

# Eine Methode zur formalen Spezifikation und Umsetzung von Bezeichnungskonventionen für fachkonzeptionelle Informationsmodelle

Patrick Delfmann, Sebastian Herwig, Łukasz Lis, Armin Stein

European Research Center for Information Systems  
Westfälische Wilhelms-Universität Münster  
Leonardo-Campus 3  
48149 Münster

[vorname.nachname]@ercis.uni-muenster.de

**Abstract:** Bei der verteilten Erstellung von fachkonzeptionellen Informationsmodellen können, insbesondere aufgrund der Subjektivierung, beim Vergleich und der anschließenden Zusammenführung von Teilmodellen eine Reihe von Vergleichskonflikten (z. B. Namens-, Detaillierungs- oder Typkonflikte) auftreten, deren Auflösung Gegenstand zahlreicher methodischer Arbeiten ist. Diese betrachten jedoch ausschließlich die Problembehandlung zur Zeit der Konsolidierung. In diesem Artikel wird eine Methode vorgestellt, die Probleme, welche bei der nachträglichen Auflösung von Namenskonflikten aufgrund semantischer Diversitäten auftreten, bereits während der Modellierung verhindert. Die Methode ist generisch und somit für jede Modellierungstechnik und jede natürliche Sprache anwendbar.

## 1 Motivation

Für die erfolgreiche und effiziente Entwicklung von Informationssystemen sind im Vorfeld erstellte fachkonzeptionelle Modelle eine allgemein anerkannte Grundlage [KK84, Ka88]. Durch den gewachsenen Umfang der Projekte werden diese Modelle jedoch in der Regel nicht mehr von einzelnen Personen an einem Ort, sondern sowohl zeitlich als auch räumlich verteilt entwickelt, was eine Effizienzsteigerung gegenüber der Einzelentwicklung mit sich führen soll [Br03]. Außerdem werden zur optimalen Erfassung der fachlichen Anforderungen zunehmend auch Fachvertreter an der Erstellung der Modelle beteiligt, so dass Ergebnisse unterschiedlicher Modellierungsparadigmen und -kenntnisse in die Modelle einfließen [Fi04, WJ04]. Hierdurch entstehen Teilmodelle, die anschließend zu einem Gesamtmodell integriert werden müssen. Diesem Umstand trägt beispielsweise die Entstehung von Entwicklungscommunities wie die Open Model Initiative Rechnung, welche darum bemüht ist, Modelle nicht nur unternehmensintern, sondern auch gemeinschaftlich verteilt über die Unternehmensgrenzen hinweg zu erstellen [FSK07].

Durch die Diversität der Modellierer unterscheiden sich die Modelle häufig in Bezug auf die verwendeten Begriffe sowie in Bezug auf den Abstraktions- und Detaillierungsgrad, wie empirische Studien belegen [HS06]. Diese Missstände führen dazu, dass ein erschwerender Vergleich der Modelle vor einer Zusammenführung notwendig ist, da ansonsten bei der Integration der Modelle im Zuge der Erstellung einer Gesamtbetrachtung des Modellierungsgegenstandes u. a. *Namenskonflikte* (*Homonym-, Synonym- und Abstraktionskonflikte*), *Typkonflikte* und *Detaillierungskonflikte* auftreten können [Pf08]. Um diese Konflikte zu lösen, muss eine aufwändige manuelle Anpassung unter Einbeziehung sämtlicher an dem Projekt beteiligter Modellierer und Fachvertreter vorgenommen werden, was den Gedanken der Effizienzsteigerung ad absurdum führt. Eine Vielzahl von wissenschaftlichen Beiträgen beschäftigt sich mit Vorschlägen zur Durchführung dieser Aufgabe (vgl. Abschnitt 2).

Dieser Beitrag betrachtet die Vermeidung von Namenskonflikten schon vor der Integration der Teilmodelle. Es wird eine Methode vorgeschlagen, mit der schon während der Modellierung verhindert werden kann, dass es bei einer an die Modellierung anschließenden Zusammenführung der Teilmodelle zu Namenskonflikten kommt. Dies geschieht durch ein automatisiertes Anleiten des Modellierers bei der Bezeichnung verwendeter Modellelemente mittels Bezeichnungskonventionen. Diese basieren auf bereits im Vorfeld des Modellierungsprojektes festgelegten Begriffs- und Phrasenstrukturkonventionen. Während Begriffskonventionen durch Begriffsmodelle abgebildet werden können, die neben Konventionen zur Verwendung von Substantiven auch solche zur Verwendung weiterer Wortarten spezifizieren, bietet sich für die Definition von Phrasenstrukturkonventionen die Verwendung der Head-driven Phrase Structure Grammar (HPSG) an [PS94]. Hierbei handelt es sich um eine formale Grammatik zur Spezifikation der Syntax natürlicher Sprachen. Somit ist die Methode einerseits von der verwendeten Modellierungssprache, andererseits aber auch von der verwendeten Landessprache unabhängig.

Der Artikel ist folgendermaßen strukturiert: Zunächst werden in Abschnitt 2 verwandte Arbeiten sowie insbesondere der durch diesen Beitrag zu deckende methodische Entwicklungsbedarf diskutiert. Des Weiteren wird die im Rahmen des Beitrags angewendete Forschungsmethode vorgestellt. In Abschnitt 3 erfolgt die konzeptionelle Spezifikation der Methode zur Vermeidung von Namenskonflikten, deren Anwendung in Abschnitt 4 exemplarisch anhand eines Demonstrators gezeigt wird. Abschnitt 5 vermittelt einen Ausblick auf notwendigen, zukünftigen Forschungsbedarf, der sich u. a. den hier noch nicht behandelten Konfliktarten widmen muss.

## **2 Verwandte und zu leistende Arbeiten; Forschungsmethode**

### **2.1 Verwandte Arbeiten und methodischer Entwicklungsbedarf**

Die Ursprünge der Betrachtung von Namenskonflikten sind in Arbeiten der 1980er und 1990er Jahren zu suchen. Als Motivation wurde damals die Notwendigkeit der Integration von Unternehmensdatenbanken gesehen, wobei in einem ersten Schritt zunächst die

vorhandenen Datenbankschemata in ein konsolidiertes Schema überführt werden müssen [BL84, BLN86, BKK91, LB01]. Bei den in diesem Zusammenhang betrachteten Modellierungssprachen handelt es sich häufig um Dialekte des Entity-Relationship-Modells (ERM) [Ch76]. Die Autoren begegnen der Problematik zumeist über einen Vergleich der Bezeichner der in den Datenbankschemata verwendeten Schemaelemente. Eine Automatisierung des semantischen Vergleichs der Elemente wird durch die Verfasser grundsätzlich ausgeschlossen, so dass immer eine manuelle Anpassung durchgeführt werden muss. Semiautomatisierte Lösungen machen es sich zunutze, dass bei der Modellierung mittels ERM in der Regel ausschließlich einzelne Substantive verwendet werden, was die Komplexität des Vergleichs auf einem relativ geringen Niveau hält. Wesentlich komplexer stellt sich die Aufgabe bei der Verwendung gängiger Prozessmodellierungssprachen wie bspw. Ereignisgesteuerte Prozessketten (EPK) [KNS92] dar, bei denen ganze Satzphrasen üblich sind, die zusätzlich zu Substantiven sämtliche mögliche Wortarten enthalten können. Die oben angeführten Beiträge betrachten diese Problematiken in ihren Ansätzen nicht.

