deutschen Sprache werden Termini der Rechtssysteme Italien, Deutschland, Österreich und Schweiz sowie manchmal der EU eingefügt. Die Begriffe sind anhand von Querverweisen verlinkt. Es handelt sich um hierarchische Beziehungen wie z.B. Oberbegriff, Unterbegriff, Nebenbegriffe. Alle terminologischen Daten sind im Internet über *bistro* – das Informationssystem für Rechtsterminologie – abrufbar.

1.3. Nutzergruppen und Aufgabenbereiche

Die terminologischen Daten werden in erster Linie von Übersetzern, Juristen und Studenten aber auch von Dolmetschern und Lehrkräften konsultiert¹. Es handelt sich um

¹ Vgl. Ergebnisse der Online-Umfrage zu *bistro*.

eine heterogene Benutzergruppe. Je nach beruflicher Ausrichtung und akutem Bedarf sucht jeder Einzelne nach unterschiedlichen Informationen, wie z.B. nach expliziten Definitionen von Fachtermini, nach Übersetzungen oder nach Verwendungsbeispielen. Für die Terminologieerarbeitung stellt das eine besondere Herausforderung dar, da es nicht einfach ist, alle geforderten Informationen in einem einzigen terminologischen Eintrag unterzubringen. Neben dem einfachen Zugriff auf terminologische Daten können außerdem weitere Angebote, wie terminologische Werkzeuge, für einige Nutzergruppen von Interesse sein.

1. 4. Elektronische Datenverarbeitung

Während terminologische Daten lange Zeit auf Papier verfasst und verbreitet wurden, bieten Computeranwendungen und das Internet neue Möglichkeiten der Datendarstellung und -verarbeitung. Vorzüge, wie die Loslösung von alphabetischen Anordnungen, die dynamische Datendarstellung und unterschiedliche Möglichkeiten des Datenzugriffs mittels mächtiger Suchverfahren, für die elektronische Wörterbücher gelobt werden (De Schryver 2003:157), sind genauso für elektronische Terminologiesammlungen zu nennen. Die Struktur eines terminologischen Eintrags in virtuellen Datensammlungen oder auf Online-Plattformen ist nicht so festgeschrieben, wie es für gedruckte Formen erforderlich ist. Unterschiedliche Darstellungsformen können dynamisch erzeugt werden, Detailinformationen können auf Knopfdruck ein- und ausgeblendet werden und Abfragen erfolgen über automatische Suchen. Darüber hinaus eröffnen computerlinguistische Methoden Möglichkeiten, die Datensammlungen in Anwendungen einzubinden und dort zur Unterstützung des Nutzers einzusetzen. Sammlungen von Fachbegriffen können beispielsweise in ein Termerkennungswerkzeug oder in ein Programm zur automatischen Termextraktion einfließen.

Die Einbindung des Computers stellt eine große Bereicherung und Erweiterung der Möglichkeiten dar. Dennoch können automatische Verfahren der Komplexität von Sprache nur bis zu einem bestimmten Punkt Rechnung tragen, da sie vordefinierten Abläufen und starren Verfahrensmustern folgen.

1. 5. Überblick über *bistro* – das Informationssystem für Rechtsterminologie

bistro wurde von den WissenschaftlerInnen des Instituts für Fachkommunikation und Mehrsprachigkeit an der Europäischen Akademie Bozen entwickelt. Das kostenlose Online-Informationssystem (www.eurac.edu/bistro) kann als Hilfsmittel für die Recherche,

die Analyse und die Übersetzung von Rechts- und Verwaltungstermini und -texten in den Sprachen Italienisch, Deutsch und Ladinisch genutzt werden.

Das System beinhaltet folgende Komponenten:

- **EINE TERMINOLOGISCHE DATENBANK:** In der Datenbank sind ca. 50.000 Termini der italienischen Rechtsordnung (italienische, deutsche und ladinische Rechtstermini) sowie der österreichischen, schweizer und bundesdeutschen Rechtsordnungen (Rechtstermini in deutscher Sprache) erfasst. Teile der deutsch-italienischen Terminologie wurden durch die Paritätische Terminologiekommission normiert und sind für Südtirol verbindlich festgelegt.
- **DIE ONLINE-VERSION DES TERMINOLOGISCHEN WÖRTERBUCHS ZUM HOCHSCHULWESSEN ITALIEN-ÖSTERREICH:** Das Wörterbuch enthält ca. 1000 terminologische Einträge zu ca. 3000 Termini des italienischen und österreichischen Hochschulwesens.
- **DAS ZWEISPRACHIGE KORPUS CATEx:** Das CATEx (Computer Assisted Terminology Extraction) besteht aus einer Sammlung von italienischen Gesetzestexten und deren Übersetzungen ins Deutsche.
- **DAS DREISPRACHIGE KORPUS CLE** (vgl. Streiter et al. 2004): Das CLE (Corpus Ladin d'EURAC) besteht aus einer Sammlung von italienischen, deutschen und ladinischen Rechts- und Verwaltungstexten. Die Texte wurden von den Gemeinden des Gadertals und des Grödnertals sowie vom Ladinischen Pädagogischen Institut und vom Amt für Sprachangelegenheiten der Provinz Bozen zur Verfügung gestellt.
- **ZWEI WERKZEUGE ZUR AUTOMATISCHEN TEXTANALYSE:** Eine Terminologieerkennung, die die Termini, die in der Datenbank enthalten sind, in Texten farblich markiert und eine Termextraktion zur Erstellung von Termlisten aus Texten.

In den folgenden Abschnitten werden wir auf die Termsuche sowie auf die Termerkennung und Termextraktion näher eingehen. Wir werden die Funktionsweisen dieser Komponenten erläutern und Problembereiche sowie Lösungsmöglichkeiten aufzeigen.

