

言語知識の電子化辞書における情報構造の研究

横井俊夫

1996年

言語知識の電子化辞書における  
情報構造の研究

横井 俊夫

1996 年

## 序文

大規模な情報や知識の構造を解明し、共有性に富む情報ベースや知識ベースを構築するための技術を確立することは、国家情報基盤、国際情報基盤の整備の礎となるきわめて重要な課題である。本論文では、言語に関する知識、すなわち、言語知識を対象に情報構造の解明を試み、電子化辞書として具体的に実現した研究について述べる。

情報そのものに強い関心が寄せられるようになった背景には、高機能化を求めて情報処理技術が意味処理に向かい始めたこと、大量情報の集積と利用という急成長を始めた領域から有効に対応するための新しい技術への強いニーズが生じ始めたことがある。これら背景のもとに具体的な課題としての共通認識が生まれ始めた分野に自然言語処理と知識処理がある。自然言語処理技術の本格的な研究開発のためには、多様で複雑な言語現象を体系的にとらえた大規模な言語データが重要であることが共通認識となってきた。知識処理においては、実用化システムへの過剰な期待への反省として共有性・再利用性に富む知識ベースコンテンツの着実な研究・開発が重要であることが共通認識となってきた。

これらの共通認識のもとにこの十年来各所で多彩な取り組みがなされてきた。しかし、それらの諸活動には、本格的なものに向けて多くの課題、問題が残されている。自然言語処理のための大規模言語データに関する活動については、自然言語処理技術という工学からの要請を十分にバランス良く反映したものになっていないこと、大規模な実現へのプロセスを十分に考慮したものとはなっていないこと、統合的な観点からの設計がなされていないことなどである。知識処理における大規模知識コンテンツ構築技術に関する活動については、本研究に直接関連するものとして、知識表現メディアにおける議論



が未整理である点が挙げられる。すなわち、知識処理から見ても、自然言語を知識表現メディアとして重視しなければならないが、自然言語による文書ベースと知識表現言語による知識ベースとのつながりが明らかにされていない点である。

このような状況に対して本研究は、言語知識の統合的な情報構造を明らかにすることに焦点を当てる。そして、この情報構造が、今後の自然言語処理技術の動向や効率的な開発プロセスの実現に十分に対応できること、また、文書ベースと知識ベースを統合化する新たな提案へと結び付くものであることを明らかにする。ただし、このような試みとしては初めてのものであることから、基本仕様、概念仕様の提案にとどまるものである。

情報構造は仕様の記述と EDR 電子化辞書という実現事例の提示によって説明される。基本とする情報構造は、語、句、文、文章、文書という記述の単位、表層から概念、さらにより深層という記述レベル、そして、日本語、英語等の言語の種類によって特徴付けられる。表層と概念のレベルに対し、より詳細な情報構造が述べられる。説明は基本となる情報単位である辞書項目の内容を詳細化することによって進められる。辞書項目は、語、文、文章、文書という構成単位、それぞれにつき説明がなされる。また、基本構造、表層レベル、概念レベルのそれぞれに実現事例として EDR 電子化辞書が対応付けられる。述べられる情報構造は、EDR 電子化辞書の成果を再整理することによって得られたものであり、情報構造の実現としても EDR 電子化辞書は第一步を印したというところのものである。また、本研究の言語知識の情報構造の提案は、このような仕様のものを総体として実現すべきであるという考えによるものではない。実際の言語知識コンテンツの開発・整備は各所でそれぞれの目的に沿って進められることになる。そのような多くの努力が納まるべき枠組みを明らかにし、効果的な協力や共有化を可能にしようという考えによるものである。そのためにも、提案の妥当性を、自然言語処理、世界知識の情報構造、開発工程、それぞれの観点から検討し、他の事例との比較・検討、対応付けも行う。

情報処理技術は、今まで大規模なハードウェアや大規模なソフトウェアの構造を明ら



かにし、それに基づいて実用性の高いシステムを実現してきた。次なる課題は、大規模な情報や知識そのものの構造を明らかにすることである。しかしながら、情報や知識の構造に関する研究はまだ取り組まれ始めたごく初期の段階にある。すなわち、ハードウェアがアーキテクチャという観点から取り組まれ始めた頃、コンパイラやOS等のソフトの試作・研究が始まった頃に対応する。したがって、本研究も、このような新しい分野に輪郭を与えるための試みであり、出発点のひとつとなるものである。

## 目 次

|                       |    |
|-----------------------|----|
| 1. 序論                 | 1  |
| 1.1 研究の背景および目的        | 1  |
| 1.2 研究の位置付け           | 5  |
| 1.3 論文の構成             | 17 |
| 2. 言語知識の基本情報構造        | 19 |
| 2.1 情報構造への要求仕様        | 19 |
| 2.2 要求仕様の詳細化に関する技術的判断 | 21 |
| 2.3 情報構造の基本仕様         | 24 |
| 2.4 EDR 電子化辞書との対応     | 28 |
| 3. 表層レベルの情報構造         | 32 |
| 3.1 表層辞書項目            | 32 |
| 3.2 語表層辞書             | 35 |
| 3.3 文表層辞書             | 38 |
| 3.4 文章表層辞書・文書表層辞書     | 40 |
| 3.5 EDR 電子化辞書における実現   | 41 |
| 3.5.1 辞書仕様            | 41 |
| 3.5.2 数値データ           | 45 |

|                       |     |
|-----------------------|-----|
| 4. 深層レベルの情報構造         | 51  |
| 4.1 概念と概念辞書項目         | 54  |
| 4.2 語概念辞書             | 56  |
| 4.3 文概念辞書             | 60  |
| 4.4 文章概念辞書・文書概念辞書     | 63  |
| 4.5 EDR 電子化辞書における実現   | 64  |
| 4.5.1 辞書仕様            | 64  |
| 4.5.2 数値データ           | 70  |
| 5. 情報構造の妥当性           | 76  |
| 5.1 自然言語処理からの検討       | 76  |
| 5.2 世界知識の情報構造からの検討    | 79  |
| 5.3 開発工程からの検討         | 80  |
| 6. 他の事例との比較           | 81  |
| 7. 結論                 | 87  |
| 付録. EDR 電子化辞書における実現仕様 | 90  |
| 付.1 仕様の記述形式           | 90  |
| 付.2 全体の構成             | 92  |
| 付.3 日本語単語辞書           | 94  |
| 付.4 英語単語辞書            | 98  |
| 付.5 概念辞書              | 104 |
| 付.6 日英対訳辞書            | 114 |
| 付.7 英日対訳辞書            | 117 |



|                   |     |
|-------------------|-----|
| 付.8 日本語共起辞書.....  | 120 |
| 付.9 英語共起辞書.....   | 124 |
| 付.10 日本語コーパス..... | 125 |
| 付.11 英語コーパス.....  | 129 |
| 参考文献              | 133 |
| 本論文に関する原著論文       | 137 |
| 図一覧               | 143 |
| 表一覧               | 144 |

## 第 1 章

### 序論

#### 1.1 研究の背景および目的

大規模で高度に構造化された情報そのものはたず役割が情報処理技術の高機能化にとってきわめて重要であることが強く認識されるようになった[Fuchi and Yokoi 94]。ハードウェア、ソフトウェアという言葉に対し、インホウェア (infoware) という言葉も使われるようになってきている。このように情報そのものの構造を解明し、大規模な情報・知識ベースの構築や利用の技術を研究開発することへの要請が存在することが、最も広い視点からとらえた時の本研究の背景である。

情報そのものに強い関心が寄せられるようになった背景には、多くの要因があるが、代表として次の2つがあげられる。

- 1) 意味処理への関心：情報処理の高機能化を求め情報の意味の処理への関心が高まってきた。情報の意味を扱うためには情報が表現している多様で複雑な対象世界を捕捉しなくてはならない。これは単純なモデルや簡単なアルゴリズムで捕捉できるものではなく、対象世界の構造を、それを表現した大規模な情報の構造として分析していく手法が必要になる。
- 2) 大量情報の集積と利用：情報化の進展は大量の情報の蓄積をもたらす。マルチメディア技術やネットワーク技術の進歩と普及によって、集積される情報の範囲と量は急速に拡大されつつある。コンピュータが対象にしている情報表現メディアが広がり、物理的な距離にわずらわされることなく情報にアクセスでき

るようになったからである。これにともなう、この大量情報をいかに有効利用するか、あるいは有用な利用が可能なように情報の蓄積を行うにはどのようにすべきか、また、様々な情報障害にいかに対処するか、等々の課題が急ぎ取り組むべきものとなってきている。このような課題に取り組むには情報が持つ内部構造を解明し、膨大な情報を体系付け、秩序立てる枠組みを作り出さなくてはならない。