Weitere Ansätze zur Behandlung von Namenskonflikten betrachten die Einbindung von Ontologien [Gr93, Gu98] zum semantischen Vergleich der in den Modellen verwendeten Bezeichner. So können bspw. Domänenontologien an die in den Teilmodellen vorkommenden Begriffe gekoppelt werden, um Beziehungen und Ähnlichkeiten zwischen Elementen zu identifizieren und zu explizieren [Ho07]. Ein Begriffsvergleich auf syntaktischer Ebene erfolgt in [EKO07], wobei mittels eines kombinierten Ähnlichkeitsmaßes automatisiert entschieden wird, ob zwei betrachtete Modellelemente identisch sind. Dieses Vorgehen ist jedoch nicht fehlerfrei, da Homonym- und Synonymkonflikte nur durch eine semantische Betrachtung identifiziert werden können, was eine nachträgliche Überarbeitung notwendig macht. Dieses Problem beruht auch darauf, dass es bei der Kopplung mit einer „allgemein akzeptierten“ Ontologie vorkommen kann, dass syntaktisch identische Begriffe der Teilmodelle während der Modellierung mit einer unterschiedlichen Semantik versehen werden. Ein Algorithmus erkennt dann fälschlicherweise Elemente mit gleichem Bezeichner als identisch und somit integrierbar an [KO05, Sa07]. Es reicht demnach nicht aus, wenn die Modellierer sich in der gleichen Domäne bewegen.

Vielmehr ist es für diesen Fall notwendig, dass sich die Modellierer im Vorfeld des Modellierungsprojekts auf eine gültige Ontologie einigen und die hinterlegten Begriffe während der Modellierung konsequent verwendet werden. Um dies zu gewährleisten, ist eine methodische Unterstützung erforderlich, die dem Modellierer jederzeit den Zugriff auf und den Abgleich der Modelle mit der Ontologie erlauben.

Eine solche Lösung liegt bis dato aber nicht vor. Sollte eine solche Lösung existieren, besteht weiterhin das Problem, dass Begriffe falsch verwendet werden, wenn die durch die Ontologie ursprünglich vorgegebene Semantik eine andere ist, als die vom Modellierer intendierte.

Ein restriktiverer Ansatz von [Pf08] sieht vor, zur Vermeidung der Verwendung falscher Begriffe, diese bereits in die Modellierungssprache zu integrieren und anhand von Prozessbausteinen dem Modellierer den kleinstmöglichen Spielraum für Fehler einzuräumen. Die Semantik der Modellelemente ist diesen dann also inhärent. Dieser Vorschlag

wird vom Autor exemplarisch für die Domäne der öffentlichen Verwaltungen in Form der Prozessmodellierungssprache PICTURE konkretisiert. Namenskonflikte sind in dieser Sprache aufgrund der Existenz von exklusiv 26 spezifischen Prozessmodellbausteinen mit definierter Semantik von vornherein ausgeschlossen. Die Eignung solcher Sprachen ist auf überwiegend lineare Geschäftsprozesse mit klaren Strukturen beschränkt. Als kritisch anzumerken ist, dass eine Anwendung solcher vordefinierter Prozessbausteine in unstrukturierten Bereichen, die eine größere Anzahl von Bausteinen erforderlich machen würden, nicht möglich ist, da der Modellierer unter Umständen eine zu geringe Freiheit hat, den Sachverhalt adäquat darzustellen.

Einen freieren Ansatz verfolgen [Ro96, Ku00], die Satzphrasen als Bezeichner vorschlagen. Die eher praxisorientierten Arbeiten stehen im Zusammenhang mit der Modellierung von Geschäftsprozessen, deren Klarheit eine immanente Anforderung im Rahmen von Reorganisationsprojekten ist. Als Beispiele für solche Phrasenstrukturen führen die Autoren bspw. *<Substantiv> + <Verb im Infinitiv>* – also z. B. „Dienstreisantrag prüfen“ – an. Zur Verbreitung solcher Modellierungskonventionen werden z. B. Methodenhandbücher vorgeschlagen [RSD08]. Die Sicherstellung der *Umsetzung* der in den Handbüchern vermerkten Konventionen wird durch solche Ansätze allerdings nicht adressiert. Folglich besteht die Gefahr, dass solche Konventionen zwar bestehen, aber nicht konsequent eingehalten werden.

Zusätzlich fordern [KR98] Fachbegriffsmodelle – ähnlich wie [KO05, Sa07] Lexika –, die im Vorfeld des Modellierungsprojekts Unternehmensbegriffe eindeutig definieren. Diese sind jedoch, wie bereits bei den Ansätzen der Schemaintegration, lediglich durch Substantive belegt. Einen Schritt weiter gehen die Ansätze von [Pa03, RM05, BKS08]. Diese Autoren schlagen – unabhängig von der eigenen Erstellung von Fachbegriffsmodellen bzw. Lexika – die Anbindung an vorhandene Wörterbücher wie bspw. WordNet [<http://wordnet.princeton.edu>] vor. Hierbei handelt es sich um eine umfangreiche Sammlung englischer Substantive, Verben und Adjektive sowie deren semantischer Beziehungen. Hierbei geht wiederum der Domänenbezug verloren und es entstehen ähnliche Problematiken wie bereits oben für die Nutzung von Ontologien. Weiterhin werden die Überlegungen lediglich als erste Idee angeführt und nicht hinreichend spezifiziert. Bisher sind weder eine methodische Umsetzung noch eine Implementierung dieser Ansätze bekannt.

Zusammenfassend lässt sich der folgende methodische Entwicklungsbedarf ableiten, den es zur Vermeidung von Namenskonflikten beim Vergleich von konzeptionellen Informationsmodellen wie folgt zu begegnen gilt:

- *Spezifikation von Begriffskonventionen bezogen auf sämtliche Wortarten (über Substantive hinaus):* Für die Realisierung wird ein Fachkonzept der Begriffsmodellspezifikation vorgestellt.
- *Spezifikation von Phrasenstrukturkonventionen:* Hierfür wird eine Grammatik vorgestellt, die es erlaubt, Phrasenstrukturen zu spezifizieren und mit gültigen Begriffen zu verwenden.