2. Zugriff auf Termini: Funktionen zur Abfrage und Markierung von Fachausdrücken

2.1. Termsuche

Die Termsuche bezeichnet eine Funktion, die ausgehend von einem Suchbegriff den gezielten Zugriff auf alle zugehörigen Informationen der Terminologiesammlung ermöglicht. Gesuchte Fachbegriffe und orthographisch ähnliche Termini können aufgefunden und zusammen mit vollständigen terminologischen Einträgen sowie Übersetzungen angezeigt werden.

2. 1. 1. Termsuche bei normierter Terminologie

Bei der Arbeit mit mehrsprachiger, normierter Terminologie bestehen folgende Anforderungen an die Termsuche: Alle Begriffe und die zugehörigen terminologischen Informationen müssen über eine Stichwortsuche recherchierbar sein. Eine Ausweitung der Suche auf ähnliche Termini sollte möglich sein (De Schryver 2003:175). Konzeptuelle Unterschiede mehrdeutiger Begriffe müssen transparent gemacht werden. Für jeden Terminus muss der Normierungsstatus angegeben sein und nicht normierte Begriffe sollten auf normierte Synonyme hinweisen.

2. 1. 2. Termsuche in bistro

Die Suche nach Fachausdrücken wird über die Eingabe von Stichwörtern in ein einfaches Suchfenster gestartet (Abbildung 1). Es kann sowohl nach einfachen Fachausdrücken (z.B. „Kündigung“) als auch nach Mehrworttermini (z.B. „Kündigung aus Rachsucht“) gesucht werden. Im Ergebnis wird der gesuchte Begriff zusammen mit seinen möglichen Übersetzungen angezeigt (Abbildung 2). Zu jeder Übersetzung ist dabei das Fachgebiet angegeben, so dass Homonyme² unterschieden werden können. Den Angaben über den exakten Suchterminus folgt eine Liste von Termini, in denen das gesuchte Wort innerhalb eines Syntagmas vorkommt. Eine Suche nach dem Begriff „Kündigung“ liefert dementsprechend auch Informationen über „fristlose Kündigung“ oder „Kündigungsschutz“.



Abbildung 1: Eingabemaske für die Termsuche

2 In diesem Artikel wird der Begriff „Homonym“ in Sinne Wüsters als „eine Benennung, die dieselbe Form [...] wie eine andere Benennung, aber eine andere Bedeutung [hat]“ (1991:88) verwendet. Es wird keine Unterscheidung zwischen Homonymie und Polysemie gemacht.

Kündigung N-f	licenze N-f	civil proced.law
	licenziamento individuale N-m	labor law
(einseitige Willensäußerung zur Beendigung eines Rechtsverhältnisses)	denuncia N-f	labor law
	denunzia N-f	labor law
	licenze N-f	civil law
	rinunzia N-f	civil law
	disdetta N-f	law of obligations

Abbildung 2: Liste exakter Treffer bei der Suche nach „Kündigung“

Begriffe, die von der Paritätischen Terminologiekommission normiert wurden, sind grün unterlegt, während nicht normierte Begriffe unmarkiert bleiben. Darüber hinaus erlaubt es eine gesonderte Abfragefunktion, die Suche auf die Sammlung der normierten Termini zu beschränken. Falls die beschränkte Suche keine Ergebnisse liefert, kann sie im Folgeschritt abermals auf die gesamte Terminologiesammlung ausgeweitet werden.

Ausgehend vom gesuchten Terminus oder von der jeweiligen Übersetzung gelangt man in einem weiteren Schritt zum vollständigen terminologischen Eintrag. Wie oben beschrieben (siehe Abschnitt 1.2) enthält der terminologische Eintrag alle Informationen, die über einen Begriff zusammengetragen wurden. Er zeigt dazu alle verschiedenen Entsprechungen in den einzelnen Rechtsordnungen und Sprachen. Zu jedem Begriff ist der Sprachgebrauch, eine klare Definition sowie ein Kontext zur Veranschaulichung der praktischen Verwendung angegeben. Außerdem umfasst der Eintrag nochmals die Angaben zum Normierungsstatus und zu den Rechtsbereichen, denen die Termini angehören. Kontexte und Definitionen werden mit den entsprechenden Quellen angegeben.

2.2. Termerkennung

Die Termerkennung bezeichnet eine Funktion, die in einem beliebigen Text alle Termini einer Terminologiedatenbank identifiziert (Streiter et al. 2004:208). Die identifizierten Begriffe werden im Text hervorgehoben und mit der Datenbank verlinkt, so dass ein direkter Zugriff auf die Übersetzungen und terminologischen Einträge möglich ist³. Durch die automatische Markierung der Termini im Text ist für den Benutzer auf einen Blick ersichtlich, welche Termini bereits in der Datenbank vorhanden sind.

2.2.1. Termerkennung für normierte Terminologie

Die primäre Anforderung an jede Termerkennung ist es, alle Termini der Datenbank im Text zu identifizieren und in Farbe zu markieren. Zum einen erfordert das, flektierte Formen oder Mehrworttermini mit Einschüben auf Einträge der Datenbank zurückzuführen. Zum anderen muss für jedes erkannte Muster geprüft werden, ob es in dem Text und

³ Vgl. eCoLoRe Project.

Kontext tatsächlich dem Fachbegriff der Datenbank entspricht oder ob die Benennung in einer anderen fachlichen oder allgemeinsprachlichen Bedeutung auftritt. Wie diese Unklarheiten aufgelöst werden können und welche Schwierigkeiten sich dabei stellen, wird in Abschnitt 4.2.1 im Detail diskutiert.

Eine zweite Forderung bezieht sich auf die Darstellung der erkannten Termini. Im Fall von normierter Terminologie sollte bei der Markierung im Text auch der Normierungsstatus eines jeden Terminus erkennbar gemacht werden, so dass der Benutzer einen schnellen Überblick über den Status der im Text enthaltenen Terminologie gewinnen kann, ohne jeden Begriff einzeln nachschlagen zu müssen.