次に自然言語処理と知識処理の観点から絞って本研究の背景を述べる。

#### 自然言語処理と言語データ

自然言語処理技術の本格的な研究開発のためには多様で複雑な言語現象を体系的にとらえた言語データが重要な役割をはたす。言語データは理論や方式を検証するための客観的なデータというより、自然言語処理の機構に直接結びつくものである。10 年程前からの機械翻訳システムの開発を契機に始まった自然言語処理研究の本格化によって、このような言語データの重要性が強く認められるようになり、様々な試みが世界各国で行われるようになった。日本においては、著者が推進した日本電子化辞書研究所の電子化辞書プロジェクト（1986 年度～1994 年度）が包括的な言語データを目指した最初の試みである。その他に用途等を限定した言語データの研究・開発としては、NTT 情報通信研究所における機械翻訳システム ALT 用辞書、日本科学技術情報センター（JICST）における機械翻訳用対訳辞書、情報処理振興事業協会（IPA）の日本語辞書（IPAL）が日本における代表的な取り組みである。

ヨーロッパにおいては、言語事情の複雑さを反映して多くの試みが行われている。辞書に関するものとしては Aquilex、Genelex、Multilex 等が代表例である。コーパスやテキストデータに対しても多くの試みがある。British National Corpus はコーパスの分野で世界を代表する試みである。アメリカにおいてもテキストの検索や要約を中心に



自然言語処理への関心の高まりがあり、言語データ研究開発の多くの試みがなされている。辞書に関しては Comlex, シソーラスを含め単語の意味の定義を試みるものに関しては WordNet, コーパスに関しては Penn Treebank が代表例である。

以上のような試みによって現状の言語データの開発が進む一方で、その経験を整理して今後の言語データ研究開発を方向付けるための模索が始まっている。言語データは自然言語処理技術の進展を先導するものでなければならない。求められている言語データは、辞書、シソーラス、コーパス、テキストデータ等を体系的に統合化したものでなければならない。そして、本来、十分に大規模であり、高い精度を有し、低いコストで短期間に開発できるものでなければならない。このような言語データの実現のためには新たに多くの課題に取り組まねばならない。その中で、まず、急がねばならないのは、どのような言語データでなければならないのか、すなわち、統合的な言語データの情報構造の解明とそれに基づく大規模な実現方法の研究・開発である。

#### 知識処理と大規模知識ベース

知識処理は人工知能研究の中で現実的な成果が期待できる分野としてここ 10 年程多彩な取り組みがなされてきた。知識処理の一般的な機構の研究と処理系の実現、もう一方では各種の応用システムの開発、実用化が進められた。しかしながら、知識獲得の隘路から知識ベースの機構が当初期待された程には機能しないことが明らかとなった。その解決のためには、十分な有用性を持ち、共有性・再利用性の高い知識ベースコンテンツの研究開発が重要な課題となってきた[Yokoi 94]。

このような知識ベースコンテンツに向けての試みは役割付けやアプローチから 3 種類ほどに分類される。第一のものは、純粋な人工知能研究の立場から人間の知能の核となる常識の解明を目指すものである。これに関しては常識を表現する言語や機構を何にするかによって色々なアプローチがあり、多様な試みが必要であるが、自然言語からのアプローチを有力視する意見がある。第二のものは、より実用的な汎用知識ベースの実

現を目指し、各分野に対するオントロジー<sup>1</sup>を実現しようというものである。オントロジーは対象世界をどう見るかの世界観、その世界観に基づき記述するための用語のセット、そして、その用語によって記述された知識内容から構成されるものとみなされる。この試みも、中心的な課題である知識獲得の機構を自然言語処理の機構の上に構成しようという提案が注目され始めている。第三のものは、メディア処理と知識処理を融合し、非常に汎用性の高い大規模知識ベースを実用性の高いレベルから基礎的な研究への試みまでを含めて研究開発しようとするものである。この試みでは自然言語メディアを最も汎用な知識表現メディアと見なし、メディア処理と知識処理を結び付けようとするものである。

いずれにしても、自然言語による知識表現を有望な手法と考え、自然言語処理を知識処理の大きな構成要素と見なす考え方は共通のものである。そして、この考え方をさらに進めると、自然言語処理のための言語データを知識処理が目指す大規模知識ベースコンテンツの一つの実現例と見なすという考え方に至る。すなわち、統合的な言語データの情報構造を一般的な知識の情報構造（知識構造）の基底部分に対応付けようという考え方である[Yokoi 95]。

以上のような背景分析から、統合的な大規模言語データの情報構造の研究を言語知識、すなわち言語に関する知識の情報構造の研究と称することにする。

本研究の目的は大規模な言語知識が持つべき情報構造を明らかにすることである。言語知識とは、言語に関する知識のことであり、これからの自然言語処理の研究や自然言語処理応用システムの開発に重要な役割をはたすように整えられた言語に関するデータや知識の体系的な集積である。言語知識は電子化辞書と呼ばれるものに具体化される。ここでいう電子化辞書とは、通常いわれる辞書ばかりではなく、シソーラスやコーパスやテキストデータまでを対象にし、形態や構文にかかわる知識から意味や文脈にかかわる知識までを含むものである。この事実を強調するために汎電子化辞書という言葉も用

---

<sup>1</sup> 哲学の用語である ontology（存在論）を借用したものである。本来の意味に用いている訳ではない。

いられる。

知識の情報構造とは、知識に内在する論理的な構造のことである。すなわち、どのような要素知識アイテムから成り立っているのか、知識アイテムがどのように関係し合っているのか、知識アイテムがどのようなまとまりを形成しているのかなどという知識の構造のことである。言語知識の情報構造を、電子化辞書の論理仕様であるということもできる。論理仕様は具体的な検索方法や処理方法から独立したものである。この論理仕様を特定のデータベース機構や知識ベース機構に対応付けることにより電子化辞書の具体的な仕様が得られることになる。言語知識の情報構造は、マクロなレベルからミクロな細部を詳細に規定するレベルまでさまざまな段階がある。本研究が対象とするのはマクロなレベルのものである。ソフトウェアシステムやハードウェアシステムにおける基本仕様、概念仕様に相当するものである。

情報処理技術は、今まで大規模なハードウェアや大規模なソフトウェアの構造を明らかにし、それに基づいて実用性の高いシステムを実現してきた。これからの課題として大規模な情報や知識そのものの構造を明らかにすることがある。しかしながら、情報や知識の構造に関する研究はまだ取り組まれ始めたごく初期の段階にある。すなわち、ハードウェアがアーキテクチャという観点から取り組まれ始めた頃、コンパイラやOS等のソフトの試作・研究が始まった頃に対応する。従って、本研究では、すべての出発点という意味から、基本仕様、概念仕様というレベルに焦点を合わせ、詳細な仕様に関しては今後の研究に委ねることにする。

## 1.2 研究の位置付け

背景および目的の説明に加えて本研究の位置付けをまとめる。すなわち、本研究が対象とする分野がかかえる課題が何であり、状況はどのようなであったのか、それに対し本研究がいかなる新しい解決策を試みたのかを説明する。



## 分野がかかえる課題

本研究が対象とする分野、すなわち、自然言語処理と知識処理の分野それぞれがかかえる課題について前節の要点を整理し補足すると次のようになる。まず、自然言語処理分野の要点を整理し補足すると以下のようになる。自然言語処理分野における最大の課題は大規模な言語データに基づく本格的な自然言語処理技術、文書処理技術の確立である。文書の検索、翻訳、要約に関する高度な技術の実現には言語表現（言語で表現されたもの）に対する同一性の判定や同一内容のものへの変換を行う高度な処理機能が必要である。そして、この処理機能は、処理のレベルをより深い意味のレベルで行えるようにすること、また、処理の単位を単語から文、文章へと拡大していくことによってより高機能なものになる。しかしながら、言語現象は非常に複雑、多様である。したがって、処理のレベルを深め、処理の単位を大きくすることを単純な規則や少量のデータによって実現することは不可能である。処理機能に対応した言語現象を包括的にとらえる大規模な言語データが不可欠となる。いうなれば、これは大規模言語データに基づく意味処理、文脈処理の機能実現へのアプローチである。

これに対し、もうひとつの典型的なアプローチとして人工知能分野における言語理解と呼ばれるものがある。言語理解においては対象を限定し、世界や話者のモデルを作り上げ、その上に意味処理や文脈処理の仕組みを精緻に組み上げるというものである。しかしながら、多くの人工知能研究の例に漏れず、言語理解によるアプローチもスケールアップが難しくトイ・プロブレムに留まったままになっている。文書処理においては実問題に対応できる頑健な意味・文脈処理が求められる。そのため、世界や話者のモデルとしては一般性はあるが粗い近似のものをを用いて、浅いが広範囲を覆うことのできる機能に注目することになる。そのような機能を実現する土台となるのが単語から文、文章、文書にまたがる体系的に整備された言語データである。