- *Kopplung der Konventionen mit fachkonzeptionellen Modellierungssprachen:* Eine fachkonzeptionelle Spezifikation zur Integration der Bezeichnungskonventionen mit Modellierungssprachen soll diesem Mangel begegnen.
- *Überwachung der Einhaltung dieser Konventionen:* Hierfür wird ein Verfahren vorgestellt, das diese methodische Unterstützung ermöglicht.

## 2.2 Forschungsmethode

Der vorliegende Beitrag folgt dem Design-Science-Forschungsansatz [He04], der im Kern die Konstruktion von *wissenschaftlichen Artefakten* (1) wie z. B. Methoden, Sprachen und Softwarewerkzeugen zum Inhalt hat. Im Rahmen einer solchen Forschungsmethode ist sicherzustellen, dass die durchgeführten Arbeiten ein *relevantes Thema* (2) behandeln. Diese Relevanz ist zu belegen. Darüber hinaus ist sicherzustellen, dass die konstruierten Forschungsartefakte einen *innovativen Beitrag* (3) zur bereits in der Disziplin vorhandenen Wissensbasis darstellen, d. h., dass vergleichbare Lösungen nicht bereits im Vorfeld existieren. Die konstruierten Artefakte sind im Rahmen eines Design Science-Ansatzes zu *evaluieren* (4), um sicherzustellen, dass die gesetzten Forschungsziele tatsächlich erreicht werden. Der Charakter eines Design Science-Ansatzes ergibt sich für diesen Beitrag aus der Konstruktion des *wissenschaftlichen Artefakts* (1) in Form einer Methode zur Vermeidung von Namenskonflikten in der konzeptionellen Informationsmodellierung. Dieses Thema stellt ein *relevantes (2) Forschungsfeld* dar, wie sich aus der Konsolidierungsproblematik ergibt, die durch die verteilte Modellierung induziert wird (vgl. Abschnitt 1). Die *Innovation* (3) zeigt sich darin, dass Lösungen für die diskutierten Probleme beim Vergleich von Informationsmodellen bisher nicht vorliegen und hier entsprechende Lösungsansätze entwickelt werden (vgl. Abschnitt 3). Die *Evaluation* (4) der entwickelten Methode erfolgt anhand eines Demonstrators, der die grundsätzliche Anwendbarkeit illustriert (vgl. Abschnitt 4).

## 3 Fachkonzept

Zur Unterstützung der formalen Spezifikation und automatisierten Durchsetzung von Bezeichnungskonventionen wird im Folgenden das Vorgehen der konventionsunterstützten verteilten Modellierung beschrieben und die notwendigen fachkonzeptionellen Spezifikationen vorgenommen.

### 3.1 Vorgehensmodell

Die Syntax der in einem bestimmten Modellierungskontext (z. B. Domäne, Projekt, Unternehmen) verwendeten Bezeichnungen leitet sich im Allgemeinen von zwei Komponenten ab. Zum einen findet eine Grammatik einer konkreten natürlichen Sprache Anwendung, so dass generell nur grammatikalisch korrekte Phrasen gebildet werden dürfen. Zum anderen wird diese natürlichsprachliche Grammatik weiter an die Belange des konkreten Anwendungskontexts angepasst, indem domänenspezifische Begriffs- und

Phrasenstrukturkonventionen in die Syntax einfließen (vgl. Abbildung 1). Bei der Bezeichnung „Prüfe eingehenden Dienstreiseantrag“ sorgt bspw. die natürlichsprachliche – in diesem Fall deutsche – Grammatik für die Kongruenz (Übereinstimmung) des Numerus, Genus und Kasus zwischen den Wortformen „prüfe“ und „Dienstreiseantrag“. Gleichzeitig tragen die domänenspezifischen Bezeichnungskonventionen dazu bei, dass nur die zugelassenen Begriffe verwendet werden (nur „Dienstreiseantrag“, nicht „Dienstreiseformular“) und dass die Struktur der Bezeichnung der erlaubten Phrasenstrukturkonvention entspricht (<Verb im Imperativ> + <Substantiv im Singular>).

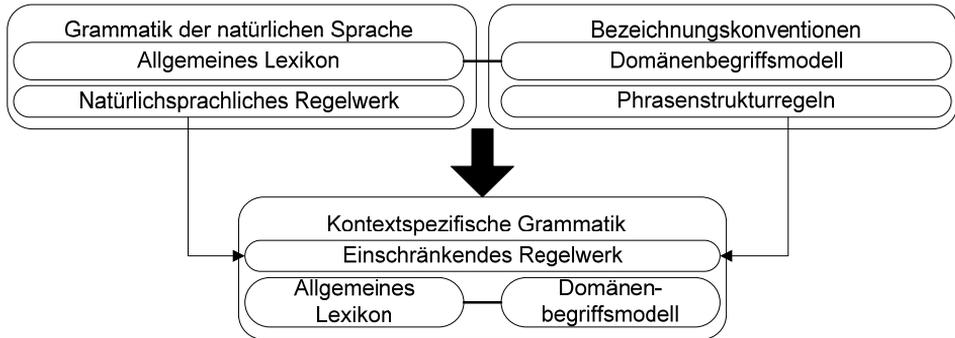


Abbildung 1: Anpassung der natürlichen Sprache durch anwendungsbezogene Bezeichnungskonventionen

Die Grammatik der für Bezeichnungen verwendeten natürlichen Sprache ist durch einen Formalismus beschrieben, der in der Regel aus einem Lexikon und einem sprachlichen Regelwerk besteht [Ca01]. Diese natürlichsprachliche Grammatik ist durch domänenspezifische Bezeichnungskonventionen, die sich wiederum aus Begriffs- und Phrasenstrukturkonventionen zusammensetzen, zu ergänzen. Begriffskonventionen werden mit Hilfe eines Domänenbegriffsmodells formalisiert, welches an das Lexikon der natürlichen Sprache anknüpft. Des Weiteren werden die gültigen Phrasenstrukturregeln mittels sogenannter Phrasentypen definiert. Somit wird die natürliche Sprache durch eine kontextspezifische Grammatik gezielt auf die im Modellierungsprojekt gültige Menge von Bezeichnungen angepasst. Dies erlaubt die spätere Überprüfbarkeit der Bezeichnungen und die Durchsetzung von Strukturkonventionen. Der fachkonzeptionelle Aufbau der Spezifikation von Bezeichnungskonventionen wird im Abschnitt 3.2 vorgestellt.