2.2.2. Termerkennung in bistro

Um die Termerkennung zu starten wird ein Text in das Webinterface *bistros* eingegeben. Dafür gibt es zwei Möglichkeiten: Der Text wird direkt in ein Textfeld kopiert oder es wird die Internet-Adresse angegeben, unter der der Text zu finden ist. Zusätzlich muss die Sprache des Textes ausgewählt werden. Das Programm ermittelt nun das Fachgebiet des Textes und identifiziert die enthaltenen Termini (Abbildung 3). Alle Termini, die in der terminologischen Datenbank enthalten sind, sind in roter Farbe angezeigt und sind anklickbar.

- Die von der Paritätischen Terminologiekommission normierten und für Südtirol gültigen Termini haben die grün unterlegte Anmerkung BAd und BAi (BAd = *termine tedesco pubblicato nel Bolletino Ufficiale*/ deutscher Terminus publiziert im Amtsblatt; BAi = *termine italiano pubblicato nel Bolletino Ufficiale*/ italienischer Terminus publiziert im Amtsblatt)
- Die von der Paritätischen Terminologiekommission abgelehnten Termini haben zusätzlich die rot unterlegte Anmerkung AB (AB = *termine tedesco rifiutato*/ abgelehnter deutscher Terminus).



Abbildung 3: Ausgabe der automatischen Termerkennung

Fährt man mit dem Mauszeiger über einen der markierten Begriffe, erscheint ein kleines Fenster mit den Übersetzungen des Begriffs. Klickt man auf einen Begriff, geht ein neues Browserfenster auf, in dem *bistros* Termsuche für diesen Begriff gestartet wird. Wie in Abschnitt 3.1.2 zur Termsuche beschrieben, wird der Begriff zusammen mit seinen Übersetzungen präsentiert und es kann auf den vollständigen terminologischen Eintrag mit Definition und Kontexten zugegriffen werden.

2.3. Termextraktion

Die Termextraktion bezeichnet eine Funktion, die basierend auf digitalisierten Fachtexten eine Liste von Termkandidaten ermittelt. Die extrahierten Termkandidaten müssen von einem Terminologen bestätigt/ verifiziert werden (Cabré Castellví et al. 2001:54).

Die automatische Extraktion von Termlisten ist neben dem Einsatz in der automatischen Sprachverarbeitung (z.B. maschinelle Übersetzung, Textindizierung) auch für die Erarbeitung von Glossaren und terminologischen Wörterbüchern von Bedeutung (Cabré Castellví et al. 2001:53). Die automatische Termextraktion vereinfacht und beschleunigt den Prozess der Termsammlung und ist damit insbesondere für Terminologen und Übersetzer von Nutzen.

2.3.1. Termextraktion in *bistro*

Ebenso wie bei der Termerkennung kann bei der Termextraktion in *bistro* der Text entweder in ein Textfeld kopiert oder über die URL angegeben werden. Zusätzlich besteht die Möglichkeit die Termextraktion auf ausgewählte Subsektionen der lokalen Korpora anzuwenden. Nach Angabe der Sprache des Textes und einer Obergrenze für die Anzahl zu extrahierender Termini wird die Termextraktion gestartet. Es wird eine Liste von Termkandidaten ausgegeben. Die Termini sind nach ihrer „Qualität“ sortiert. Das heißt, die Termini, die entsprechend des Extraktionsalgorithmus den stärksten „Termcharakter“ aufweisen, sind am Anfang der Liste zu finden, die weniger sicheren Kandidaten am Ende der Liste. Zu jedem Begriff ist dessen Häufigkeit im Text angegeben. Die in der Datenbank vorhandenen Termini werden je nach Termstatus farblich markiert (vgl. Termerkennung in Abschnitt 1.6). Ausgehend von den extrahierten Termini kann der Benutzer per Mausklick zur Termsuche *bistros*, zur terminologischen Datenbank des Generalsekretariats des Rates der Europäischen Union (TIS) oder zur alignierten Korpusuche innerhalb von CATEX (italienisch-deutsches Korpus) oder CLE (italienisch-deutsch-ladinisches Korpus) gelangen. Weiterhin kann für den Begriff eine Google-Suche oder eine KWIC-Suche (*Key Word In Context* Suche) in den Korpora gestartet werden.

3. Herausforderungen beim automatischen Zugriff auf terminologische Informationen

3.1. Kritische Aspekte der Termsuche

Die Termsuche ermöglicht es, aus der Terminologiesammlung, das heißt aus allen terminologischen Einträgen, die in der Datenbank gespeichert sind, über eine Suchmaske Termini, deren Übersetzung sowie vollständige Einträge abzurufen. In der Termsuche stellen insbesondere der Umgang mit Homonymen, die Hervorhebung normierter Daten und die Suche nach ähnlichen Wörtern eine Herausforderung dar.

3.1.1. Homonyme

Die Suche nach bestimmten Termini wird durch die Präsenz von Homonymen erschwert. Wenn der Benutzer ein Suchwort eingibt, hat er üblicherweise einen Terminus eines bestimmten Fachgebiets im Sinn. Ohne Disambiguierung werden ihm aber alle Homonyme angezeigt, auch die, die für ihn nicht relevant sind. Als Beispiel eines solchen Falls betrachten wir die Suche nach dem italienischen Terminus *atto*. Dieser Terminus kommt in der Datenbank in unterschiedlichen Fachgebieten mit unterschiedlichen Bedeutungen

vor. Mit jeder Bedeutung sind dementsprechend auch unterschiedliche Übersetzungen verknüpft. Je nach Fachgebiet (Zivilrecht, Verwaltungsrecht, Strafprozessrecht oder Strafrecht) müssen folgende Bedeutungen unterschieden werden: *atto* im Sinne von „Urkunde“, *atto* im Sinne von „Handlung“, *atto* im Sinne von „Rechtshandlung“, usw.

