

Kohyponymie in elektronischen Wörterbüchern

Zusammenfassung: Die Kodierung der Hyponymierelation in elektronischen Wörterbüchern und Wortdatenbanken zeigt zahlreiche Inkonsistenzen, für die hier Lösungen vorgeschlagen werden sollen. Ich beginne mit einer Analyse der Probleme bei der Kodierung der Hyponymierelation für Nomina in existierenden Wörterbüchern, namentlich die fehlende Trennung verschiedener Ebenen der Spezifität und die Vermischung unterschiedlicher Unterordnungskriterien. Beides lässt sich in der Analyse lexikografischer Praxis vor allem an der Heterogenität von kodierten Kohyponymen festmachen. Im Anschluss an die Problemanalyse werden Strategien zur korrekteren lexikografischen Erfassung der Hyponymierelation aufgezeigt. Zentrale Idee dabei ist eine lokale Etikettierung der Hyponymierelation. Ein wichtiges Kriterium für die Erarbeitung der Vorschläge war die einfache lexikografische Umsetzbarkeit.

1 Einführung

Dieser Artikel beschäftigt sich mit Inkonsistenzen in der Kodierung der Hyponymierelation in der lexikografischen Praxis, v.a. bei der Erstellung und Wartung elektronischer Wörterbücher und Wortdatenbanken. Er gliedert sich in zwei Teile: Zunächst erfolgt eine Analyse der Schwierigkeiten bei der Kodierung der Hyponymierelation in bestehenden Wörterbüchern und Wortdatenbanken. Im darauf folgenden zweiten Teil werden Vorschläge zur korrekteren Erfassung dieser Sinnrelation gemacht. Zur Auffrischung des Hintergrundwissens jedoch zunächst eine kurze Abklärung des Begriffs der Hyponymie und verwandter Begriffe.

1.1 Theorie der Sinnrelationen und Hyponymie

Sinnrelationen erlauben die Strukturierung des Lexikons aufgrund von Bedeutungsbeziehungen zwischen Lexemen. Die Bedeutungsbeschreibung durch den Bezug auf die Bedeutung anderer Lexeme ist eine der am leichtesten in die Praxis umsetzbaren Grundideen der lexikalischen Semantik, und fast alle Wörterbücher - elektronische wie herkömmliche - machen in der einen oder anderen Weise Gebrauch von ihr. Die verschiedenen Bedeutungsrelationen haben dabei unterschiedliche Relevanz. Die Hyponymie als die Unterordnungsrelation zwischen Lexembedeutungen ist sicher die neben der Synonymie wichtigste Sinnrelation, um eine lexikografische Gliederung des Wortschatzes vorzunehmen - zumindest im Bereich der Nomina (vgl. Cruse 1986: 88-92).

Es existieren einige sprachlichen Tests für diese Relation. Der übliche Testrahmen für das Vorliegen einer Hyponymiebeziehung zwischen zwei Nomina X und Y sind *ein X ist ein Y* oder *ein X ist eine Art Y* (vgl. Lyons 1977:292f). Aufgrund dieser Testrahmen wird die Hyponymiebeziehung in der englischsprachigen Forschungsliteratur auch IS_A-Relation genannt. Hyponymie kann auch über eine Folgerungsbeziehung zwischen Satzbedeutungen definiert werden (vgl. Lyons 1977: 292-293). Wenn in einer logischen Repräsentation $X(e)$ substituiert werden kann durch $Y(e)$, und das Ergebnis stets wahr bleibt – wie etwa die Aussage *ein Vogel fliegt* stets folgert aus *ein Spatz fliegt* – ohne dass der umgekehrte Schluss zulässig ist, liegt zwischen den Bedeutungen X und Y eine hyponymische Unterordnungsbeziehung vor. Eine solche Substitution ist in negierten Kontexten nicht möglich (s. a. Cruse 1986: 88-92).