知識処理分野における最大の課題は共有し広く利用できる大規模な知識コンテンツの構築技術の確立である。ただし、ここでの知識処理は通常より広範囲の内容を含む言

業として用いられる。すなわち、体系的に蓄積された大量の情報を適切に利用し、新しい情報を作成し、その情報から新しい価値を生みだし、さらに、それらを再利用のために蓄積するという一連の情報処理活動にかかわるものをすべて含むものとして用いる。知識ベース、ハイパーメディア、テキストベース、データベース等に関連する技術を含むものである。ただし、以下の説明では知識ベースとハイパーメディアに関連する技術を中心にする。

現在、大量の情報が蓄積され広く共同利用できる環境が現実により作り出されているのは、インターネット上のハイパーメディア・システムである。WWW(World Wide Web)上のGopherやMosaicによって地球規模での環境が作り出され急速に成長しつつある。ここでの情報・知識は自然言語、画像、音声など人間が従来から用いてきたメディア（ヒューマンメディア）によって表現されている。これによって従来からの情報の蓄積が直接的に利用でき、新しい情報を容易に作成し、共有化できることになる。そして、高速のネットワークによって物理的な距離が無視できるようになり、各所に分散するものが一体化され大規模な情報・知識スペースが出現することになる。ただし、このスペースは単純なハイパーメディア構造という初歩的な情報構造を持つだけである。このままでスペースの成長にともなって有効利用が急速に難しくなることは明らかであり、抜本的な対策への模索が始まっている。もうひとつの観点は、知識ベースシステム（エキスパートシステム）における知識獲得のボトルネックをどう克服するかという課題に根差すものである。エキスパートシステムの期待過剰であった部分には反省を要するが、正當に評価すべき部分は適切に発展させねばならない。ここでの知識・情報は、知識表現言語、エキスパートシステムシェルなど、コンピュータ向けの人工言語（コンピュータメディア）によって表現されている。専門家が有する知識を引き出しこの人工言語で適切に表現し尽くすという作業がボトルネックとなる。しかしながら、知識ベースシステムの有用性は、専門家が持つ高度な（高次の）知識によって達成されるというより、単純な知識の大規模な集積によって達成されると考えるべきである。このような考え方に基づいて、知識表現におけるヒューマンメディアとコンピュータメディアの連携、そし

て、それを利用しての大規模な知識ベースの実現が重要な課題となる。

#### 解決に向けての活動の状況

各分野の課題に対し、解決や改善に向けて行われている諸活動の状況を概観する。大規模な言語データの研究・開発に関しては、従来からの辞書学・言語学のものに加え、自然言語処理への利用を指向した活動がここ十年来活発化してきている。欧州、米国、日本が活動の中心であるが、韓国や中国などアジア地域においても活動の萌芽がみられるようになってきている。

欧州では、EC (EU) の機械翻訳プロジェクト EUROTRA 計画が順調とは言い切れなかったことへの反省として言語データの整備に関心が向けられるようになった。大学等の研究機関が中心となり ESPRIT 計画の中で進められたものに Acquilex、Multilex の2つのプロジェクトがある。Acquilex[Briscoe et al. 90]は、第一期(1989年10月-1992年7月)はイタリアのビサ大学、第二期(1992年6月-1995年5月)は英国のケンブリッジ大学が中心となった。機械可読辞書から自然言語処理システムに利用可能な語彙(単語)に関する知識を半自動的に抽出するための技術の確立と評価を目的としたものである。辞書の基本仕様は HPSG に基づくユニフィケーション文法の枠組みに従っている。辞書の作成技術や仕様の研究が中心であるため、実際に大規模な辞書を目指すには、このままでは難しいとの批判がある。半自動的にとはいえ、既存の辞書から知識を抽出することによって辞書を作成していくという設定には現状のままでは無理があるようである。基本部分は人手によってまず精度の高いものを作り上げるという手順が基本的に必要であると思われる。また、あまりに仕様が緻密すぎるため不必要に手順が多くなりすぎて、実際の言語現象に適応すると相当の無理が生ずることとなるとされる。

Multilex(1990年12月-1993年12月)[Serasset 94]は、欧州各国の機関が協力し、各国で利用できる多言語多機能辞書の仕様を策定し、辞書データを作成し、検索な



どのソフトウェアを開発し、応用システムによって有用性を検証しようというものである。しかし、目標の割には投入された資源が少なく、そのため、辞書の基本仕様の策定とごく基本的なソフトウェアの開発にとどまっている。欧州の各国語に適応できるオーソドックスな単語の辞書仕様の提案となっている。

企業が中心となり EUREKA 計画の中で進められたものに Genelex プロジェクト (1990 年 12 月 - 1994 年 12 月) [Normier and Nossin 91] がある。共通の辞書仕様のもとに、フランス語、イタリア語、ポルトガル語等のローマン系言語の単語に関する汎用辞書を開発・整備しようというものである。各機関が保有する辞書を包括した辞書データベースを開発するという現実的な要請に沿うものである。自然言語処理用というより機械可読辞書のデータベース化という段階のものである。かなりの資源が投入され、実際の実現も見ているようであるが、企業が主導していることから成果の詳細な公表が遅れている。

各国それぞれが進めるプロジェクトもある。その代表例が英国の BNC (British National Corpus) プロジェクト (1991 年 1 月 - 1994 年 3 月) [BNC 92] である。英語に関する言語現象 (書き言葉、話し言葉) を網羅するように設計された大規模コーパス (文例とその出典) の開発である。すでに目標としたものが開発され、広く利用に供されている。内容としては、素材としてのデータの集積であり、解析データに関しては品詞のタグ付けと一部構文解析データの付与が行われている。どのような解析データをどのように付与していくのか、どのように加工し利用するかは、今後の課題ないしは、利用者への課題である。

欧州は欧州統合という課題からも言語に関する問題に非常に敏感であり、多くの努力を払ってきている。言語学、辞書学、用語学等の長い伝統もあり、辞書やコーパス等に関しても人文科学的言語データとしては、豊かな蓄積を有している。それだけに、本研究が対象とするような工学的言語データに対しても言語学者や辞書学者サイドの発言力が強く、コンピュータ科学的、自然言語処理的発想が弱まる傾向にある。実際の実現より仕様の詳細な議論に深入りする傾向や工学的なバランス感覚にそぐわない傾向な

どがみうけられる。

米国においては、欧州ほど多彩ではないが大規模な言語データの研究・開発・供給に関して ARPA の支援のもとに一貫した体制が作られている。LDC (Linguistic Data Consortium) がペンシルバニア大学に CLR (Consortium for Lexical Research) がニューメキシコ州立大学にそれぞれ設立されている。歴史的に見ると、米国においては話し言葉に対するものの方に比重が大きい。欧州や日本に比べると、機械翻訳研究への要求の希薄さから、書き言葉に対する関心が相対的に低い。しかしながら、文書検索や文書からの知識抽出や文書要約への関心が高まるにしたがって、書き言葉に対する言語データの開発・整備への活動も活発化してきている。ARPA による TREC、MUC、TIPSTER 等の文書処理関連プロジェクトと連携している。代表的な活動は、Comlex、WordNet、Penn Treebank である。

Comlex[Grishman et al. 94]はニューヨーク大学計算機学科において開発された英語の単語に関する構文情報にかかわる記述にしばった辞書である。既存辞書を利用して短期間に基本部分を実現した(1994年5月第0版)。LDCの会員にとりあえず利用できる辞書を提供するのが目的で、改良、拡張や WordNet との統合等は今後の課題となっている。

WordNet[Miller et al. 93]はプリンストン大学の G.A. Miller 等によって 1985 年より進められている心理言語学の立場から作成された一種のシソーラスである。語の連想の仕組を明らかにするため人間の心理的な語彙の記憶のシステムを実現するのが目的である。ただし、内容的には、本研究の単語に関する概念辞書の考え方に非常に近いものとなっている。自然言語処理への利用も検討されており、そのために最小限度必要な表層語情報との接続作業が Comlex との間で計画されている。

Penn Treebank[Marcus et al. 93]は、ペンシルバニア大学で 1989 年より開発が続けられているコーパスである。ACL (American Computational Linguistics) の DCI (Data Collection Initiative) の収集資料を対象にしたもので、約 600 万トークンの規模のものが実現されている。全体に品詞のタグ付けがなされており、2/3 に対しては構文構造