atto _{N-m}	Nachweis _{N-m}	civil law
	schriftlicher Bericht _{N-m}	civil law
	Urkunde _{N-f}	civil law
	Nachweis _{N-m}	admin.law
	schriftlicher Bericht _{N-m}	admin.law
	Urkunde _{N-f}	admin.law
(giuridico)	Rechtshandlung _{N-f}	law of obligations
	juristische Handlung _{N-f}	law of obligations
	menschliche Handlung _{N-f}	law of obligations
	Rechtshandlung _{N-f}	law of obligations
(documento)	Urkunde _{N-f}	meta
	Nachweis _{N-m}	meta
	schriftlicher Bericht _{N-m}	meta
	Urkunde _{N-f}	meta
	Akt _{N-m}	penal proced.law
	Akte des Verfahrens _{N-f}	penal proced.law
	Prozeßakt _{N-m}	penal proced.law
	Prozeßakte _{N-f}	penal proced.law
	Prozeßhandlung _{N-f}	penal proced.law
	Prozessakte _{N-f}	penal proced.law
	Prozesshandlung _{N-f}	penal proced.law
	Verfahrenshandlung _{N-f}	penal proced.law
	Verfahrensschrift _{N-f}	penal proced.law
	Handlung _{N-f}	penal law
	Handlung _{N-f}	penal law

Tabelle 1: *atto* in seinen unterschiedlichen Bedeutungen

Da im ersten Suchschritt keine Disambiguierung vorgenommen wird, kommt es in der Anzeige der Ergebnisse zu einer Vervielfachung der Daten: Der Suchterminus kommt mehrfach vor und ist jeweils mit einer anderen Übersetzung verknüpft (siehe Tabelle 1 für eine Auswahl der Ergebnisse).

Der Benutzer wünscht sich bei der Suche nach einem Terminus aber meist eine klare Antwort und einen eindeutigen Übersetzungsvorschlag. Eine mögliche Lösung für diesen

Fall wäre, den Benutzer bereits vor der Präsentation der Ergebnisse darauf hinzuweisen, dass der Begriff ambig ist. Dem Benutzer würden nicht die Übersetzungen, sondern die Fachgebiete, denen der Begriff angehört, präsentiert werden und er müsste eines daraus auswählen (für eine Beispielanzeige siehe Tabelle 2).

atto _{N-m}	civil law
	admin.law
	law of obligations
	penal proced.law
	penal law

Tabelle 2: Suchbegriff mit Angabe der Fachgebiete, in denen er auftritt

Die Einschränkung des Fachgebietes, zum Beispiel die Einschränkung auf das Verwaltungsrecht, würde den Begriff disambiguieren und somit eine gezielte Ergebnispräsentation ermöglichen (Tabelle 3). Es würden ausschließlich die für das Fachgebiet gültigen Übersetzungen angezeigt werden.

atto _{N-m}	admin.law	Nachweis _{N-m}
		schriftlicher Bericht _{N-m}
		Urkunde _{N-f}

Tabelle 3: *atto* mit Übersetzungen im Bereich des Verwaltungsrechts

Durch dieses Vorgehen würde der Disambiguierungsschritt vorgezogen. Das kann für den Benutzer jedoch ebenfalls problematisch sein, falls er das Fachgebiet nicht zu benennen weiß.

Da eine Disambiguierung basierend auf der Benennung allein nicht möglich ist, lässt sich das Problem der Homonymieabfrage nicht vollständig automatisch lösen. Bei der Implementierung der Termsuche muss entschieden werden, welche der Darstellungsvarianten für den Benutzer die am meisten geeignete ist.

3. 1. 2. Markierung des Normierungsstatus

Die von der Paritätischen Terminologiekommission normierten Begriffe sind für Südtirol bindend und müssen gegenüber nicht-normierten Äquivalenten bevorzugt verwendet werden. Der Sonderstatus der normierten Terminologie sollte aus der Darstellungsweise der Suchergebnisse ersichtlich sein.

Zur Zeit werden in *bistro* normierte Termini grün unterlegt, um den Benutzer darauf hinzuweisen, welche Übersetzung in Südtirol verwendet werden soll. Sucht der Nutzer

nach einem Terminus für den eine normierte Übersetzung existiert, kann er diese ohne Schwierigkeiten finden. Wird allerdings ein (deutscher) Suchbegriff eingegeben, der nicht normiert wurde, der aber ein normiertes Äquivalent besitzt, bleibt dies dem Benutzer zunächst verborgen. Zum Beispiel ergibt die Suche nach *decreto legislativo* alle deutschsprachigen Entsprechungen (Abbildung 4); die Suche nach „Legislativdekret“ resultiert ausschließlich in *decreto legislativo* (Abbildung 5). Zur Unterstützung sollte der Benutzer auf das normierte Synonym hingewiesen werden, falls er einen Terminus sucht, der ein Äquivalent hat, das normiert wurde. Im Beispielfall sollte der Benutzer durch eine Anmerkung und einen Verweis darauf aufmerksam gemacht werden, dass anstelle des nicht normierten „Legislativdekret“ das normierte Äquivalent „gesetzesvertretendes Dekret“ zu verwenden ist.

decreto legislativo N-m	Gesetzesverordnung N-f	admin.law
	gesetzesvertretendes Dekret N-n	admin.law
	gesetzesvertretende Verordnung *	admin.law
	Legislativdekret N-n	admin.law
	gesetzesvertretendes Dekret N-n	meta

Abbildung 4: *decreto legislativo* mit allen deutschen Übersetzungen

Legislativdekret N-n	decreto legislativo N-m	admin.law
	decreto legislativo N-m	meta