Das Domänenbegriffsmodell legt die für den konkreten Anwendungszweck erlaubten Begriffe fest und kann entweder von Grund auf neu aufgebaut bzw. aus ggf. vorhandenen Begriffsmodellen generiert werden. Im Gegensatz zu herkömmlichen Begriffsmodellen, die ausschließlich Substantive beinhalten, ist das Begriffsmodell zusätzlich auf die Wortarten der Verben und Adjektive auszudehnen. Alle übrigen Wortarten sind weitgehend domänenunabhängig, so dass deren Aufnahme in das Begriffsmodell nicht notwendig ist. Diese sind aber weiterhin im Lexikon der natürlichen Sprache enthalten. Das Domänenbegriffsmodell ist über Synonym-, Homonym- und Wortbildungsbeziehungen mit dem Lexikon und ggf. weiteren computerlinguistischen Diensten zu verknüpfen. Da solche Dienste für verschiedene natürliche Sprachen bereits vorliegen und wie z. B. Canoo [<http://canoo.net>], GermaNet [<http://www.sfs.uni-tuebingen.de/lsd>] und

Wortschatz-Lexikon [<http://wortschatz.uni-leipzig.de>] teilweise Online-Zugriff bieten, ist deren technische Anbindung grundsätzlich möglich. Im Falle einer späteren Verletzung der Begriffskonventionen durch den Modellierer können somit bspw. synonyme, erlaubte Begriffe ermittelt und vorgeschlagen werden. Zumindest das Domänenbegriffsmodell ist mit textuellen Beschreibungen der Begriffe zu hinterlegen, so dass die Bedeutung eines jeden Begriffs durch die Modellierer nachgeschlagen werden kann. Es ist darauf zu achten, dass sich ein Begriffsmodell während seines Einsatzes nicht verändern darf, um die Konsistenz bisheriger Anwendungen zu erhalten. Sollen Online-Dienste verwendet werden, ist hier bspw. die Möglichkeit einer lokalen Kopie bzw. eines versionierten Dienstes zu prüfen.

Die Spezifikation der Bezeichnungskonventionen muss für jeden Anwendungskontext (z. B. Domäne, Modellierungsprojekt, Unternehmen) einmalig durchgeführt werden, wobei ggf. bestehende Konventionen wiederverwendet werden können. Dies geschieht in Abhängigkeit von Modellierungssprachen, da für diese jeweils unterschiedliche Konventionen denkbar sind. Beispielsweise werden Funktionen einer EPK mit *Verrichtungen* bezeichnet, während Bezeichnungen von Ereignissen *Zustände* angeben [KNS92]. Für jeden Modellelementtyp sind im Folgenden die Bezeichnungskonventionen in Form von Phrasentypen anzugeben (z. B. für Funktionen in EPK <Substantiv> + <Verb im Infinitiv> sowie für Ereignisse <Substantiv> + <Hilfsverb 3. Person> + <Verb im Partizip Perfekt> [Ro96, Ku00]), wobei Alternativen je Elementtyp möglich sind. Um eine praktische Umsetzbarkeit zu gewährleisten, müssen die Phrasenstrukturen in einer solchen Form definiert werden, die sich dem Formalismus der natürlichsprachlichen Grammatik vereinbaren lässt.

Die während der Modellierung eingegebenen Bezeichnungen werden durch einen Automatismus direkt gegen die kontextspezifische Grammatik geprüft, wobei zwei Aspekte berücksichtigt werden. Zum einen ist zu untersuchen, ob die Struktur der vom Modellierer gewählten Bezeichnung den einschränkenden Regeln der zweckbezogenen Grammatik entspricht. Zum anderen ist zu prüfen, ob die gewählten Begriffe durch die Grammatik zugelassen sind. Handelt es sich bei einem Begriff um ein Substantiv, Verb oder Adjektiv – d. h. um Wortarten, die im Begriffsmodell vertreten sind – wird überprüft, ob das entsprechende Wort im Begriffsmodell erlaubt wurde. Vertreter sonstiger Wortarten werden gegen das natürlichsprachliche Lexikon geprüft. Im gesamten Prüfprozess finden computerlinguistische Methoden und Werkzeuge (u. a. Grammatikformalismen und darauf basierende Parser) Anwendung.

Sind die Begriffe und ihre Struktur gültig, so gilt die gewählte Modellelementbezeichnung als konform mit der einschränkenden kontextspezifischen Grammatik. Im Falle der Verletzung eines oder beider Aspekte werden ausgehend von der grammatikalischen Struktur der Bezeichnung und den vom Modellierer gewählten Wörtern alternative Phrasenstrukturen oder Begriffe angeboten, die mit den Bezeichnungskonventionen konform sind. Die Entscheidung, ob die angebotenen Phrasen als Bezeichnung ausgewählt werden, liegt beim Modellierer. Dabei kann dieser die Bedeutungen der einzelnen Begriffe aus den jeweiligen Beschreibungen der Begriffsmodelle entnehmen und daraufhin entscheiden, ob diese zu seiner intendierten Bezeichnung passen. Alternativ hat der er die

Möglichkeit, einen der vorgeschlagenen Phrasentypen als Schablone zu wählen und die darin enthaltenen Platzhalter mit erlaubten Wörtern zu füllen.

### 3.2 Modellierungstechnische Spezifikation

Die modellierungstechnische Spezifikation liefert die Grundlage für die Deklaration und Durchsetzung von Bezeichnungskonventionen und wird mit Hilfe von ERM in (min,max)-Notation [SS83] modelliert. Jedem *Elementtyp* einer Modellierungssprache muss mindestens eine *Phrasenstrukturkonvention* zugeordnet werden, welche die Form eines *Phrasentyps* hat. Ein Phrasentyp beschreibt den strukturellen Aufbau einer Phrase, die als Bezeichnung verwendet werden kann. Er setzt sich rekursiv zusammen aus weiteren Phrasentypen oder *Worttypen*, die atomare Elemente einer Phrase sind und als Platzhalter für die Konkretisierung durch Wörter fungieren. Ein Beispiel für einen Worttyp ist *<Substantiv im Nominativ Singular>*, ein Beispiel für einen Phrasentyp ist *<Substantiv im Nominativ Singular>+<Verb im Infinitiv>*. Die Zusammensetzung von Phrasentypen wird durch den *Phrasenstrukturkonventions-Aufbau* spezifiziert, der einem Phrasentyp entweder Subphrasentypen oder Worttypen mit ihrer *Position* im übergeordneten Phrasentyp zuordnet (vgl. Abbildung 2).