Die konverse Relation zu Hyponymie ist die Hyperonymie. Ist das Lexem X Hyponym zu Y , ist folglich Y Hyperonym für X . Ist Z ebenfalls (unmittelbares) Hyponym zu Y , sind X und Z Kohyponyme.

Die Hyponymierelation zwischen zwei Lexemen ist oftmals nur in einem bestimmten semantischen Bereich oder in einem bestimmten Kontext relevant oder gültig. So wird aus der Sicht eines Zoologen – und damit in einem zoologischen Text – die Hyponymiebeziehung *Huhn - Hausgeflügel* eine weniger große Rolle spielen, und dafür die Kategorie *Hühnervögel*, die in der Allgemeinsprache nicht relevant ist, von Interesse sein.

1.2 Relevanz der Hyponymierelation

Hyponymie ist eine zentrale Relation für die Beschreibung diverser sprachlicher Regularitäten:

- Eine Reihe anderer semantischer Relationen wie Synonymie und Opposition lassen sich nur für Kohyponyme beschreiben - d.h. eine Strukturierung des Lexikons aufgrund der Hyponymierelation ist Voraussetzung bzw. geht einher mit der Kodierung anderer semantischer Relationen.
- Selektionsrestriktionen operieren in vielen Fällen auf Mengen von Lexemen, die Kohyponyme sind oder (mittelbare und unmittelbare) Hyponyme eines bestimmten Lexems sind. Es lässt sich zeigen, dass Selektionsrestriktionen und -präferenzen von Verben häufig für hyponymisch definierbare Klassen von Nomina gelten (Langer 1996).
- Bei der anaphorischen Wiederaufnahme von bereits über einen sprachlichen Ausdruck in den Diskurs eingeführten Entitäten werden häufig Hyperonyme zu diesem Ausdruck verwendet.

Daraus folgt, dass der Zugriff auf lexikalische Information zur Hyponymierelation in fast allen computerlinguistischen Anwendungen von Relevanz ist, die auf semantische Information angewiesen sind. Hier einige Beispiele (s.a. Fellbaum 1998):

- In maschinellen Übersetzungssystemen kann/sollte zur Anaphernresolution und zur semantischen und syntaktischen Disambiguierung auf die hyponymische Gliederung des Wortschatzes zurückgegriffen werden.
- In (nicht rein statistischen) Textretrieval-Systemen kann der Zugriff auf Hyponyme und Hyperonyme die Ausbeute erhöhen; eine zu wenig spezifische und nicht korrekte Kodierung der Hyponymie führt allerdings zu Präzisionsproblemen.
- In der automatischen Textklassifikation und im Dokumentenclustering ist der Zugriff auf semantisch verwandte Wortschatzbereiche von hoher Relevanz.

Eine korrekte Kodierung der Hyponymie in den lexikalischen Ressourcen, auf die Systeme der maschinellen Sprachverarbeitung zurückgreifen, kann damit als eine nicht unwesentliche Grundbedingung zur Performanzverbesserung gesehen werden¹.

2 Die Hyponymie in der lexikografischen Praxis am Beispiel WordNet

2.1 Hyponymie in Wörterbüchern

Da die Hyponymierelation mit der Synonymie die wichtigste Sinnrelation zur Strukturierung des Wortschatzes ist, wird sie in fast allen Wörterbüchern mehr oder weniger systematisch erfasst.

Gedruckte Wörterbüchern und Thesauri sind für menschliche Benutzer bestimmt. Es gilt allgemein für solche Wörterbücher, dass bestimmte Typen fehlender Information vom Benutzer ergänzt werden können - daher erfolgt die Beschreibung der Hyponymierelation in der Regel nicht formalisiert und häufig relativ unsystematisch. Unregelmäßigkeiten in der Kodierung fallen in gedruckten Wörterbüchern weniger leicht auf, da in diesen eine Betrachtung der Struktur unter verschiedenem Blickwinkel nicht ohne weiteres möglich ist - so ist es etwa im Dudenband "Sinn- und sachverwandte Wörter" (Müller 1997) nicht möglich auf alle Hyponyme eines Lexems zu-