Abbildung 5: „Legislativdekret“ mit allen italienischen Übersetzungen

3. 1. 3. Suche nach ähnlichen Wörtern

Bei der Termsuche werden auch orthographisch ähnliche Termini angezeigt. Die Ähnlichkeit wird anhand von kürzeren Buchstabenfolgen des Suchbegriffs bestimmt, indem diese mit Buchstabenfolgen der Termini der Datenbank verglichen werden. Diese Art der fuzzy-Suche⁴ erlaubt es, auch Begriffe zu finden, die beispielsweise durch Vokalveränderung im Plural von der gesuchten Form abweichen (z.B.: Gut – Transitgüter). Andererseits werden durch diese Art der fuzzy-Suche auch Wortformen gefunden, die mit dem gesuchten Begriff nichts zu tun haben. Sucht man beispielweise nach *atto* (Urkunde) erhält man als Ergebnis auch *contratto* (Vertrag) oder *aspettativa di fatto* (Anwartschaft). Die Ausgabe solcher falscher Treffer ließe sich vermeiden, indem der Abgleich vom Suchwort mit den Termini der Datenbank strenger geführt würde. Schwieriger ist es „falsche fuzzy matches“ auszuschließen, die Beispielen wie denen in Tabelle 4 ähneln. Sucht man

4 Die fuzzy-Suche ist eine „ungenau“ Suche, die es ermöglicht, Wörter zu finden, die dem gesuchten Terminus ähneln. Wenn man *atto* sucht, findet man auch *atto amministrativo*.

zum Beispiel nach *interessi* (Zinsen) werden unter anderem die in Tabelle 4 aufgelisteten Formen gefunden. Der Terminus *conflitto di interessi* (Interessenkonflikt) steht in keiner inhaltlichen Beziehung zu *interessi* und sollte nicht in der Ergebnisliste aufscheinen. Falsche Treffer von diesem Typ auszuschließen ist quasi unmöglich, da sich ausschließlich aus der Bedeutung der Begriffe erschließen lässt, dass sie in keiner inhaltlichen Beziehung zum Suchwort stehen.

<i>interessi</i> _{N-p}	Zinsen _{N-p}	law of obligations
<i>conflitto di interessi</i> _{N-m}	Interessenkonflikt _{N-m}	civil proced.law
	Interessenkonflikt _{N-m}	civil proced.law
	Interessenwiderstreit _{N-m}	civil proced.law
	Widerstreit mit Interessen _{N-m}	civil proced.law
<i>interessi attivi</i> _{N-mp}	Aktivzinsen _{N-p}	civil law
<i>interessi passivi</i> _{N-mp}	Passivzinsen _{N-mp}	civil law
<i>interessi compensativi</i> _{N-mp}	Ausgleichszinsen _{N-mp}	civil law
<i>acquisizione degli interessi</i> _{N-f}	Interessenerhebung _{N-f}	admin.law
<i>interessi usurari</i> _{N-mp}	Wucherzinsen _{N-mp}	law of obligations
<i>saggio degli interessi</i> _{N-m}	Zinssatz _{N-m}	law of obligations
<i>interessi moratori</i> _{N-mp}	Verzugszinsen _{N-mp}	law of obligations
<i>interessi legali</i> _{N-mp}	gesetzliche Zinsen _{N-mp}	law of obligations

Tabelle 4: Ergebnisliste der Suche nach *interessi* (Auszug)

3. 2. Kritische Aspekte der automatischen Termerkennung

Die Funktionsweise eines Terminologieerkennungswerkzeugs ergibt sich aus dem Zusammenwirken von drei Komponenten: der Datenbasis, dem Identifikationsalgorithmus und dem Modell für die Präsentation der Ergebnisse.

Die Datenbasis, das heißt die Terminologiesammlung, auf die sich die Erkennungsfunktion stützt, bildet das Herzstück des Termwerkzeugs. Je mehr Fachausdrücke in der Datenbank erfasst sind, desto mehr Fachausdrücke können im Text erkannt werden. Und je mehr Informationen zu jedem einzelnen Terminus vorhanden sind, desto differenzierter kann der Erkennungsalgorithmus arbeiten und desto reicher können die identifizierten Begriffe annotiert werden.

Der Identifikationsalgorithmus ist die „intelligente“ Komponente des Systems und damit der Motor des Werkzeugs. Der Algorithmus regelt die Erkennung der Termini im Text, dazu gehört sowohl die Erkennung auf Wortebene (Erkennung von Wörtern, die Termini oder Teile von Termini sind) als auch die korrekte Identifizierung auf kontextueller Ebene (Entscheidung, ob Wörter im textuellen Kontext Termini sind). Probleme,

die bei der Erkennung auf Wortebene und auf kontextueller Ebene auftreten, werden im nächsten Abschnitt ausführlich beleuchtet.

Sind alle Termini des untersuchten Textes identifiziert und mit den korrespondierenden Einträgen in der terminologischen Datenbank verlinkt, sorgt das Modell zur Datendarstellung für die benutzerfreundliche Präsentation der Ergebnisse. Es muss unterschieden werden, welche Informationen direkt im Text zu kodieren sind, welche halbversteckt gespeichert werden und welche Informationen nur durch einen sukzessiven Suchschritt recherchierbar sein sollen.

Die identifizierten Fachausdrücke sollten direkt im Text hervorgehoben werden. Auch Unterschiede im Normierungsstatus zählen zu den relevanten Informationen und sind im Text kenntlich zu machen (z.B. durch unterschiedliche Farbgebungen). Weniger vorrangige Informationen können versteckt bleiben und beispielsweise nur beim Überfahren eines Wortes mit der Maus sichtbar werden (in *bistro* sind Übersetzungen auf diese Weise repräsentiert). Detaillierte Informationen zu den Begriffen, wie zum Beispiel der vollständige terminologische Eintrag, würden die Lesbarkeit des Textes deutlich verschlechtern und sollten auf gezielte Abfrage in einem separaten Fenster präsentiert werden.

3. 2. 1. Identifikation von Termini im Text

Ein grundlegendes Problem bei der Erkennung von Fachbegriffen hängt damit zusammen, dass Fachausdrücke in Datenbanken in der Regel in der Grundform abgespeichert sind (Substantive im Nominativ Singular, Verben im Infinitiv, Adjektive im Positiv, usw.), sie im Text hingegen häufig in flektierter oder deklinierter Form vorkommen und demnach ohne weitere Bearbeitung nicht erkannt werden.