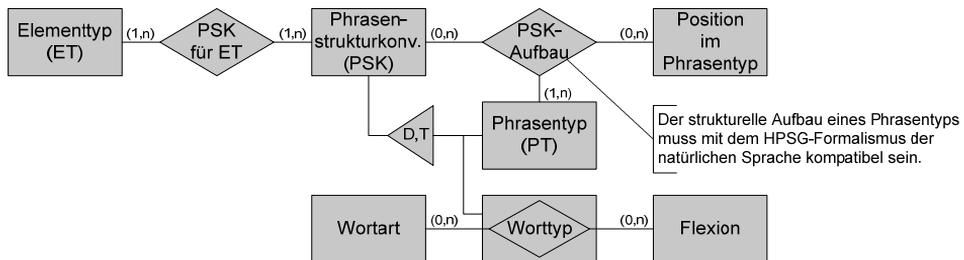


Abbildung 2: Spezifikation von Phrasenstrukturkonventionen

Ein Worttyp spezifiziert die gewünschte *Wortart* (*Substantiv, Verb, Adjektiv, Adverb, Artikel, Pronomen, Präposition, Konjunktion* oder *Numerale*) sowie die erwartete *Flexion*, d. h. die Beugungsform. Als Beugungsformen kommen u. a. *Kasus, Numerus, Tempus, Genus, Modus, Person* und *Komparation* in Frage, die i. d. R. kombiniert werden. Zu beachten ist, dass nicht jede Beugungsform auf jede Wortart anwendbar ist. Mit der in einem Phrasentyp enthaltenen sachlogischen Folge von Worttypen, deren Wortart und Flexion bestimmt ist, lassen sich beliebige Phrasenstrukturkonventionen definieren. Hierbei ist zu beachten, dass die deutsche Sprache oder auch jede andere beliebige natürliche Sprache einer Syntax unterworfen ist, die der Spezifikation von Bezeichnungskonventionen als Grundlage dient.

Deklarierte Strukturregeln sind dabei als Ergänzung der zu Grunde liegenden deutschen Sprachsyntax anzusehen und schränken demgemäß die Bezeichnungsfreiheit des Modellierers, die durch die natürliche Sprache gegeben ist, zusätzlich ein. Um den Abgleich zwischen der natürlichen Sprache und den anwendungsbezogenen Phrasenstrukturregeln zu ermöglichen, wird gefordert, dass beide Syntaxspezifikationen in kompatiblen Formalismen vorliegen (vgl. Annotation in Abbildung 2). Somit ist es möglich, direkt bei

der Spezifikation der Strukturregeln deren Konformität mit der natürlichen Sprache zu überprüfen und eventuelle Widersprüche zu signalisieren.

Zur formalen Beschreibung der Syntax natürlicher Sprachen sind aus dem Bereich der Computerlinguistik unterschiedliche Ansätze bekannt. Die Gruppe der Unifikationsgrammatiken bezeichnet eine Klasse derartiger Ansätze, welche derzeit eine weite Verbreitung und Akzeptanz erfahren [Ca01]. Diesen Ansätzen ist gemein, dass die Deklaration grammatikalischer Regeln stets in Bezug auf Phrasenstrukturen sowie deren Bestandteile – Phrasentypen als komplexe Einheiten oder Wörter als atomare Einheiten – geschieht. Die Formalisierung von Grammatikregeln erfolgt hierbei unter Verwendung von Merkmalsstrukturen. Eine Merkmalsstruktur repräsentiert eine Menge hierarchisch strukturierter Merkmal-Wert-Kombinationen, die die grammatikalischen Eigenschaften einer Phrasenstruktur sowie deren Bestandteile beschreiben. Die sich hierbei ergebenden Phrasenstrukturregeln geben an, welche grammatikalischen Eigenschaften und welche Position Wörter oder Phrasentypen als einzelne Bestandteile einer Phrasenstruktur einnehmen müssen, um den syntaktischen Sprachregeln zu entsprechen, denen diese Phrasenstruktur unterliegt. Exemplarisch kann eine Phrasenstruktur aus einem *Phrasentyp* bestehen, welcher sich aus zwei Worttypen zusammensetzt, die über die Merkmale *Wortart* und *Flexion* beschrieben werden (vgl. Abbildung 3). Das Merkmal *Flexion* setzt sich abhängig vom Worttyp aus weiteren Merkmalen zusammen. Für die Wortart *Verb* kann es sich um die Merkmale *Person*, *Modus* und *Tempus*, für *Substantiv* um *Numerus*, *Genus* und *Kasus* handeln. Die hierbei angegebene Folge der Worttypen markiert deren Position innerhalb der Phrasenstruktur.

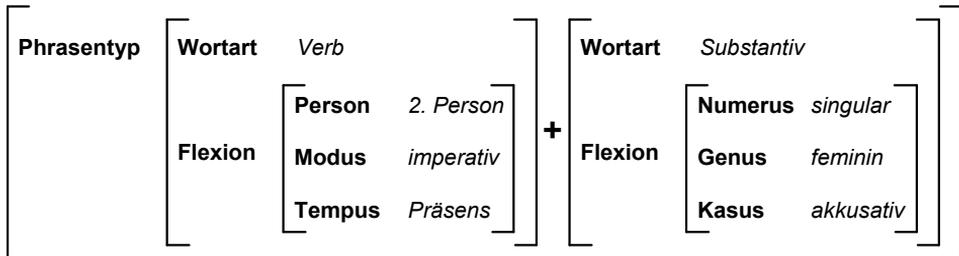


Abbildung 3: Exemplarische Darstellung einer Phrasenstruktur

Mit der *Head-driven Phrase Structure Grammar (HPSG)* [PS94] ist ein weithin etablierter Vertreter der Unifikationsgrammatiken gegeben, der bei der formalen Spezifikation von Sprachregeln dem grundlegenden Prinzip dieser Klasse von Grammatiken folgt. Hierzu bedient sich die HPSG neben einem grammatikalischen Regelwerk insbesondere einem Lexikon. Auf Grundlage von HPSG ist bereits eine Anzahl formaler Sprachspezifikationen verfügbar. Die formale Spezifikation der deutschen Sprachsyntax in HPSG liefert bspw. [Mu99].

Zur Sicherung der Syntaxkonformität wird diese Spezifikation bei der Definition von Phrasenstrukturkonventionen zu Grunde gelegt. Die Formalisierung definierter Konventionen erfolgt zu diesem Zweck ebenfalls in HPSG. Folglich werden die mit den Phrasenstrukturkonventionen verbundenen Grammatikregeln sowie zugehörige Lexikoneinträge mithilfe von Merkmalsstrukturen formalisiert. Die zugehörigen Lexikoneinträge

umfassen hierbei die zur Verfügung stehenden Wörter mit ihren grammatikalischen Eigenschaften. Die erlaubte Form eines Wortes sowie die Position, die es innerhalb einer Phrasenstruktur einnehmen darf, werden dabei durch die entsprechenden Merkmal-Wert-Kombinationen der Phrasenstrukturkonvention vorgegeben. In Bezug sowohl auf die Konformitätsprüfung im Rahmen der Spezifikation von Phrasenstrukturkonventionen als auch auf die spätere Überprüfung der bei der Modellierung angegebenen Bezeichnungen sind somit für HPSG verfügbare Parsing-Methoden wie bspw. das *Babel-System* [Mu96] anwendbar.