の解析データが付与されている。基本的な仕様は BNC に非常に近いものであるが、言語現象を系統的にとらえるという立場からのコーパスにはなっていない。

日本においては、日本語の特殊性から日本語ワープロ、機械翻訳等の自然言語処理システムへの関心が高く、言語データの整備もこれにそって進められてきた。そのためもあり、全体としては企業における個別の努力が主になってきたきらいがある。しかしながら、より広い立場からの活動もこの十年程にいくつか始まってきている。日本における活動の代表は、本研究の内容となっている EDR 電子化辞書プロジェクトであるが、それ以外にも色々な機関で努力が続けられている。その内で内容が公表され、利用実績のあるものの内で代表的なものは ALT 辞書、IPAL、JICST 対訳辞書、分類語彙表である。

ALT 辞書(日英機械翻訳システム ALT への利用を主目的としているためとりあえずこう呼ぶ)[池原ほか 93] は、NTT 情報通信研究所で開発が続けられているもので、大規模な日本語単語からの辞書を実現している。自然言語処理、特に、機械翻訳において表層に近いレベルの意味解析を実用規模で実現することを目的として開発され、その目的にかなう完成度に達している。対訳情報は日英に限られるが、日本語の単語に関してはかなり完備した辞書であり、慣用文型を含む豊富な文型パターン情報によって文に関する辞書への手掛かりを与えている。

IPAL[情報処理振興事業協会 87] は、情報処理振興事業協会技術センターで開発が続けられているもので、日本語の単語について基本的な部分が実現されている。大規模な辞書を実現することより語数を制限し、基本的な語に対し緻密な分析と記述を行うことを当面の目標としている。ただし、その緻密さが言語的興味に引きずられがちであり、自然言語処理からの要求にはかならずしも合致しないものになってきている。

JICST 対訳辞書[芦崎 94] は、日本科学技術情報センターにおいて開発が進められているもので、同センターが開発した日英機械翻訳システムに使用されている。日本語の単語に対する構文情報、英語への対訳情報を中心に科学技術、医学の専門用語を含む実用規模の辞書を実現している。



分類語彙表[国立国語研究所 64]は、国立国語研究所で作成された日本語語彙に対するシソーラスである。広く公開されており、適時改良が行われていることから、このような言語データとしては、かなり標準的なものとして利用されている。自然言語処理の意味データとしても利用されているが、基本的な発想は言語学からのものである。

次に、知識処理分野における、共有化される大規模な情報・知識コンテンツの構築技術に関する活動状況を概観する。このような技術の核となるは、情報や知識を表現するメディアである。そこでメディアの観点から活動状況を分析する。コンピュータメディアそのものにかかわる活動、コンピュータメディアからヒューマンメディアを指向する活動、ヒューマンメディアからコンピュータメディアを指向する活動、ヒューマンメディアにかかわる活動の4つに分類する。全体を概観するが、重点を置くのは二番目と三番目である。

コンピュータメディアそのものにかかわる活動としては、洗練された計算機構の上に高度な言語を開発すること、標準とする知識表現言語を設定すること、異なった言語間での変換を可能にすること、知識ベースシステム間のプロトコルを標準化することなどがある。高度な言語の開発は、第5世代コンピュータ・プロジェクトのロジック・プログラミングに関する研究・開発が代表例である。後の三つの活動は、米国の ARPA による Knowledge Sharing Efforts に代表される。これには、次に述べるオントロジーに関する活動も含まれている。

コンピュータメディアからヒューマンメディアを指向する、ないしは、ヒューマンメディアとの接点を持つに至ったコンピュータメディアにかかわる活動としてまず挙げられるのは、オントロジー策定にかかわるものである。オントロジーとは対象をどのように見るのかという世界観とそれを記述するための基本語彙のセットのことである。これは、その分野の専門的な用語に対する用法や意味の体系を策定するもので、意味的な用語辞書へと繋がるものである。色々な分野に対する試みが行われている。

大規模な共有知識ベースを実際に実現しつつあるものに Cyc プロジェクト[Guha and Lenat 94]がある。米国 MCC において Lenat 等によって推進されたもので、エキ

スパートシステムの知識獲得ボトルネックの解消のため機械学習が可能となるための基となる常識的知識の集積を机上タイプ百科事典を対象に実現するというのが当初の目標であった。これには、若干無理な設定が含まれている。現在は、一般的なものとして時間、空間、因果関係、信念等を、専門分野として計算機、アパレル、法律、化学等を対象に知識の集積を行い、データベースの制約記述や自然言語の意味処理への利用を前面にかかげている。そして、NII（国家情報基盤）への貢献を目標に自然言語処理や電子化辞書との接点を模索する方向に向かいつつある。

人工知能の基礎的な研究として人間が持つ常識を解明し、常識ベースを実現しようという動きがある。常識というものが十分に大きな有限の知識アイテムによって成り立っているという想定に立つ。色々なアプローチがあるが、自然言語からのアプローチが有力なものとしてあげられている。まだ、初期の試みの段階である。

ヒューマンメディアからコンピュータメディアを指向する活動には、文書処理に関するものと文書の構造処理に関するものが挙げられる。ヒューマンメディアによる知識表現は人間による理解のしやすさを求めると、画像や音声を含む、いわゆるマルチメディアによる知識表現となる。しかし、知識の内容をコンピュータに理解しやすくするという要素を加味すると、自然言語がメディアの中心となる。まとまった知識を表現する自然言語による文書（文章）に対する内容検索、内容分類、内容抽出、内容要約等がコンピュータによる理解の機構の実現である。この機構は、自然言語で表現された知識の内容の一部（適切な部分）をコンピュータメディアに変換する機構と考えることができる。この機構実現に向け、文書（テキスト、ドキュメント）処理技術の研究開発の活動が大きくなりつつある。このためには、高度な言語データ、すなわち、大規模な言語知識ベースの実現が不可欠である。文書の論理的な構造に関する知識が与えられれば、コンピュータによる知識内容の理解は一段と進むことになる。SGMLなどの文書構造記述言語（一種のコンピュータメディア）の利用が拡がりつつある。さらに、SGMLの利用の普及や構造記述の高度化等を進める必要がある。これらには、現実の文書を含む言語データの詳細な分析が必要である。

ヒューマンメディアにかかわる活動は、用語等を手掛かりに情報や知識の体系や構造を研究してきた伝統的な学問分野の活動である。すなわち、情報学、図書館学、用語学などにおける活動である。知識の分類、体系、関係などを人間による洞察をもとに研究してきたが、最近、自然言語処理や知識処理等の計算機科学との接点を作り、新しい展開を求める活動が起こってきている。

### 本研究の役割

対象とする分野がかかえる課題に対する諸活動の問題点を整理し、それらの問題点のどのような部分の解決を本研究が試みたのか、すなわち本研究の役割を説明する。

本格的な自然言語処理、文書処理技術のための大規模な言語データに関する活動の問題点は次の3点にまとめられる。

#### (1) 自然言語処理技術という工学からの要請を十分にバランス良く反映したものとはなっていない。

言語学における理論的な研究の動向と、自然言語処理における処理アルゴリズム、処理システムの研究・開発の動向とは、必ずしも一致しない。これは他の分野においてもよく見受けられる純粋科学と工学との離反の例である。理論的な手続きが性能や効率を前提とする工学的な手続きにそのまま利用できるわけではない。理論的には不可能なものでも、工学的な近似としては十分実現可能なものがある。言語学においてはあらゆる言語に適用できる理論を追求するが、工学としては各言語の特性を活用して効率の良い処理アルゴリズムを求めることが第一義となる。また、言語学では言語現象は無限の多様性があるものとするが、工学では適切な有限で実効的に捕捉出来るもの、近似出来るものと看做す。

#### (2) 大規模な実現へのプロセスを十分に考慮したものとはなっていない。

対象とする言語現象を十分に網羅する規模の実現が不可欠である。言語データ各部分の



間の依存関係、制約条件を明らかにし、適切な構築の手順が明らかにされねばならない。そして、既存の資源の有効活用とコンピュータによる効果的な支援技術の開発が重要である。既存の資源としては、辞書データ、用例データ、テキストデータ等のデータ資源と作成作業を行う人的資源がある。確保できる作業者の質と量に対する正確な予測が必要である。そして、その作業者を適切に支援する高機能なシステムが必要である。この支援システムによって、作業者の個性に左右されることなく均一な結果が得られるようになること、集積の量が増大しても作業効率が落ちないこと、さらには、増大するにしたがって効率が変わらず増大していくことなどが実現されねばならない。高機能な支援システムの実現には最新の自然言語処理技術が必要であり、それ自身としても興味深いテーマである。