Dieses Problem kann gelöst werden, indem in einem ersten Schritt alle gebeugten Wortformen auf ihre Grundformen (Lexeme) zurückgeführt werden, um in einem zweiten Schritt mit den Termini der Datenbank abgeglichen werden zu können. Dieser Vorgang der Rückführung von gebeugten Formen auf Grundformen wird Lemmatisierung genannt (Baker et al. 2006:104).

Im Fall von Phraseologismen stellt sich die Situation schwieriger dar. Die Lemmatisierung jedes einzelnen Wortes ist nicht notwendigerweise zielführend, da Mehrwortausdrücke oft gebeugte Formen enthalten und auch in der Datenbank in dieser Form gespeichert sind. In der Wendung „missbräuchliche Verwendung von echten Siegeln und Geräten“ steht beispielsweise das Wort „Siegel“ im Dativ Plural und der Ausdruck könnte nicht identifiziert werden, wenn nach allen Grundformen gesucht würde. Für Beispiele dieser Art müsste für jedes Wort der Sequenz sowohl die aufgefundene Form als auch deren Grundform beim Abgleich mit der Datenbank in Betracht gezogen werden.

Eine weitere Schwierigkeit im Umgang mit Rechtskollokationen ergibt sich aus der schon genannten Wortsequenz. Kollokationen können in authentischen Texten über größere Textabschnitte verteilt und eventuell durch Einschübe unterbrochen sein, während sie in der Datenbank als lineare Sequenz gespeichert sind. Der Identifikationsalgorithmus muss Einschübe tolerieren, wobei bezüglich der Anzahl und Länge von tolerierten Einschüben abgewogen werden muss, bis zu welchem Ausmaß es sich noch um Termini handeln kann und ab welchem Punkt man das Risiko eingeht auch Textsequenzen als Termini zu identifizieren, die *de facto* keine Termini sind. In der Datenbank könnte zum Beispiel die Kollokation „einen Vertrag unterzeichnen“ eingetragen sein. In einem Text könnten folgende Sätze vorkommen:

- (1) „Der Vertrag wurde am 12.12.2007 unterzeichnet.“
- (2) „Der Vertrag wurde verhandelt und eine Reihe von Urkunden wurden unterzeichnet.“

In Beispiel (1) ist die Kollokation vorhanden und sollte identifiziert werden. Beispiel (2) hingegen enthält die Kollokation nicht, und sollte vom System nicht erkannt werden, obwohl alle Wörter, die zur Kollokation gehören, in dem Beispielsatz vorkommen.

Die Markierung basierend auf Länge und Anzahl der Einschübe vorzunehmen, stellt eine Vereinfachung dar. Um solche Fälle mit hoher Korrektheit zu verarbeiten, müsste eine syntaktische Analyse des Eingabetextes vorgenommen werden.

3.2.2. Homonyme

Wie bereits im Abschnitt zur Termsuche erläutert, stellen Homonyme Schwierigkeiten dar, solange sie nicht in ihre unterschiedlichen Bedeutungen aufgelöst (disambiguiert) werden.

Für die Anwendung der automatischen Termerkennung können zwei Fälle unterschieden werden: Homonyme bei denen eines der beiden Homonyme ein Terminus ist, das andere ein Begriff der Allgemeinsprache und Homonyme bei denen beide Homonyme Termini sind, sie aber unterschiedlichen Fachgebieten angehören.

Typische Beispiele für den ersten Fall sind Begriffe wie *stato*, *fatto*, *dati* sowie *bene* oder „gut“, die sowohl Termini bezeichnen als auch allgemeinsprachliche Perfekt Partizipien oder Adjektive sind. Werden diese Homonyme nicht disambiguiert, kommt es zu folgenden Fehlern: *stato* wird als Terminus markiert (italienische Entsprechung von „Staat“), obwohl es als Partizip Perfekt von *stare* (stehen) auftritt und keine Markierung erhalten sollte; „gut“ wird als Terminus markiert (also als deutsche Entsprechung zum italienischen *bene giuridico*), obwohl es als Adjektiv auftritt und unmarkiert bleiben muss-

te⁵. In all diesen Beispielen werden Wörter der Allgemeinsprache fälschlicherweise der Fachsprache zugeordnet.

Die zuverlässige Disambiguierung mehrdeutiger Wörter ist eine noch unbewältigte Aufgabe der automatischen Sprachverarbeitung (Kennedy 1998:225). Für die oben aufgeführten Hymonyme kann vereinfachend eine Disambiguierung mit Hilfe der Wortklasseninformation (z.B. Adjektiv vs. Substantiv) vorgenommen werden. Dieses Vorgehen setzt aber voraus, dass die Termini in der Datenbank mit Wortklasseninformationen annotiert sind und dass die Wortklassen der Wörter im Text automatisch bestimmt werden können. Eine automatische Wortklassenbestimmung kann von Part-of-Speech-Taggern geleistet werden (Carstensen et al. 2004:224). In anderen Fällen ist die Disambiguierung mittels der Wortklasseninformation jedoch nicht möglich: *parte* wird als Entsprechung zu „Partei“ annotiert, obwohl es im Text in der Bedeutung „Teil“ auftritt und *terzo* wird als Entsprechung zu „Dritter“ annotiert, obwohl es in der Bedeutung von „ein Drittel“ auftritt. Für diese Fälle müssten Regeln gefunden werden, die die Bedeutungen basierend auf dem Satzkontext auflösen. In der Praxis ist es allerdings kaum möglich alle derartigen Fälle korrekt zu behandeln.

Kürzel, die Wörtern der Allgemeinsprache entsprechen (z.B. *SE* für *Servizio educativo* – Erziehungsdienst), sind ebenfalls problematisch und könnten aufgelöst werden, indem man Regeln zur Groß- und Kleinschreibung einfügt.