¹ Die Relevanz semantischer Relationen für zahlreiche Anwendungen zeigt eine Durchsicht der Bibliografie zu Publikationen, die auf die semantische Datenbasis WordNet Bezug nehmen (Mihalcea 2003). Eine ausführliche Darstellung aller Anwendungen würde den Rahmen dieses Artikels bei weitem sprengen.

zugreifen, ohne den gesamten Band durchzuarbeiten. Die meisten Thesauri, d.h. Wörterbücher für die die Unterordnungsrelation zentrales Strukturierungskriterium ist, kodieren nicht die ausschließlich die Hyponymierelation, sondern eine Mischung aus hyponymischer Unterordnung und thematischer Unterordnung.

Elektronische Wörterbücher sind nicht - oder nicht ausschließlich - für menschliche Benutzer gedacht. Anwendungen, die auf solche Wörterbücher zurückgreifen, können fehlende Information nicht einfach ergänzen, sondern arbeiten mit voll spezifizierter, formal kodierter Information. Und anders als in gedruckten Wörterbüchern lässt sich die Information in elektronischen Wörterbüchern ohne weiteres in verschiedenen Ausgabemodi betrachten. Wenn hier semantische Relationen formal erfasst werden, ist es natürlich möglich, von einem bestimmten Lexem ausgehend alle verwandten Wörter ausgeben zu lassen. Deshalb sind auch Inkonsistenzen und Fehler der Kodierung in elektronischen Wörterbüchern und Wortdatenbanken sehr viel einfacher zu erkennen als in gedruckten Wörterbüchern.

Was lässt sich nun über die Kodierung der Hyponymierelation in elektronischen Wörterbüchern sagen? Auf den ersten Blick scheint die Hyponymie eine recht einfach zu erfassende Relation, da die hyponymische Hierarchisierung von Wortbedeutungen weitgehend einer ontologischen Hierarchisierung der Denotate entspricht. Bei näherer Betrachtung der Relation und ihrer lexikografischen Umsetzung wird man allerdings schnell auf die Schwierigkeiten bei ihrer Kodierung aufmerksam. Sie werden deutlich, wenn man stichprobenartig Hyponyme eines Lexems in elektronischen Wörterbüchern und Thesauri vergleicht. Diese Hyponyme, also Lexeme mit einem gemeinsamen Hyperonym werden als Kohyponyme bezeichnet. Der Begriff des Kohyponyms impliziert, dass die Lexeme sich auf derselben Ebene der Spezifität befinden sollten, zudem sollten Kohyponyme komplementär sein, d.h. keine Überschneidungen in der Denotatsmenge aufweisen. Die Realität sieht jedoch völlig anders aus: bei der Auswahl von Hyponymen eines Lexems bzw. einer Menge von Synonymenmenge scheinen zahlreiche verschiedene Unterordnungskriterien zum Zuge zu kommen, und verschiedenste Ebenen der Wortschatzhierarchie vermischt, so dass sehr allgemeine Terme mit hochspezifischen Termen auf einer Ebene auftreten. Die als Kohyponyme kodierten Lexeme sind aus der Sicht des Semantikers und Lexikologen betrachtet also keine wirklichen Kohyponyme. was heißt, dass es in der Praxis der Lexikografie letztlich nötig ist, zwischen echten Kohyponymen und kodierten Kohyponymen zu unterscheiden. All dies lässt den Schluss zu, dass die Hyponymierelation in elektronischen Wörterbüchern nicht korrekt erfasst wird, da eine vollständige und korrekte Kodierung der Hyponymie eigentlich auch dazu führen müsste, dass kodierte und tatsächliche Kohyponyme weitestgehend übereinstimmen.