Unflektierte Wörter werden unabhängig von ihrer Wortart als *Lexeme* bezeichnet (bspw. das Verb „prüfen“), flektierte Wörter als *Wortform* (z. B. die 3. Person Singular „prüft“ des Lexems „prüfen“). Die bei der Modellierung verwendeten Wortformen (auf Instanzebene) sind ihren Wortarten und Flexionsformen, d. h. ihren Worttypen (auf Typebene) zuzuordnen (vgl. Abbildung 4).

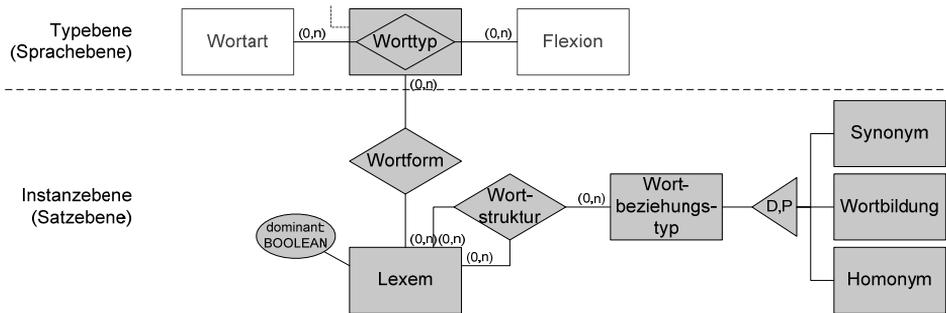


Abbildung 4: Spezifikation von Begriffskonventionen

Die Kombination aus Lexem und Worttyp ergibt damit die *Wortform*. Für die Spezifikation von Begriffsmodellen werden erlaubte Wörter in Form von Lexemen abgelegt und über diverse Beziehungstypen verknüpft. So lässt sich definieren, welche Lexeme sich zueinander *synonym* oder *homonym* verhalten bzw. durch *Wortbildung* aus anderen Lexemen gebildet werden (bspw. wird das Substantiv *Rechnungsprüfung* aus den Substantiven *Rechnung* und *Prüfung* komponiert, welche ihrerseits durch Substantivierung aus *rechnen* bzw. *prüfen* entstehen). In der sich hieraus ergebenden *Wortstruktur* werden die Begriffe der Modellierungsdomäne abgelegt und als *dominant* gekennzeichnet. Dies bewirkt, dass beim Auftreten synonyme Begriffe derjenige präferiert wird, der als *dominant* gekennzeichnet wurde.

Über die Beziehungstypen werden allgemein verfügbare Lexika und Dienste (vgl. Abschnitt 3.1) an das Domänenbegriffsmodell angekoppelt, um Modellierern bei Verwendung eines für die Domäne ungültigen Wortes eine gültige Alternative aus dem Domänenbegriffsmodell anbieten zu können.

## 4 Praktische Umsetzung

Zur Überprüfung der Umsetzbarkeit der fachkonzeptionellen Methode sowie zu Präsentationszwecken wurde eine prototypische Implementierung in Form eines Demonstrators realisiert. Der Funktionsumfang des Demonstrators orientiert sich an dem oben angeführten Vorgehensmodell.

Vor der eigentlichen Modellierung müssen zunächst das Domänenbegriffsmodell sowie das verwendete kontextspezifische Regelwerk erzeugt werden. Abbildung 5 visualisiert das Anlegen von Verben, wobei lediglich die Flexionsform des Infinitivs angegeben werden muss.

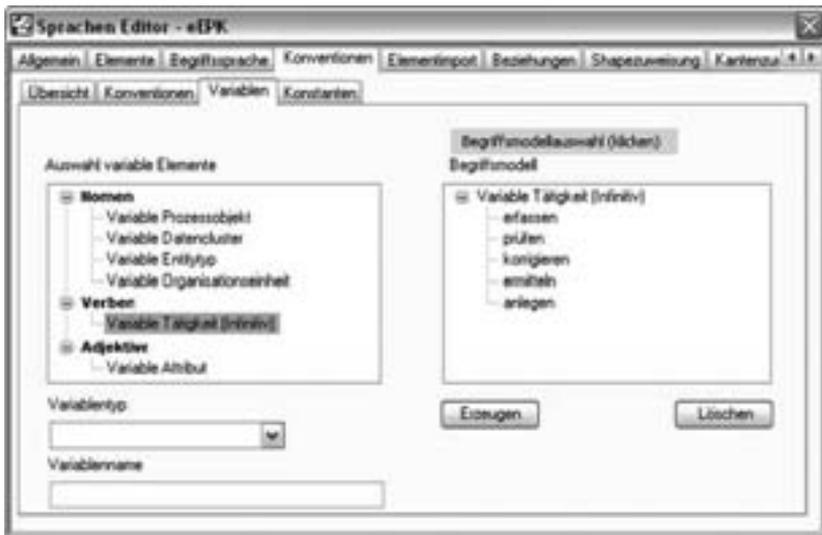


Abbildung 5: Spezifikation von Begriffskonventionen (Verben)

Für die spätere Verwendung weiterer Flexionsformen wird entweder auf externe Lexika mit hinterlegten Flexionsformen zurückgegriffen oder automatisiert eine entsprechende Form gemäß der Grammatik gebildet. Neben dem manuellen Anlegen der Begriffe ist es auch möglich, diese aus entsprechend formatierten Dateien – bspw. aus frei verfügbaren Lexika – zu importieren.

Die Oberfläche zum Anlegen von Substantiven (vgl. Abbildung 6) entspricht in ihrem Aufbau derjenigen für das Anlegen von Verben. Hierbei ist es jedoch zusätzlich möglich, durch Drag & Drop Beziehungen wie „ist Synonym für“ oder „ist ein(e)“ zwischen Begriffen zu erstellen.



Abbildung 6: Spezifikation von Begriffskonventionen (Substantive)

Im nächsten Schritt werden Phrasenstrukturkonventionen festgelegt und den Sprachelementen, für die sie gültig sind, zugeordnet. Abbildung 7 zeigt exemplarische Phrasenstrukturkonventionen für Ereignisse der EPK. In dem hier präsentierten Beispiel wurde für Bereitstellungsereignisse die Struktur *<Substantiv> + <„ist“> (+ <„nicht“>) + <Verb im Partizip Perfekt>* und für Auslöseereignisse die Form *<Substantiv> + <„ist“> (+ <„nicht“>) + <Verb im erweiterten Infinitiv>* gewählt.