Auch orthographische Varianten von Wörtern müssen berücksichtigt werden. Im Deutschen können beispielsweise „ss“ und „ß“ alternierend vorkommen. In gleicher Weise können Umlaute gelegentlich umgeschrieben sein („oe“ anstatt „ö“, „ae“ anstatt „ä“, usw.). Die Termerkennung sollte die möglichen Varianten von Umlauten und „ss/ß“ aufeinander abbilden.

Im zweiten oben angeführten Fall muss vermieden werden, dass ein Wort zwar korrekt als Terminus erkannt, aber mit der Bedeutung des falschen Fachgebiets verknüpft wird. Ein Beispiel: Im Text kommt *atto* in der Bedeutung „Rechtshandlung“ vor, aber das System schlägt als mögliche Übersetzung „Urkunde/ schriftlicher Bericht“ vor, was der Bedeutung im verwaltungssprachlichen Kontext entspricht.

In welcher fachspezifischen Bedeutung der Terminus im Text auftritt, richtet sich maßgeblich danach, welchem Fachgebiet der Text zuzuordnen ist. Eine Disambiguierung der Begriffe ist demnach durch die Bestimmung des Fachgebiets zu erreichen. Eine Möglichkeit besteht darin, den Benutzer das Fachgebiet des Textes explizit angeben zu lassen, eine andere Möglichkeit ist es, eine automatische Textklassifizierung vorzunehmen. Die automatische Klassifizierung kann für den Benutzer praktischer sein, birgt aber, wie alle

5 Im Italienischen ist eine Unterscheidung entsprechend der Groß- und Kleinschreibung nicht möglich. Im Deutschen kann diese Unterscheidung auch nicht als zuverlässiges Kriterium verwendet werden, da am Satzanfang alles groß geschrieben wird.

automatischen Verfahren, die Gefahr von Fehlern. In der Termerkennung *bistros* erfolgt die Klassifizierung anhand der im Text vorhandenen Termini. Wenn im Text eine hohe Anzahl an Termini vorkommt, die in der Datenbank mit dem Fachgebiet Familienrecht gekennzeichnet sind, wird dieser Text als Text des Familienrechts gekennzeichnet. Die Probleme, die bei dieser Zuweisung auftreten können, beziehen sich auf jene Texte, die Termini beinhalten, die mehreren Fachgebieten zugeordnet werden können, zum Beispiel Verwaltungsrecht und Arbeitsrecht, oder Verwaltungsrecht und Familienrecht. Ebenso verfälschen Begriffe der Allgemeinsprache, die als Termini erkannt wurden, das Ergebnis.

Enthält ein Text Termini eines fremden Fachgebiets, ist die korrekte Identifizierung des richtigen Homonyms quasi unmöglich. In diesem Fall könnte man einen Zwischenschritt einfügen und verlangen, dass der Benutzer ein Fachgebiet für den Terminus bestimmt. Eine weitere pragmatische Lösung wäre es, dem Benutzer alle Informationen zu beiden Homonymen zur Verfügung zu stellen und ihm so die Möglichkeit zu geben, sich das richtige Homonym auszusuchen.

3.3. Kritische Aspekte der automatischen Termextraktion

Eine Termextraktion zeichnet sich dadurch aus, dass sie aus einem Text eines bestimmten Fachgebiets alle Fachbegriffe extrahiert (Vivaldi & Rodríguez 2007:225), während Begriffe der Allgemeinsprache nicht als Termini ausgegeben werden. Dementsprechend können zwei Typen von Fehlern unterschieden werden: (1) extrahierte Begriffe sind keine Termini und (2) im Text vorhandene Termini werden nicht extrahiert⁶.

Zwischen den zwei Fehlertypen besteht eine wechselseitige Abhängigkeit. Um Fehler vom Typ (2) zu minimieren, muss häufig eine erhöhte Fehlerquote für Typ (1) in Kauf genommen werden und umgekehrt.

Automatische Termextraktionswerkzeuge wie auch die Termextraktion *bistros* stützen sich auf linguistisches oder statistisches Wissen oder verbinden die Informationen in einem hybriden Ansatz (Cabré Castellví et al. 2001:54). Um Termkandidaten zu extrahieren wird die Zusammengehörigkeit komplexer linguistischer Einheiten (*unithood*) sowie deren Termhaftigkeit (*termhood*) bewertet (Kageura & Umino 1996:260)⁷. Da die Programme anders als die Termerkennung darauf ausgelegt sind, unbekannte Termini aufzufinden, sind die Bewertungskriterien immer relativ. Ein Begriff gelangt in die Liste der extrahierten Termkandidaten, sobald seine „Bewertungsnote“ einen Schwellenwert überschritten hat, das heißt sobald es wahrscheinlich genug ist, dass es sich um einen Terminus handelt.

6 Die Fehlerraten werden üblicherweise in den Evaluationsmaßen *precision/ recall* bzw. *noise/ silence* ausgedrückt (vgl. Vivaldi & Rodríguez 2007:229 ff.).

7 Kageura & Umino (1996) definieren *unithood* als Grad der Stabilität syntagmatischer Kombinationen und Kollokationen und *termhood* als Grad zu dem eine linguistische Einheit einen domänenspezifischen Begriff repräsentiert.

Dieser Schwellenwert wird bei der Entwicklung des Termextraktionswerkzeugs festgelegt und je strenger er definiert ist, desto sicherer muss ein Termkandidat sein, um extrahiert zu werden. Ein höherer Schwellenwert führt also zur Minimierung der Fehler vom Typ (1) aber zur Maximierung der Fehler vom Typ (2) (Vivaldi & Rodríguez 2007:229). Wie der Schwellenwert zu definieren ist, sollte im Hinblick auf die Nutzer und deren Ansprüche entschieden werden.