2.2 WordNet

Zur Illustration soll hier die Kodierung der Hyponymie in WordNet (Fellbaum 1998) herangezogen werden. Die Wahl dieser elektronischen Wortdatenbank zu Illustrationszwecken hat mehrere Gründe: Zum einen ist WordNet sicher die bekannteste elektronische Wortdatenbank, und aufgrund der freien Verfügbarkeit können die hier gemachten Aussagen ohne weiteres nachvollzogen und durch weiteres Material ergänzt werden können. Zum anderen hat WordNet Modellcharakter für andere Projekte der elektronischen Lexikografie - etwa für GermaNet und EuroWordNet (Kunze 2001) oder MultiWordNet (Pianta/Bentivogli/Girardi 2001).

Es soll keineswegs gesagt werden, dass die Hyponymie in WordNet im Vergleich zu anderen Datenbanken oder elektronischen Wörterbüchern besonders unsystematisch kodiert wäre. Im Gegenteil - abgesehen von EuroWordNet ist uns keine Ressource bekannt, die den Wortschatz einer Sprache so umfassend semantisch kodiert, und insbesondere kein zugängliches elektronisches Wörterbuch, das bei vergleichbarem Kodierungsumfang ein ähnlichen oder höheren Qualitätsstandard erzielt.²

Exemplarisch für die Kodierung der Hyponymiebeziehung in WordNet nachfolgend die Hyponyme für "bird" (vollständige Liste, WordNet Version 1.7.1):

bird
 => dickeybird, dickybird
 => cock
 => hen
 => nester
 => night bird
 => bird of passage
 => protoavis
 => archaeopteryx, archeopteryx, Archaeopteryx lithographica
 => Sinornis
 => Ibero-mesornis
 => archaeornis
 => ratite, ratite bird, flightless bird
 => carinate, carinate bird, flying bird
 => passerine, passeriform bird
 => nonpasserine bird
 => bird of prey, raptor, raptorial bird
 => gallinaceous bird, gallinaceon

² Die Kodierung der "classes d'objets" für ein umfassendes französisches Wörterbuch (Gross/Clas 1997) beruht auf einem völlig anderen Ansatz zur Wortsemantik, und ist von daher nicht vergleichbar.

- => parrot
- => cuculiform bird
- => coraciiform bird
- => apodiform bird
- => caprimulgiform bird
- => piciform bird
- => trogon
- => aquatic bird
- => twitterer

Welche Lexeme sind nun hier als Hyponyme von "bird" kodiert? "Dickeybird" ist ein informaler Ausdruck zu Bezeichnung kleiner Vögel; "cock" und "hen" sind eine Unterteilung aufgrund des Kriteriums Geschlecht; andere Kriterien hier sind: Zugverhalten ("bird of passage"), Nachtaktivität ("night bird") und schließlich die naturwissenschaftlich-taxonomische Unterteilung, die allerdings auch nicht systematisch ist: "Nonpasserine bird" (Nicht-Sperlingsvögel) ist wohl unnötig und zudem der Überbegriff für zahlreiche andere hier auftretende Begriffe (etwa "cuculiform bird"). Auch ist die Aufzählung der Ordnungen der Vögel nicht vollständig und es treten zwei nicht weiter qualifizierte ausgestorbene Vogelarten auf ("sinornis" und "archaeopteryx"), die eigentlich auf eine völlig andere Spezifitätsebene gehören. "Parrot" und "trogon" sind neben Vogelfamilien, -gattungen und -arten gleichzeitig Ordnungen der Vögel, und damit auf dieser Hierarchiestufe nicht unbedingt fehl am Platze, ihre Erfassung entspricht aber nicht der Art der Erfassung anderer Vogelordnungen ("psittaciform bird" für "parrot" wäre konsequent).

Inkonsistenzen zeigen sich auch, wenn man die Hierarchie nach oben geht - vom Begriff "gull" (Möwe) sind es fünf Stufen bis "bird", wobei die biologisch-taxonomische Unterteilung sich nur im Begriff "larid" niederschlägt, während die anderen Hierarchiestufen auf einer Unterteilung aufgrund des Lebensraums beruhen. Dagegen wird "Parrot" (Papagei) direkt unter dem Begriff "bird" angesiedelt.