- (3) 統合的、総合的な観点からの設計がなされていない。単語に関する辞書、単語の意味にかかわるシソーラス、文データとしてのコーパス、文書データとしてのテキストベース、それぞれが個別にアプローチされている。語、文、文章、文書というそれぞれの構成単位に関するもの、表層情報から深い意味の情報に至るもの、異なったそれぞれの言語に関するもの、これらを統合化した言語知識のアーキテクチャが必要である。自然言語処理技術の今後の動向や効果的な開発プロセスの実現要求から生ずる要請がこのようなアーキテクチャを強く求めているからである。

大規模な知識コンテンツの構築技術に関する活動の問題点は多数挙げられるが、本研究にかかわるものは、次の1点にまとめることができる。

- (4) 自然言語を知識表現メディアとする文書ベースとコンピュータメディアである知識表現言語による知識ベースとの繋がりが明らかにされていない。人間の知的作業を支援するコンピュータシステムにとっては提示する知識は人間が理

解すればよく、コンピュータは適切な支援が行える範囲の理解で十分である。コンピュータが人間に勝る能力は、処理速度と記憶容量と信頼性におけるものである。このためには、人間の理解を目的とする文書ベースとコンピュータの理解を目的とする知識ベースを統合化するアーキテクチャが必要である。これによって、技術進歩にしたがって文書ベースの知識を知識ベースの知識へと順次変換していける機構も明らかとなる。

以上の4つの問題点の解決に向けての本研究の役割をまとめると次のようになる。

本研究が焦点を合わせる課題は、統合的な言語データ（言語知識）のアーキテクチャを明らかにすることである。このアーキテクチャは今後の自然言語処理技術の動向や効果的な開発プロセスの実現に十分対応できるようにするための基本的な条件を満たすものである。さらに、文書ベースと知識ベースを統合化するものとしても新たな提案を行うものである。ただし、本研究では開発支援システム自身は直接対象とはしない。これは今後の課題である。

本論文で述べるアーキテクチャは、このような試みとしては初めてのものであることから、説明は論理的な仕様に対する基本仕様のレベルにとどまる。そこで、アーキテクチャと言う言葉より情報構造と言う言葉を用いる。情報構造は、仕様の記述と EDR 電子化辞書という実現事例の提示によって述べられる。この言語知識の情報構造の提案は、このような仕様のものを総体として実現すべきであるという観点からのものではない。実際の言語知識の開発・整備は、各所でそれぞれの目的に沿って進められる。そのような多くの努力が納まるべき枠組みを明らかにし、効果的な協力や共有化を可能にしようとの観点からのものである。EDR 電子化辞書も、あくまでもこの情報構造の部分的な実現である。

### 1.3 論文の構成

本論文は、序論と結論を含め全体として7章で構成されている。第1章の序論に続いて、第2章では、情報構造を全体的に把握できるようにするため骨格となる基本仕様の説明がなされる。まず、基本的な要求仕様とその要求仕様の詳細化に関する技術的判断の要点が列挙される。次に、それに基づく全体構造の記述が行われ、続いて、EDR 電子化辞書がどのように対応しているのかが概観される。全体構造は、記述レベル、記述の単位、言語の種類によって特徴付けられるサブ辞書群によって構成されること、辞書項目という基本単位が互いに関係付けられることによって構造が形作られること、等が説明される。説明は、言語知識という言葉よりも電子化辞書という言葉を用いてなされる。

第3章では、記述のレベルに関して表層レベルにおける情報構造が説明される。表層レベルのサブ辞書である語表層辞書、文表層辞書、文章表層辞書、文書表層辞書と順を追って説明がなされる。各サブ辞書に対する表層辞書項目の仕様を詳細化することによって進められる。ただし、文章表層辞書と文書表層辞書は概観するにとどめる。また、説明は日本語と英語を対象になされている。最後に、実現事例として EDR 電子化辞書の表層レベル対応部分が辞書仕様と数値データの両面から概観される。

第4章では、概念レベルにおける情報構造が説明される。概念レベルのサブ辞書である語概念辞書、文概念辞書、文章概念辞書、文書概念辞書の順に説明がなされる。各サブ辞書に対し、概念というものがいかなる設定に基づくものなのか、それを表現する概念辞書項目の仕様はどのようなものなのかという点について説明が進められる。同様に、文章概念辞書と文書概念辞書は概観するにとどめる。実現事例として、EDR 電子化辞書の概念レベルへの対応部分が辞書仕様と数値データの両面から述べられる。

第5章では、情報構造の妥当性が検討される。自然言語処理技術の今後の動向にそうごと、世界知識の情報構造と連続的につながること、効率の良い開発工程が存在することの3点から検討結果が概観される。ただし、定性的な概観にとどめる。

第6章では、国内外の諸活動との比較がなされる。本研究が、諸活動の問題点の解決



を試みていること、また、諸活動の成果が情報構造の妥当性を傍証するものであることの2点から概観が述べられる。第7章では、結論として成果のまとめと、今後の展開に向けての指針の説明を行う。

付録では、EDR 電子化辞書の各辞書の辞書レコードについて第一版の仕様が例を含めて詳述される。これによって、2章、3章、4章での情報構造との対応付けや実現内容が具体的に示される。ただし、情報構造の方は、言語知識が本来持つべき論理的な構造であるが、EDR 電子化辞書の方は現実的な要請を加味したものとなっている。その部分をどう補正して対応付けるかは各章に要点が述べられている。

## 第2章

### 言語知識の基本情報構造

#### 2.1 情報構造への要求仕様

言語知識（言語に関する知識）の情報構造の基本的な枠組みを明確にするために、情報構造への要求仕様、あるいは、この情報構造を実現する電子化辞書への要求仕様を要請事項の形式で列挙する。ただし、ここでいう要求仕様とは、通常いわれる程の精密さではなく、概観的なものである。

##### (1) 十分な有用性を持つこと。

今後の自然言語処理の研究やシステム開発に十分な有用性を発揮できることである。自然言語処理の今後の技術動向に適切に沿うものでなくてはならない。形態素・構文処理に対しては頑健性の達成、意味・文脈処理への本格的な取り組み、コーパスベースなど大量データ解析に基づく言語処理、文書処理という観点からの取り組み等が今後を代表的する技術動向である。これらの技術動向に沿うためには、必要とする言語現象に対し十分な網羅性を達成すること、現実の言語現象を忠実に捕捉すること、表層の情報から深層・意味にかかわる情報までを適切に扱うことができること、大量のコーパスやテキストデータに対して対応できることなどが達成されねばならない。

##### (2) 十分な柔軟性、汎用性が得られること。

状況の変化に適応できる柔軟性や色々な状況に対応できる汎用性を十分に持つことである。このためには、安定して確保できる部分や共通の合意の得られる部分を土台にして、そうでない部分は巾を持たせて関連付けておくことである。そのようにして、言語理論や処理方式の多様性への対応、多言語への対応、多様な応用への対応などが可能になる。さらに、十分に予測できない将来の技術変化への対応や客観的な検証を可能にすることへの対応も保証されることになる。

(3) 十分な実現可能性を有すること。

大規模で十分な精度を持つものが低コストで実現できることである。情報構造がモジュール化されており漸進的な実現が可能であることである。それを利用し、目的とする機能やシステムにとって、必要な内容と規模と精度のものが容易に達成できるようにすることである。コンピュータによる十分に強力な開発支援機能が用意できることも重要であり、しかも、規模の拡大や精度の向上にともなってこの支援機能自身の能力が増大していくことである。これによって、規模の拡大が開発効率を落とさぬこと、出来うれば上昇させていくという自己増殖的な仕組みを実現することが可能になる。このような仕組みや基本的な開発素材の蓄積という環境が整えられることによって、あまり訓練を受けていない作業者も安定した効率の良い作業ができるようになる。

なお、本研究の言語知識の情報構造はこのような仕様を不足なく満たす完備した言語データの開発を進めようということからのものではない。実際の言語データは開発コストと利用価値のバランスをとりながらそれぞれに必要な部分と量をそれぞれが開発していくことになる。本研究はそのような個別の努力が共通に納まるべき枠組みを定めることを目的としたものである。これにより、個々の努力が協調し合うための共通基盤を得られるようにすること、これが要求仕様の根幹をなす。



## 2.2 要求仕様の詳細化に関する技術的判断

要求仕様から言語知識の情報構造の仕様を具体化していくにあたって考慮した技術的判断の要点を列挙する。すなわち、言語知識をどのようにすべきかの諸方式や言語知識をどのように使うべきかの諸方式に対してどのような判断を下すべきか、あるいは、下したのかを説明する。これによって、言語知識の情報構造が立脚する技術的観点が明確に示されることになる。本節の説明は第5章の情報構造の妥当性の説明に呼応するものであり、あわせて本研究の妥当性を明らかにすることを目的としたものである。妥当性の立証は他と比較する数値データや有効性を実証する利用例によって示されるのが通常である。しかしながら、本研究のように比較データを得るのが容易ではなく、また、まだ利用例が少ない場合には、本節や第5章のような記述で代替するのが適切と思われる。