Abbildung 7: Spezifikation von Phrasenstrukturkonventionen

Ausgehend von den spezifizierten Bezeichnungskonventionen wird nun während der Modellierung geprüft, ob die vom Modellierer gewählten Bezeichnungen in ihrer Struktur und in der Begriffsauswahl den Vorgaben entsprechen (vgl. im Folgenden Abbildung 8). Eine korrekte Bezeichnungsvergabe bleibt für den Modellierer unbemerkt. Erkennt der Algorithmus jedoch eine Konventionsverletzung, wird die eingegebene Bezeichnung analysiert. Identifizierte Wörter werden ausgehend von ihrer Flexionsform in ihre Grundform überführt und ggf. durch ihr konventionstrees Synonym, das im Domänenbegriffsmodell als dominant markiert wurde, ersetzt. Ein nicht verzeichnetes Wort wird derzeit nicht akzeptiert.

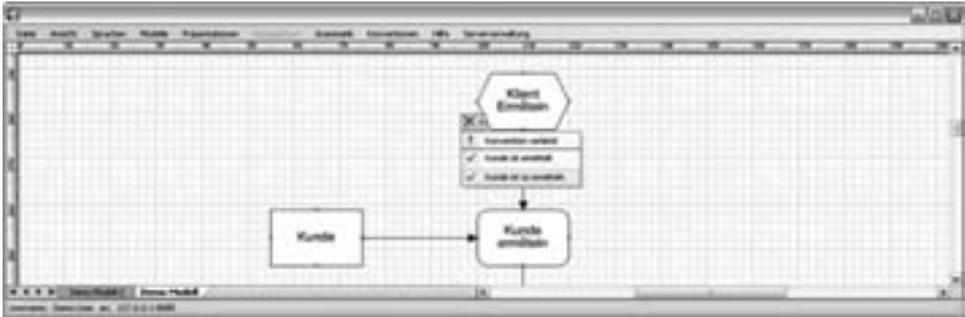


Abbildung 8: Anleitung des Modellierers zur korrekten Bezeichnung

Die Menge der Lexeme, die durch dieses Vorgehen identifiziert werden konnten, werden dem Modellierer – ggf. nach wiederholter Flexion – als Alternativen angeboten. Im Beispiel werden statt der Bezeichnung „Klient ermitteln“ die Vorschläge „Kunde ist zu ermitteln“ und „Kunde ist ermitteln“ generiert. Der Modellierer entscheidet, welche der angebotenen Alternativen er wählt bzw., ob er die angebotenen Alternativen verwirft und eine neue Bezeichnung wählt.

## 5 Fazit und Forschungsperspektiven

Im vorliegenden Beitrag wurde die methodische Verankerung von Bezeichnungskonventionen im Rahmen der fachkonzeptionellen Modellierung vorgestellt. Diese Methode dient als Ansatz für eine Verbesserung der Integrationsmöglichkeiten verteilt erstellter Informationsmodelle. Zwei Bestandteile ermöglichen den in Wissenschaft und Praxis bekannten und in Abschnitt 2 vorgestellten Integrationsproblemen zu begegnen. Zum einen wird durch die bereits im Vorfeld stattfindende Erstellung von Bezeichnungsrestriktionen die während des Modellierungsprojekts verfügbaren Bezeichnungsmöglichkeiten reduziert, sodass bei einer späteren Konsolidierung der Teilmodelle bereits Informationen über den Inhalt der Modellelemente verfügbar sind. Zum anderen wird durch die stete Überprüfung der Eingaben des Modellierers sichergestellt, dass die erstellten Modelle diesen Anforderungen entsprechen. Ein erster Demonstrator belegt die Durchführbarkeit der Methode.

Grenzen der Methode liegen in dem nicht zu vernachlässigenden Aufwand, der mit der Spezifikation der notwendigen Begriffsmodelle, deren Verknüpfung mit allgemeinen Lexika sowie der Festlegung der Phrasenstrukturkonventionen einhergeht. Daher wird die Methode trotz der Wiederverwendbarkeit dieser Artefakte wohl vorwiegend Anwendung in großen bis sehr großen Modellierungsprojekten und Begriffsgemeinschaften finden.

Als zukünftiger Forschungsbedarf lässt sich zunächst die Weiterentwicklung der Methode in Hinsicht auf die Toleranz im Umgang mit fehlerhaften oder fehlenden Begriffen prognostizieren. Es wird bspw. zu klären sein, ob das Domänenbegriffsmodell zur Laufzeit erweitert werden darf oder nicht, und wenn ja, welche Konsequenzen dies für evtl. bereits bestehende Teilmodelle mit sich bringt. Der Demonstrator soll in weiteren Arbeiten verbessert werden, so dass auch die Anbindung an Online-Dienste ermöglicht werden kann. In diesem Rahmen ist auch die technisch fehlerfreie Integrationsfähigkeit bestehender Dienste zu prüfen.

Die Ergebnisse aus quantitativen empirischen Evaluationen in Forschung und Lehre werden im Sinne der Design Science ebenfalls in die Verfeinerung der Methode einfließen. Insbesondere ist zu belegen, dass die verfolgte Zielsetzung der Vereinfachung der Modellkonsolidierung im Rahmen der verteilten Modellierung auch erreicht werden kann. Weiterhin ist im Rahmen von qualitativen Untersuchungen festzustellen, ob zwei Phrasen, die strukturell wie inhaltlich identisch aufgebaut und deren Wortbedeutungen ebenfalls bekannt sind (d. h. es handelt jeweils sich um die gleiche Phrase), von verschiedenen Akteuren im Rahmen der verteilten Modellierung auch als gleichbedeutend empfunden werden.

## Literaturverzeichnis

- [BL84] Batini, C., Lenzerini, M.: A Methodology for Data Schema Integration in the Entity Relationship Model, in: IEEE Transactions on Software Engineering, 10 (1984) 6, S. 650–663.
- [BLN86] Batini, C., Lenzerini, M., Navathe, S. B.: A Comparative Analysis of Methodologies for Database Schema Integration, in: ACM Computing Surveys, 18 (1986) 4, S. 323–364.
- [BKK91] Bhargava, H. K., Kimbrough, S. O., Krishnan, R.: Unique Name Violations, a Problem for Model Integration or You Say Tomato, I Say Tomahto, in: ORSA Journal on Computing, 3 (1991) 2, S. 107-120.
- [BKS08] Bögl A., Kobler M., Schrefl M.: Knowledge Acquisition from EPC Models for Extraction of Process Patterns in Engineering Domains, in: Proceedings der Multikonferenz Wirtschaftsinformatik 2008 (MKWI 2008), München 2008.
- [Br03] vom Brocke, J.: Referenzmodellierung. Gestaltung und Verteilung von Konstruktionsprozessen, Berlin 2003.
- [Ca01] Carstensen, K. U. et al.: Computerlinguistik und Sprachentheorie – Eine Einführung, Heidelberg, Berlin 2001.
- [Ch76] Chen, P. P.-S.: The Entity-Relationship Model – Toward a Unified View of Data, in: ACM Transactions on Database Systems, 1 (1976) 1, S. S. 9-36.