- gull, seagull, sea gull
 - => larid
 - => coastal diving bird
 - => seabird, sea bird, seafowl
 - => aquatic bird
 - => bird

- parrot
 - => bird

Versuchen wir, die Beobachtungen für WordNet zu systematisieren und zusammenzufassen: Der Aufbau der hyponymischen Hierarchie hat folgende problematische Eigenschaften:

- Spezifitätsebenen gehen kunterbunt durcheinander - unter den unmittelbaren Hyponymen eines Begriffs finden sich in vielen Fällen ganz allgemeine Begriffe bis hin zu sehr spezifischen Begriffen, ohne dass eine Ebenenunterscheidung gemacht wird.
- Die Unterteilung ist unvollständig - in unsere Beispiel fehlen in der biologisch-taxonomischen Unterteilung zahlreiche Ordnungen, aber auch andere Begriffe müssten durch Kohyponyme ergänzt werden (so fehlt "resident bird" - Standvogel, das Antonym zu "bird of passage").
- Unterteilungskriterien werden nicht unterschieden - es ist nirgends spezifiziert, dass ganz unterschiedliche Kriterien der Aufstellung der Unterbegriffe zugrunde liegen. In unserem Beispiel ließen sich u.a. die Unterteilungskriterien zoologisch-taxonomische Klasse, Lebensraum, Lebensweise und Geschlecht zeigen, ohne dass diese verschiedenen Kriterien markiert wurden.
- Das Ergebnis der fehlenden Spezifikation unterschiedlicher Unterteilungskriterien ist die Ununterscheidbarkeit von kompatiblen und nicht-kompatiblen Unterbegriffen. So kann ein Zugvogel natürlich gleichzeitig ein Sperlingsvogel sein, ein Sperlingsvogel kann aber nicht ein Raubvogel (zumindest im zoologischen Sinne) sein.

Diese Probleme der hyponymischen Strukturierung des Wortschatzes treten prinzipiell bei der Erfassung der Hyponymierelation in elektronischen Wörterbüchern auf, und sind im Grundsatz sprachenunabhängig.

3 Verbesserungsmöglichkeiten

Im folgenden Abschnitt sollen Vorschläge zur korrekteren lexikografischen Erfassung der Hyponymierelation gemacht werden. Die genannten Inkonsistenzen bei der Kodierung der Hyponymie wurden bereits mehrfach erkannt, etwa in Mendes/Chaves (2001), die das Problem in Bezug auf WordNet im Rahmen der Theorie des generativen Lexikons von Pustejovski (1995) analysieren. Bisher wurden diese Lösungen allerdings meines Wissens nicht lexikografisch umgesetzt, und es scheint mir auch fraglich, ob die Vorschläge ohne weiteres umsetzbar sind. Im Gegensatz zu diesen eher aus der Theorie kommenden Verbesserungsansätzen entstammen die folgenden Vorschläge der lexikografischen Praxis. Die hier vorgestellten Ideen sind bereits in der

semantischen Kodierung für die Nomina des elektronischen Wörterbuchs CISLEX eingeflossen (Langer 1996).

Die Fragen, auf die eingegangen werden soll, sind die folgenden³:

1. Wie erkennt und trennt man verschiedene Unterteilungs- und Unterordnungskriterien, die zu kompatiblen kodierten Kohyponymen (oder eher Schein-Kohyponymen) führen. Dies ist für fast alle Anwendungen elektronischer Wörterbücher relevant, denn nur wenn eine solche Unterscheidung der verschiedenen Unterteilungskriterien gewährleistet ist, lässt sich sinnvoll entscheiden, wo und wie eine Mehrfachklassifizierung von Lexemen unter verschiedenen Zweigen einer hyponymischen Struktur möglich ist - und dies hat u.a. gravierende Konsequenzen für Anwendungen elektronischer Wörterbücher in der Textklassifizierung und der Bedeutungsdisambiguierung
2. Wie grenzt man verschiedene Spezifitätsebenen des Wortschatzes voneinander ab

3.1 Unterscheidung verschiedener Unterordnungskriterien

Zur Lösung des ersten oben genannten Problems schlage ich eine lokale Etikettierung der Hyponymierelation vor. Jede Hyponymangabe wird mit einem Identifikator versehen, der verschiedene Unterordnungskriterien für das Hyperonym trennt.