以下に4つの要点についての技術的判断を説明する。ただし、後で定義される用語等を用いるが、これらに関しては常識的な理解に納まる範囲内での用法に限ることとする。

機械翻訳の方式：基本的な方式としてトランスファ方式と中間言語方式があげられる。機械翻訳システムは、この2方式のいずれかであるか、あるいは、いずれに重きをおくかが何らかの形で特徴付けられている。トランスファ方式では、構文処理の結果に対して表層レベルでの変換規則を適用して対象言語への翻訳を行う。一方、中間言語方式では、意味処理の過程を経て言語にニュートラルな中間言語に変換し、中間言語表現から対象言語表現を生成し翻訳を行う。ただし、言語に完全にニュートラルな理想的な中間言語は存在しないため、言語に依存する部分に対して中間言語間での変換規則が必要になる。

したがって、いずれの方式にしても、原文を原言語に依存する中間表現に変換し、その変換結果を訳言語に依存する中間表現に変換し、そして、それを訳文へと変換すると

いう過程を経ることになる。トランスファ方式では、表層文（原文や訳文）と中間表現との距離が短いため、中間表現への変換精度が高い。しかし、中間表現で失われる情報が少ないかわりに、不必要な情報も多く含むため中間表現間での変換が複雑になる。中間言語方式では、表層文と中間表現との距離があるため、変換精度が落ちることになる。中間表現での近似度は粗くなるが、含む情報が整理されるため中間表現間の変換も整理されることになる。このように対応付けてみるといずれの方式にもそれぞれの役割があることがわかる。また、実際に開発されたシステムをみても、特徴はそれぞれに感じられるものの方式上の優劣は無いというのが現状である。

したがって、言語知識の情報構造としては、両方式に対してもにバランス良く対応できるような仕様となるようにする。それぞれに利用する言語知識のレベルは異なるものの両方式は相似形の仕組みを持っている。このことによって、両方式からの要求をあわせた情報構造をまとめあげることが出来る。ただし、変換に関する規則類にかんしては、重要な言語知識ではあるが、システムごとに個別に設計され一般化出来る段階にはない。現段階では、規則を抽出したり、検証したりするための言語知識に焦点を合わせるが妥当である。

自然言語処理の方式：自然言語処理に関しては、形態素処理、構文処理、意味処理、文脈処理という処理ステップをどう定義するか、各処理ステップをどう行うかについて多種類の方式がある。さらに、機械翻訳を始めとした各種の文書処理に至っては、方式という段階に至らないものを含めると無数の可能性が考えられる。したがって、言語知識として基本的に対象とすべきものを適切に限定しなくてはならない。基準となる方式が定まらないものに対しては、出来るだけ素データに近いものに限定する、あるいは、新しい方式より経験の積まれた方式に基づくものを採用する等が判断基準となる。

あらゆる処理に対しての基本的な方式の違いとして規則ベースに基づくものと事例ベースに基づくものがある。この2つの方式の違いについては色々な議論があるが、変換規則の透明性を確保するための2つのアプローチとみる事が出来る。実際の変換

規則は2つのアプローチが適切に組み合わせられて構成されることになる。ある構造を持った情報がある条件のもとで別の構造を持った情報に変換するという過程をどのように記述するかに関する2つのアプローチである。事例ベースは構造の複雑さをそのままの形で扱う。変換の事例の集積という形式で過程を記述する。記述しきれない事例に対しては類推によって既知事例を援用するという大きな規則で対処する。規則ベースは構成の原理を利用して構造を分解、合成するという仕組みの上に成り立つ。構造の分解、合成、要素となる構造の変換に対しては一般的な適用範囲の広い規則で過程を記述する。構成の原理に乗らない例外に対しては部分、部分の小さな事例で対処する。

事例の方がより素データに近い。したがって、情報構造としては事例ベース的な仕組みを基本となるものとして選択する。自然言語によって表現される情報の構造の基本となる構成の仕組みは、語、句、文、文章、文書という構成単位の系列である。この系列を情報構造の基本軸に選ぶことになる。規則の個別の記述は、一般性に欠けることから第一義の対象とはしない。ただし、規則記述の基本知識となる構文カテゴリー（品詞を含む）等に関しては、経験の積まれたものを採用し、安定した大規模な集積の実現を第一義に選ぶ。

概念体系構成の方式：概念（基本となる知識）の同一性や類似性に関する知識は言語知識の根幹となる。それを具体化するものが概念体系である。体系の構成についての重要な論点は、いかなる観点にたって同一性や類似性を定義するかということである。観点は多数のものが考えられる。その中で、文や文章の中で構文上の役割を考慮したうえでの意味的な共起関係を観点とするもの、記述する時に関する事象や場面を観点とするものが代表例である。本研究は前者の観点を選ぶ。自然言語処理における意味処理への利用を主目的とするからである。類義語辞典等は後者の観点に立つ。情報検索等への利用にはこの観点の方が有効である。ただし、前者の観点は後者の観点に次のようにして結び付けられる。前者の観点に基づく意味処理プログラムによって大量の文や文章を解析する。同一の文や文章の中に互いに生起することに関する統計データを収集する。こ



のデータから後者の観点に立つ体系を構成する。

深層格や概念の設定に関する方式: 言語の意味を捕らえる方式に関しても多数の議論がある。その中で、言語学からも、自然言語処理からも、深層格に基づく方式がひとつの基準になってきている。しかし、深層格に関しても多数の議論がある。深層格としてどのようなものを設定すべきか、格を決定していく手続きの不安定さにどう対処していくのか等に関してである。

表層の文の基本は述語と表層格を伴う複数の修飾語からなる。述語も修飾語も多概念(多義)であり、表層格も複数の深層格に対応付けられる。述語と各修飾語に適切な概念を選び出し、各表層格にも適切な深層格に対応付ける。このようにして得られた概念と深層格の組が深層の文である。表層の文を深層の文に変換する過程には、現在のところ、表層の文を構文解析する過程ほどの規則性は発見されていない。したがって、この過程の記述には事例ベース的な手法を選択するのが妥当である。そして、最初の段階では、深層格の種類もあまり増やさず、概念への切り分けもあまり細かにせず、すなわち、近似を粗めにするのが妥当である。これによって辞書データ作成者の作業を安定化させることが出来る。電子化辞書があるレベルにまで完成した段階で、実際の利用からのフィードバックを得ながら、順次、近似の度合いを上げていくことになる。

## 2.3 情報構造の基本仕様

言語知識の情報構造の全体の骨格となる構造[Yokoi and Yasuhara 96]を定める。情報構造とは情報・知識が内部に持つ論理的な構造である。格納や検索の仕方についての実際の扱いに関することは含めないようにしたものである。言語知識とは言語に関する情報や知識のことである。言語知識は電子化辞書というものに具現化される。電子化辞書の情報内容が言語知識である。情報構造を説明するに当たって、理解のしやすさを考慮し、言語知識という言葉より電子化辞書という言葉の方を主に用いる。ここでいう電

電子辞書とは、通常の単語の辞書ばかりではなく、シソーラスやコーパスやテキストデータを含む統合的な言語データのことである。一般化された電子化辞書のことであり、汎電子化辞書とも呼ばれているものである[横井ほか 95]。

電子化辞書の全体構造は、記述の単位、記述のレベル、言語の種類の3点で特徴付けられるサブ辞書群をもとに組み立てられている。この特徴を座標軸に対応させると電子化辞書は図2.1に示すような構造体となる。サブ辞書が基本的なモジュールとなり、これを単位に言語知識は大枠としてのモジュール化がなされることになる。その3つの特徴について説明する。

(1)記述の単位：言語表現（言語によって表現された情報）のどのような構成単位の知識を記述するのかということである。語、句、文、文章、文書という構成単位のどれについて記述しているのかということである。意味の観点からは、このような構成単位の役割は、それほど明確なものではなくなる。本来、自然言語は要素還元的ではないからである。しかし、言語現象を整理していく立場からは外見的に識別できる構成単位という見方は有力な手掛かりとなる。

(2)記述のレベル：言語表現のどのレベルの知識を記述しているのかということである。表層のレベルから意味記述にかかわる多様な深層のレベルまでがある。表層レベルは言語表現の外見的なところで判断できる知識を対象にする。形態や構文にかかわる知識は当然であるが、意味、文脈、運用にかかわる知識も表記上で記述できるものは対象にする。これらは深層レベルの知識の判断基準を与えるものである。深層のレベルは、意味の表現形式によってさらに色々なレベルが設定される。意味表現に対し色々な研究がなされてきたが、実際には部分的な適用であったり理論的な枠組みの提案であったりで、大規模な電子化辞書に適用できる段階にはない。この中で比較的に安定したレベルとして表層に近い表現形式に基づいた概念レベルというものがひとつの基準になる。