- [EKO07] Ehrig, M., Koschmider, A., Oberweis, A.: Measuring Similarity between Semantic Business Process Models, in: Proceedings of the Fourth Asia-Pacific Conference on Conceptual Modelling (APCCM) 2007, Ballarat 2007.
- [Fi04] Fischer, G. et al.: Meta-design: a manifesto for end-user development, in: Communications of the ACM, 47 (2004) 9, S. 33-37.
- [FSK07] Frank, U., Strecker, S., Koch, S.: „Open Model“ – ein Vorschlag für ein Forschungsprogramm der Wirtschaftsinformatik, in: A. Oberweis, C. Weinhardt, H. Gimpel, A. Koschmider, V. Pankratius, B. Schnizler (Hrsg.), eOrganisation: Service-, Prozess-, Market-Engineering, 8. Internationale Tagung Wirtschaftsinformatik, Band 2, Karlsruhe 2007, S. 217-234.
- [Gr93] Gruber, T. R.: A Translation Approach to Portable Ontology Specifications, in: Knowledge Acquisition, 5 (1993) 2, S. 199-220.
- [Gu98] Guarino, N.: Formal Ontology and Information Systems, in: N. Guarino (Hrsg.), Proceedings of the 1st International Conference on Formal Ontologies in Information Systems, Trento, 1998, S. 3-15.
- [HS06] Hadar, I., Soffer, P.: Variations in conceptual modeling: classification and ontological analysis, in: JAIS, 7 (2006) 8, S. 568-592.
- [He04] Hevner, A. R. et al.: Design Science in Information Systems Research, in: MIS Quarterly, 28 (2004) 1, S. 75-105.
- [Ho07] Höfferer, P.: Achieving business process model interoperability using metamodels and ontologies, in: H. Österle, J. Schelp, R. Winter (Hrsg.): Proceedings of the 15th European Conference on Information Systems (ECIS 2007), St. Gallen 2007, S. 1620-1631.
- [Ka88] Karimi, J.: Strategic Planning for Information Systems: Requirements and Information Engineering Methods, in: Journal of Management Information Systems, 4 (1988) 4, S. 5-24.
- [KNS92] Keller, G., Nüttgens, M., Scheer, A.-W.: Semantische Prozeßmodellierung auf der Grundlage „Ereignisgesteuerter Prozeßketten (EPK)“, in: A.-W. Scheer (Hrsg.), Veröffentlichungen des Instituts für Wirtschaftsinformatik, Heft 89, Saarbrücken 1992.
- [KO05] Koschmider, A., Oberweis, A.: Ontology Based Business Process Description. In: Enterprise Modelling and Ontologies for Interoperability, Proceedings of the Open Interop Workshop on Enterprise Modelling and Ontologies for Interoperability, Collocated with CAiSE'05 Conference, Porto 2005.
- [KK84] Kottemann, J. E., Konsynski, B. R.: Information Systems Planning and Development: Strategic Postures and Methodologies, in: Journal of Management Information Systems, 1 (1984) 2, S. 45-63.
- [Ku00] Kugeler, M.: Informationsmodellbasierte Organisationsgestaltung. Modellierungskonventionen und Referenzvorgehensmodell zur prozessorientierten Reorganisation, Berlin 2000.
- [KR98] Kugeler M., Rosemann, M.: Fachbegriffsmodellierung für betriebliche Informationssysteme und zur Unterstützung der Unternehmenskommunikation. In: Informationssystem Architekturen. Fachausschuss 5.2 der Gesellschaft für Informatik e. V. (GI), 5 (1998) 2, S. 8-15.
- [LB01] Lawrence, R., Barker, K.: Integrating Relational Database Schemas using a Standardized Dictionary, in: Proceedings of the 2001 ACM symposium on Applied computing (SAC), Las Vegas 2001.
- [Mu96] Müller, S.: The Babel-System – An HPSG-Fragment for German, a Parser and a Dialog Component, in: Proceedings of the 4th International Conference on the Practical Application of Prolog, London 1996.
- [Mu99] Müller, S.: Deutsche Syntax deklarativ – Head-Driven Phrase Structure Grammar für das Deutsche, Tübingen 1999.

- [Pa03] Palapoli, L. et al.: DIKE: a system supporting the semi-automatic construction of cooperative information systems from heterogeneous databases, in: *Software – Practice & Experience*, 33 (2003) 9, New York 2003.
- [Pf08] Pfeiffer, D.: *Semantic Business Process Analysis – Building Block-based Construction of Automatically Analyzable Business Process Models*, Münster 2008.
- [PS00] Phalp, K., Shepperd, M.: Quantitative analysis of static models of processes, in: *Journal of Systems and Software*, 52 (2000) 2-3, S. 105-112.
- [PS94] Pollard, C. J., Sag, I. A.: *Head Driven Phrase Structure Grammar*, in: University of Chicago Press, *Studies in Contemporary Linguistics*, London, Chicago 1994.
- [RM05] Rizopolous, N., McBrien, P.: A General Approach to the Generation of Conceptual Model Transformations, in: *Proceedings of the 17th Conference on Advanced Information Systems Engineering (CAiSE'05)*, Porto 2005.
- [Ro96] Rosemann, M.: *Komplexitätsmanagement in Prozeßmodellen. Methodenspezifische Gestaltungsempfehlungen für die Informationsmodellierung*, Wiesbaden 1996.
- [RSD08] Rosemann, M.; Schwegmann, A., Delfmann, P.: Vorbereitung der Prozessmodellierung. In: Becker, J.; Kugeler, M.; Rosemann, M. (Hrsg.): *Prozessmanagement. Ein Leitfaden zur prozessorientierten Organisationsgestaltung*. 6. Auflage, Berlin et al. 2008, S. 45-103.
- [Sa07] Sabetzadeh, M. et al.: A Relationship-Driven Framework for Model Merging, Workshop on Modeling in Software Engineering (MiSE'07) at the 29th International Conference on Software Engineering, Minneapolis 2007.
- [SS83] Schlageter G., Stucky, W.: *Datenbanksysteme – Konzepte und Modelle*, 2. Auflage, Stuttgart 1983.
- [VTM08] Vergidis, K., Tiwari, A., Majeed, B.: Business process analysis and optimization: beyond reengineering, in: *IEEE Transactions on Systems, Man, and Cybernetics*, 38 (2008) 1, 2008, S. 69-82.
- [WJ04] Wulf, V., Jarke, M.: The economics of end-user development, in: *Communications of the ACM*, 47 (2004) 9, S. 41-42.