Als Beispiel hier die Hyponyme für das Lexem "Vogel", jeweils mit dem Identifikator in Klammern vor dem Hyponym:

Vogel
 (zugverhalten) Standvogel
 (zugverhalten) Strichvogel
 (zugverhalten) Zugvogel
 (biotax) Steißhuhn
 (biotax) Laufvogel
 (biotax) Lappentaucher
 [... weitere Ordnungen in der biologischen Taxonomie]
 (lebensweise_nacht_tag) Nachtvogel
 (lebensweise_nacht_tag) Tagvogel

³ Eine Fragestellung die sich v.a. im Zusammenhang mit dem gewählten Beispiel WordNet stellt, auf die hier aber nicht eingegangen werden soll, ist die Zuordnung einzelner Lexeme zu Synonymklassen als Voraussetzung oder begleitenden Maßnahme der hyponymischen Strukturierung des Wortschatzes. Diese Zusammenfassung gleichbedeutender Lexeme sehen wir im Folgenden als gegeben an - obwohl auch diese sich einfach problematisieren lässt.

Die lexikografische Umsetzbarkeit des Vorschlags wird vor allem dadurch gewährleistet, dass lokale Entscheidungen auf der Ebene einzelner Teilbäume der hyponymischen Struktur des Wortschatzes getroffen werden, und auf eine Zuweisung zu Klassen globaler vordefinierter Unterteilungskriterien in allen Zweifelsfällen verzichtet wird. Auf der Basis eines Hyperonyms wird festgestellt, ob verschiedene Unterordnungskriterien herangezogen wurden um die Hyponyme zu erfassen. Dann werden solche Hyponyme, denen gleiche Unterordnungskriterien zugrunde liegen, zu Klassen zusammengefasst, indem die Hyponymierelation mit einem Merkmal etikettiert wird. Im einfachsten Fall ist dieses Merkmal ein beliebiger Identifikator (hier etwa "zugverhalten"), der ausschließlich dazu dient, die verschiedenen Hyponyme eines Lexems zu Klassen von echten Kohyponymen zusammenzufassen. Der Identifikator ist dabei in der Regel nur lokal (d.h. für dieses Hyponym) festgelegt, und besitzt keine weitere Semantik. Zur leichteren Wartung des Wörterbuchs ist es allerdings sinnvoll, wenn möglich einen Bezeichner zu verwenden, der von den Kodierern interpretiert werden kann. Mit dem teilweisen Verzicht auf vordefinierte Identifikatoren soll nicht gesagt werden, dass globale gültige Kriterien für die Erstellung einer Klasse von Kohyponymen nicht relevant wären. Es muss aber festgehalten werden, dass diese Kriterien (etwa die Trennung verschiedener Kriterien durch Qualia-Features wie formal, constitutive, telic wie in Mendes/Chaves 2001 vorgeschlagen) schwer zu benennen und damit verschiedenen Subtypen der Hyponymie nur schwer klassifizierbar sind. Lässt sich die Unterordnungsrelation, in unserem Beispiel etwa im Falle der biologischen Taxonomie, allerdings eindeutig spezifizieren, und hat sie auch klar eine Relevanz für größere Wortschatzbereiche, kann und sollte allerdings eine Ausnahme vom Prinzip der lokalen Etikettierung gemacht, und ein vordefinierter, globaler Identifikator verwendet werden.

Das Ergebnis der lexikografischen Umsetzung ist eine Kodierung der Hyponymierelation, die gewährleistet, dass verschiedene Kriterien zur Unterteilung auseinandergehalten werden, und sich echte Kohyponyme von Schein-Kohyponymen unterscheiden lassen, indem sie durch ein gemeinsames Unterordnungskriterium definierbar sind. So ist im gegebenen Beispiel klar spezifiziert, dass "Zugvogel" und "Standvogel" einander ausschließende Bezeichnungen sind, "Zugvogel" und "Nachtvogel" dagegen kompatible Kategorien bezeichnen.