Obwohl die Anpassung des Fehlerverhältnisses durchaus hilfreich ist, führt sie zu keiner echten Verbesserung des Werkzeugs. Um die Gesamtheit der Fehler zu reduzieren, muss der „Motor“ des Systems optimiert werden. Dies kann erreicht werden, indem der Extraktionsalgorithmus auf mehr und auf spezifischere Informationen gegründet wird. Was das im Konkreten bedeutet, hängt von der jeweils implementierten Extraktionsstrategie ab. Ähnlich wie bei der Termerkennung, kann die Integration von Wortklasseninformationen (Verb, Adjektiv, usw.) von Nutzen sein. Theoretische Einsichten zur Natur von Einschüben in Mehrworttermini sowie Variationen durch Flektion und Derivation sind notwendig, um die Extraktion komplexer Termini zu verbessern (Kageura & Umino 1996:282). Auch größere Mengen an Beispieltermini können helfen, neue Regelmäßigkeiten in Termstrukturen zu ermitteln.

4. Ausblick

Anhand des Informationssystems *bistro* haben wir diverse Komponenten zum Zugriff und zur Verwendung elektronisch gespeicherter Terminologiesammlungen vorgestellt. Wir haben diskutiert, welche Aspekte für die automatische Verarbeitung eine besondere Herausforderung darstellen und wie diese Herausforderungen gemeistert werden können.

Zum einen ist deutlich geworden, dass digitalisierte Daten verglichen mit gedruckten Daten sowohl im Hinblick auf die Darstellungsform als auch im Hinblick auf die Integration in computerlinguistische Anwendungen flexibler gehandhabt werden können. Dem Benutzer können unterschiedliche Suchmöglichkeiten geboten werden und die Ergebnisse können mit mehr oder weniger Details präsentiert und miteinander verlinkt werden. Es besteht die Möglichkeit, basierend auf den terminologischen Daten Termwerkzeuge zu erstellen. Für den Einsatz in computerlinguistischen Werkzeugen können die Daten weiter aufbereitet (z. B. mit Wortklasseninformationen annotiert) sein oder auch in ihrer ursprünglichen Form verwendet werden.

Zum anderen soll durch die ausführliche Erläuterung der einzelnen Komponenten beim Nutzer ein besseres Verständnis der Funktionsweise sowie der Grenzen automatischer Anwendungen erreicht werden. Eine kontinuierliche Evaluation der angebotenen Services ist notwendig, um diese optimieren und an neue technische Standards anpassen

zu können. Die oben ausgeführten Veränderungsvorschläge sollen als Ausgangspunkt für technische Verbesserungen der Anwendungen dienen.

Indem wir einen umfassenden Einblick in die funktionellen Zusammenhänge computerlinguistischer Anwendungen geben, wollen wir das Potential elektronischer Ressourcen für den Bereich der Terminologierecherche aufzeigen. Online-Terminologieplattformen sind eine unvergleichbar nutzerfreundliche Informationsquelle, da sie die Funktionalität gedruckter Medien übertreffen und den Anwendern über das Internet jederzeit zur Verfügung stehen.

Bibliographie

- Arntz, Reiner/ Picht, Heribert/ Mayer, Felix (2002): *Einführung in die Terminologearbeit*, 4. Aufl. Hildesheim: Olms.
- Baker, Paul/ Hardie, Andrew/ McEnery, Tony (2006): *A Glossary of Corpus Linguistics*. Edinburgh: Edinburgh University Press.
- Cabré Castellví, M. Teresa/ Estopà Bagot, Rosa/ Vivaldi Palatresi, Jordi (2001): Automatic term detection. A review of current systems. In Bourigault, Didier/ Jacquemin, Christian/ L'Homme, Marie-Claude (Hrsg.): *Recent Advances in Computational Terminology*, Bd. 2. Amsterdam/ Philadelphia: Benjamins, 53-87.
- Carstensen, Kai-Uwe/ Ebert, Christian/ Endriss, Cornelia/ Jekat, Susanne/ Klabunde, Ralf/ Langer, Hagen (Hrsg.) (2004): *Computerlinguistik und Sprachtechnologien. Eine Einführung*, 2. Aufl. Heidelberg: Spektrum Akademischer Verlag.
- De Schryver, Gilles-Maurice (2003): Lexicographers' dreams in the electronic-dictionary age. In *International Journal of Lexicography* 16:2, 143-199.
- eCoLoRe Project. *Glossar zur eContent-Lokalisierung*.
<http://ecolore.leeds.ac.uk/xml/materials/overview/glossary.xml?lang=de>
- Institut für Fachkommunikation und Mehrsprachigkeit (2005): *Ergebnisse der Online-Umfrage zu bistro*.
http://www.eurac.edu/NR/rdonlyres/D17C46E2-4CDA-4C95-83C3-DE2BBC353295/0/bistro_auswertung_final_de.pdf?pu=1
- Kageura, Kyo & Umino, Bin (1996): Methods of automatic term recognition. In *Terminology* 3:2, 259-289.
- Kennedy, Graeme (1998): *An introduction to corpus linguistics*. London/New York: Longman.
- Streiter, Oliver/ Ralli, Natascia / Ties, Isabella/ Voltmer, Leonhard (2004): Bistro, the online platform for terminology management. Structuring terminology without entry structures. In Temmerman, Rita & Knops, Uus (Hrsg.): *Linguistica Antverpiensa*, new series (3-2004) on The translation of domain specific languages and multilingual terminology. Hoger Instituut voor Vertalers en Tolken, Hogeschool Antwerpen, 203-215.

Streiter, Oliver/ Stuflesser, Mathias/ Ties, Isabella (2004): CLE, an aligned Tri-lingual Ladin-Italian-German Corpus. Corpus Design and Interface, LREC 2004, Workshop on *First Steps for Language Documentation of Minority Languages: Computational Linguistic Tools for Morphology, Lexicon and Corpus Compilation*. Lissabon, 24. Mai 2004.

Vivaldi, Jorge & Rodríguez, Horacio (2007): Evaluation of terms and term extraction systems. A practical approach. In *Terminology* 13:2, 225-248.

Wüster, Eugen (1991): *Einführung in die allgemeine Terminologielehre und terminologische Lexikographie*. 3. Aufl. Bonn: Romanistischer Verlag.