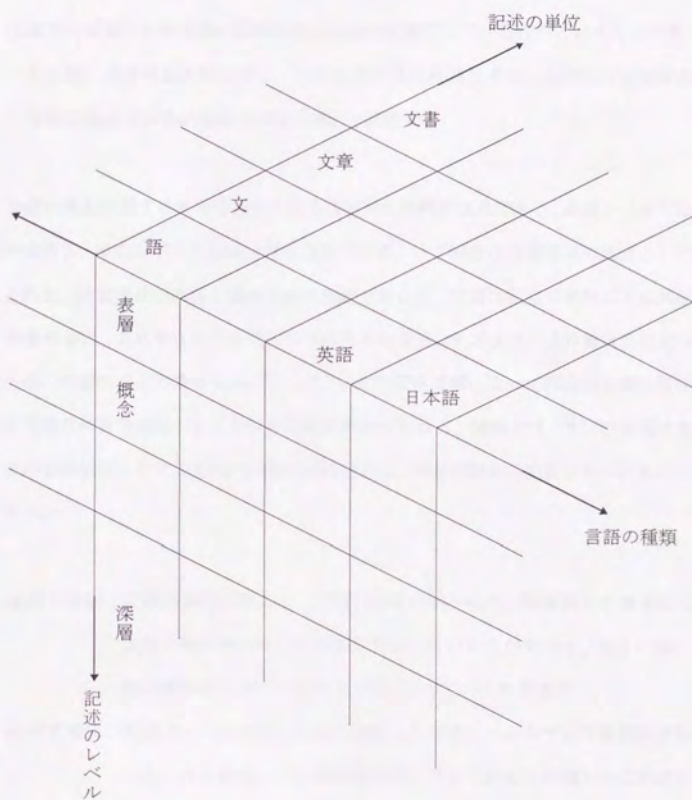


図 2.1 全体構造



(3)言語の種類：どの言語の言語表現を記述の対象にしているのかということである。

日本語、英語等基本的にはすべての言語を対象候補とする。基本的な情報構造は言語に共通であるが細部は言語の特性を反映する。

言語知識を表現するための基本となる情報単位が辞書項目である。表層レベルでは表層辞書項目、概念レベルでは概念辞書項目である。サブ辞書は辞書項目の集合として構成される。辞書項目の内容の基本部分は共通であるが、詳細は記述の単位ごとに異なる。

辞書項目は、入れ子となる多数のサブ項目からなる。サブ項目には辞書項目自身の情報や他の辞書項目との関係情報が含まれる。この関係情報によって辞書項目間の関係付けが表現され言語知識としての情報構造が形作られる。関係には、サブ辞書間にまたがる辞書間関係とサブ辞書内の辞書内関係がある。辞書間関係の主要なものは次の3つである。

(a)構成関係：記述の単位の軸にそって同じ記述レベルのサブ辞書間に定義される。

上位の構成単位のどの要素に対応しているのかを示す。あるいは、下位の構成単位のどのような文脈となっているのかを示す。

(b)対意関係：記述のレベルの軸にそって異なった記述レベルのサブ辞書間に定義される。ある記述レベルの辞書項目に対しそれより深層方向の記述レベルのもので意味表現となっているものへの対応関係である。意味表現もある種の言語による意味情報の記述であると考え、表層言語と深層言語の間の対訳関係と見なすことができる。

(c)対訳関係：言語の種類に軸にそって同じ記述レベルのサブ辞書間に定義される。異なる言語の間で同義（ほぼ同義）であるという事実が認められる対応関係である。この関係は、原則としてすべての言語対の間に定義される。

## 2.4 EDR 電子化辞書との対応

言語知識の情報構造の基本仕様に EDR 電子化辞書[日本電子化辞書研究所 95]がどのように対応しているのか、その基本的な有り様を説明する。ただし、ここで述べる基本仕様が策定されて、それにしたがって EDR 電子化辞書の開発が進められたということではない。EDR 電子化辞書の開発経験を整理し、一般化し、拡張した結果得られたのがこの基本仕様である。EDR 電子化辞書は、基本仕様の一部分の、また当面のニーズに答えるという目的からの実現である。しかし、EDR 電子化辞書は基本仕様の有効性を裏付けるに十分な実証データを提供でき、また、この基本仕様によって EDR 電子化辞書の位置付けや今後の改良・拡張の方向が明確にされることになる。

EDR 電子化辞書は、言語の種類としては、日本語と英語を対象にし、分野としては、日常一般に流通している事務文書、技術文書に対し分野共通となる部分と情報処理分野に対応する部分を対象にしている。日本語単語辞書、英語単語辞書、日英対訳辞書、英日対訳辞書、概念辞書（概念体系辞書、概念記述辞書）、日本語共起辞書、英語共起辞書、日本語コーパス、英語コーパス、日本語テキストベース、英語テキストベース、から構成されている。これらを、図 2.1 に示した情報構造の日本語と英語の部分に対応付けたものが、図 2.2 と図 2.3 である。基本語対応部分の第一版の規模に関する数値を表 2.1 にまとめる。また、EDR 電子化辞書における実現仕様が付録に詳述されている。ただし、日本語テキストベースと英語テキストベースに関する仕様は割愛する。

日本語

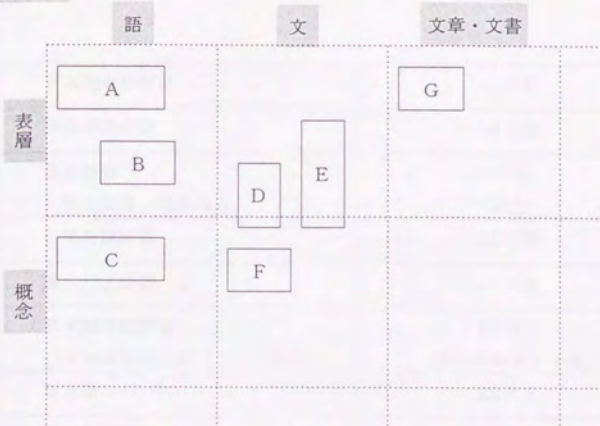
|    | 語      | 文      | 文章・文書 |
|----|--------|--------|-------|
| 表層 | イ<br>ロ | ニ<br>ホ | ト     |
| 概念 | ハ      | ヘ      |       |

- イ：日本語単語辞書
- ロ：日英対訳辞書
- ハ：概念体系辞書
- ニ：日本語共起辞書
- ホ：日本語コーパス
- ヘ：概念記述辞書
- ト：日本語テキストベース

図 2.2 EDR 電子化辞書（日本語）による実現



英語



- A : 英語単語辞書  
B : 英日対訳辞書  
C : 概念体系辞書  
D : 英語共起辞書  
E : 英語コーパス  
F : 概念記述辞書  
G : 英語テキストベース

図 2.3 EDR 電子化辞書 (英語) による実現

|                            |                       |
|----------------------------|-----------------------|
| 日本語単語辞書                    | 25 万語                 |
| 英語単語辞書                     | 19 万語                 |
| 概念辞書<br>概念記述・概念体系          | 40 万<br>概念            |
| 日英対訳辞書                     | 23 万語                 |
| 英日対訳辞書                     | 16 万語                 |
| 日本語共起辞書<br>(日本語動詞共起パターン辞書) | 90 万句<br>(基本語動詞 5 千語) |
| 日本語コーパス                    | 22 万文                 |
| 英語共起辞書                     | 46 万句                 |
| 英語コーパス                     | 16 万文                 |

表 2.1 EDR 電子化辞書の規模 (1995 年 4 月第 1 版の基本語対応部分)

## 第3章

### 表層レベルの情報構造

言語知識が表層レベル（図3.1）において持つべき情報構造[横井ほか 96a]を詳細化する。表層レベルのサブ辞書、すなわち、語表層辞書、文表層辞書<sup>1</sup>、文章表層辞書、文書表層辞書の情報構造の説明を行う。ただし、文章表層辞書、文書表層辞書については、大枠としての説明にとどめる。また、説明は日本語と英語を対象に進める。

表層レベルの言語知識が自然言語処理に果たす役割としては大きく2つある。ひとつは構文的な処理のための知識を与えることである。センテンス文法、テキスト文法、ドキュメント構造（文法）等の解析、生成のための体系的な情報を与える。もうひとつは翻訳や要約等の文書処理に対し、変換や対応付けのありのままではあるが安定した情報を与える。また、電子化辞書の開発プロセスの観点からは、表層レベルは既存の蓄積を利用でき、安定した言語直観を期待できるもので、言語知識の土台となる。