Durch die globalen Identifikatoren in den eindeutigen Fällen ist eine Isolierung von Teilhierarchien möglich, die aufgrund bestimmter Kriterien erstellt wurden - so lässt sich aus der vorgeschlagenen Kodierung die zoologische und botanische Taxonomie - soweit kodiert - eindeutig durch die Extraktion des Teilbaumes mit den Bezeichnungen "biotax" extrahieren, eine Möglichkeit, die ohne die Etikettierung nicht besteht.

3.2 Abgrenzung von Spezifitätsebenen

Durch die Einführung der Etikettierung verschiedener Unterordnungsrelationen sind die Hyponyme eines Lexems bereits zu Kohyponymklassen zusammengefasst. Innerhalb dieser Klassen ist es nun wesentlich einfacher, zu gewährleisten, dass sich alle - nunmehr potentiell echten - Kohyponyme auf derselben Spezifitätsebene befinden. Die folgenden Schwierigkeiten bei der korrekten Festlegung der Spezifitätsebenen können damit entschärft werden:

- Unschärfe Bezeichner und singuläre Hyponyme: In allen Wortschatzbereichen gibt es mehr und weniger wohldefinierte Begriffe. Für den hier als Beispiel herangezogenen Bereich der Vögel sind etwa die Begriffe "Wasservogel" oder "Küstenvogel", die auf Lebensräume abheben, alles andere als klar abgrenzbar. Ein ähnlich problematischer Begriff ist "Raubvogel". Alle diese Lexeme haben keinen Platz in der zoologischen Taxonomie. Es ist auch unklar, auf welcher Hierarchiestufe sie angesiedelt werden sollten. Diese problematischen Fälle lassen sich nicht prinzipiell lösen, doch durch die Etikettierung der Unterordnungsrelation führen problematische Einordnungen hier nicht zu größeren Inkonsistenzen im Hierarchieaufbau.
- Lexikalische Lücken: Es existiert kein Hyponym auf der richtigen Spezifitätsebene, d.h. für eine Konzept, das zur Vervollständigung einer Gruppe von Kohyponymen notwendig wäre, existiert kein Lexem⁴. Dies gilt etwa für die im vorigen Punkt genannten singulären Hyponyme zu Vogel. Eine Einführung von künstlichen Hierarchieknoten, die mit Beschreibungen gefüllt werden, ist jedoch nicht nötig, da durch die Etikettierung der Unterordnungsrelation der Status der singulären Hyponyme deutlich wird.
- Polysemie: Das Hyponym auf der richtigen Spezifitätsebene ist homonym zu einem Lexem auf einer niedrigeren Spezifitätsebene. So sind die Kuckucke eine Ordnung der Vögel, gleichzeitig ist der Kuckuck eine spezifische Vogelart. Hier muss darauf geachtet werden, dass die Unterscheidung der Spezifitätsebenen nicht verloren geht; die zwei Bedeutungen des Lexems müssen beiden relevanten Spezifitätsebenen zugewiesen werden. Im Falle der biologischen Taxonomie kann man durch geeignete Kodierung der lateinischen-wissenschaftlichen Bezeichner als Synonyme gewährleisten, dass die Ebenentrennung auch in der weiteren Kodierung erkennbar bleibt.

Es wird deutlich, dass die im vorherigen Abschnitt vorgestellte Etikettierung die aufgezählten Probleme zwar nicht vollständig löst, aber die Verwaltung der verschiede-

⁴ Zur Definition des Begriffs und zur Behandlung lexikalischer Lücken in EuroWordNet s. Bentivogli/Pianta (2000).

nen Ebenen der hyponymischen Hierarchie durch die erreichte Übersichtlichkeit wesentlich leichter macht.

4 Zusammenfassung und Ausblick

Die Hyponymierelation, die auf den ersten Blick eine einfach zu kodierende Relation zu sein scheint, erweist sich bei ihrer lexikografischen Erfassung für formal wohldefinierte elektronische Wörterbücher als hochproblematisch. Dies schlägt sich insbesondere in der Heterogenität von kodierten (aber nicht wirklichen) Kohyponymen nieder.

Es wurden Vorschläge gemacht, die eine klarere Kodierung der Hyponymie gewährleisten sollen, und die vor allem eine Erfassung von echten Kohyponymen ermöglichen sollen. Als Kohyponyme sollten nur sich einander ausschließende Hyponyme eines Lexems kodiert werden, die sich auf derselben Spezifitätsebene befinden. Zentrale Idee zur Erreichung dieses Ziels ist die Etikettierung der Hyponymierelation mit einem in den meisten Fällen nur lokal gültigen Identifikator.

Eine Frage, auf die hier nicht näher eingegangen wurde, war die Frage danach, wie es möglich sein könnte, intuitive (allgemeinsprachlich relevante) von anderen Hyponymen zu trennen - d.h. wie trennt man Hyponymrelationen, die nur für spezifische Wortschatzbereiche (etwa Fachsprache) relevant sind von anderen. Die Relevanz einer solchen Trennung kann u.a. für Anwendungen semantischer Kodierungen in Information-Retrieval-Systemen gezeigt werden. In Langer/Hickey (1998) wurde vorgeschlagen, die allgemeinsprachlich relevanten Ebenen einer hyponymischen Hierarchie wie WordNet durch Frequenzanalysen zu ermitteln - in die Kodierung des CISLEX ist diese Idee allerdings bisher nicht eingeflossen.

Literatur

- Bentivogli, L. und E. Pianta (2000): Looking for lexical gaps. In *Proceedings of the ninth EU-RALEX International Congress*, Stuttgart, Germany, August 8-12, 2000.
- Fellbaum, C. (1998): *WordNet: An Electronic Lexical Database and Some of its Applications*. Cambridge (USA): MIT Press.
- Gross Gaston, Clas André (1997): "Synonymie, polysémie et classes d'objets". *Meta* 42 (1): 147-155.
- Kunze, C. (2001): Lexikalisch-semantische Wortnetze. In: K.-U. Carstensen et al. (Eds.): *Computerlinguistik und Sprachtechnologie: eine Einführung*. Heidelberg; Berlin: Spektrum, Akademischer Verlag, 386-393.
- Langer, S. (1996). Selektionsklassen und Hyponymie im Lexikon. Semantische Klassifizierung von Nomina für das elektronische Wörterbuch CISLEX. Dissertation Universität München. CIS-Bericht.
- Langer, S. and M. Hickey (1998). Using semantic lexicons for intelligent message retrieval in a communication aid. *Journal of Natural Language Engineering* 4 (1). Special Issue on NLP for communication aids, 41-56.
- Lyons, John (1977): *Semantics*. Cambridge: Cambridge University Press.
- Mendes, S. und R.P. Chaves (2001): Enriching WordNet with Qualia Information. In: *Proceedings of the Workshop on WordNets and other Lexical resources at the II NAACL Meeting 2001*, Carnegie Mellon University, Pittsburgh, USA.
- Mihalcea, Rada (2003): WordNet Bibliographie. <http://enr.smu.edu/~rada/wnb/> [Zuletzt besucht 24.01.2003].
- Müller, W (Hrsg.) (1997): *Duden "Sinn- und sachverwandte Wörter": Synonymwörterbuch der deutschen Sprache. - Nach den Regeln der neuen dt. Rechtschreibung überarb. Neudr. der 2. Aufl.* Mannheim [u.a.]: Dudenverlag.
- Pustejovsky J. (1995): *The Generative Lexicon*. Cambridge (USA): MIT Press.
- Vossen, P. (Hrsg.) (1998): *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer.
- Pianta, E., L. Bentivogli, und C. Girardi (2001): MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, Mysore, India, January 21-25, 2002.