#### 3.1 表層辞書項目

表層辞書における情報の基本単位に対応して図3.2に表層辞書項目が定義される。表層辞書の情報の基本単位は次のような性質を持つものである。

- (1) 外見的に他と明確に区別しうる表記を持つこと（＜表記情報＞）。

<sup>1</sup> 句表層辞書は、語表層辞書と文表層辞書の両方の性質を持つ。ここでは文表層辞書に含めて説明する。



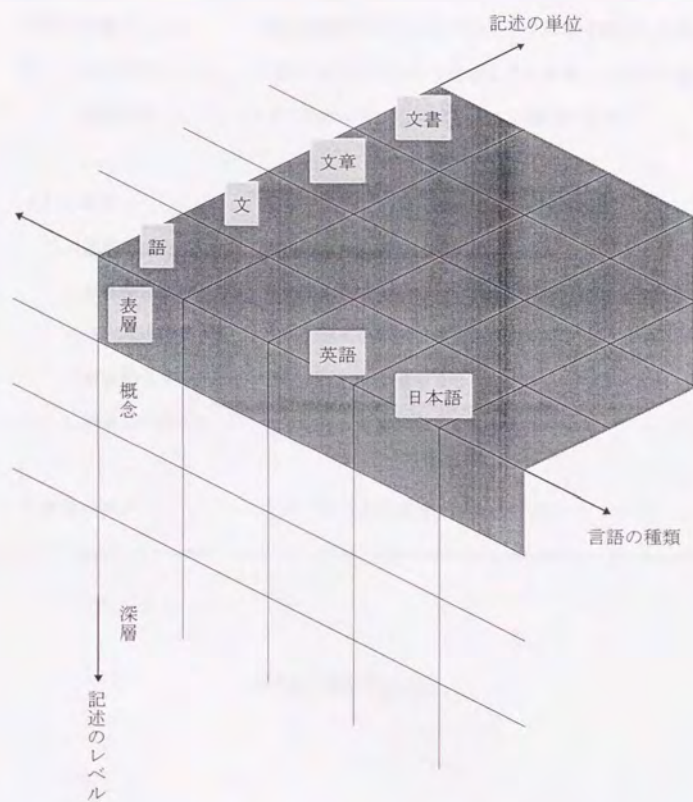


図 3.1 全体構造と表層レベル

===== [表層辞書項目] =====

- <表記情報> : 判別するための表記、標準的なものや揺れの記述
- <構成情報> : 統語的役割を中心とした上・下位の構成単位との関係
  - <要素カテゴリ> : 上位の構成単位の中でどのような要素となるかの記述
  - <構成構造> : 下位の構成単位を要素としての構造の記述
- .....
- <対応情報> : 辞書間関係、辞書内関係による対応の記述
  - <意義情報> : 意義の識別や基本的な定義
  - <意味関係情報> : 他の表層辞書項目との意味的対応
  - <構成関係情報> : 文脈となる表層辞書項目との対応
  - <対意関係情報> : 深層レベルの辞書項目との対応
  - <対訳関係情報> : 他言語の表層辞書項目との対応
- .....
- <環境情報> : 分野、状況等の運用に関する記述

図 3.2 表層辞書項目

- (2) 上位の構成構造の中でどのような要素として振る舞うのかが明確に定義できること(<構成情報>・<要素カテゴリ>)。
- (3) 対応する指示内容を確定できること(<対応情報>)。
- (4) 使用される環境(状況)を規定できること(<環境情報>)。

通常の辞書における情報単位は、表記から検索するという前提から表記の違いのみに基づいたものである。電子化辞書においては、表記、構文、意味、運用の情報をすべて等しく見なくてはならないという観点からこのような情報単位を用いる。

### 3.2 語表層辞書

語表層辞書は、語を対象として表層レベルの情報を定義する。語に対応付けられる基本的な形態素としては、語の中心的意味を表す語基(いわゆる内容形態素)と機能的・文法的意味を主に表す接辞(いわゆる機能形態素)がある。また、複数の形態素からなる派生語(語基と派生接辞によって形成される語)、複合語(語基と語基から形成される語)、接辞相当の固定した表現(「において」、「に関して」など)を含め、自然言語処理の観点から語の範囲が定められる。

表層辞書項目(図3.2)を語に対して詳細化した語表層辞書項目を図3.3に示す。<代表表記><sup>1</sup>は、語が実際に使用されるときを表記のバリエーションうち、使用頻度上、あるいは語彙体系上、代表と考えられるものである。<異表記リスト>は、その語が実際に使用されるときを表記の揺れの一覧である。日本語では、漢字の種類、送り仮名の付け方の揺れ、漢字で表すか平仮名で表すか、外来語の場合はカタカナ表記の揺れなど、様々な異表記が生ずる。英語でも、英米の綴りの違いの他、外来の語に多くの異表記がみられる。また、複合語に対し、ハイフンを付けるか付けないか、2語に分けるかなど

<sup>1</sup> 通常は発音やアクセントに関する情報も含まれるが、ここでは音声に関する事項は省略する。



===== [語表層辞書項目] =====

<表記情報>

- <代表表記> : 語を代表する表記
- <異表記リスト> : 表記のバリエーションのリスト
- <不変化部指定> : 活用語に対し、その不変化部分
- <語形成> : どのような派生語、複合語を形成するかを表示

<構成情報>

<要素カテゴリ>

<形態素情報>

- <接続情報> : 文中での環境を形態素間の接続可否の関係として表示

<構文情報>

- <統語範疇> : 構文的観点から見た時の語のクラス
- <文型情報> : 項の数と格形態を典型的な順序で示す (述語)
- <文法範疇> : 文法形式によってきまるカテゴリカルな表現の類別

<構成構造>

- <語構造> : 複数の形態素からなる語に対し、語の内部構造を表示

<対応情報>

<意義情報>

- <語義ラベル> : 語義を識別するための標識
- <語義説明> : 語義の自然言語による説明文、定義文

- <意味関係情報> : 同義語、反義語等への対応情報

- <構成関係情報> : 用例文への対応情報

- <対意関係情報> : 語概念への対応情報

- <対訳関係情報> : 訳語への対応情報

- <環境情報> : 分野、頻度、位相、語種等の語の運用に関わる情報
- =====

図 3.3 語表層辞書項目

で表記が変わる。出現の可能性のある異表記はすべて網羅される必要がある。＜不変化部指定＞は、活用語の場合、その不変化部、すなわち、文字列としての不変化部分を示す。一般には語根に対応する。この情報は屈折の語形変化表として与える方法も一般的であるが、不変部分で記述した場合は活用語尾を独立した形態素として扱うことになり、文字列の分割・連結操作のみで形態素解析・生成を行えるという利点がある。＜語形成＞は、その語が語基の場合に、どのような語を形成するかを示す。主に語彙的な関係を中心に記述し、文法的に記述可能な関係（語基と屈折接辞から形成される屈折語など）は＜接続情報＞で定義する。「切る、切れる、切らす」などの動詞の自他動の対応を含む語彙的派生関係や「釣る、釣り」などの転成も含めて扱う。これらは、明確に語基と派生接辞に分割できなくとも、形態的観点から見た関連語という情報を得るために記述しておく必要がある。

＜接続情報＞は、その語が文中でどのような形態素列の環境に現れるかを形態素間の接続可否の関係として表した情報である。この情報には、自分自身のクラスを記述する方法や、後接（あるいは前接）する形態素を直接、あるいはそのクラスによって間接的に指定する方法などがある。自分自身のクラスのみ記述する場合には、通常、接続の可否を表す文法情報が、接続テーブル等の形で必要となる。＜統語範疇＞は、構文的観点から見た時の語のクラスである。広い意味での文法範疇の一種としての品詞である。通常、品詞分類は形態・構文・意味の各観点を総合して行うことが多い。しかし、形態的観点からは既に形態情報で接続情報として形態素の分類を行っているので、ここでは構文的観点、すなわち、構文的機能・構文的意味から語および語基を分類したものを品詞分類とする。＜文型情報＞は、述語となることが出来る語基に文型を記述する。日本語では、述語が要求する項（名詞句）とその格形態（格助詞）を記述し、英語では、述語が要求する項とその位置（語順）を記述する。本来、文型は文の要素間の関係であり、文表層辞書に記述される情報である。したがって、述語と項との間の意味的な選択制限、あるいは述語が項と結び付いて表現する意味は、単語辞書の構文情報では扱わずに、文表層辞書の＜対応情報＞で扱う。ここで記述されるのは、述語の要求する表層的な格形

態のパターンによる述語の分類である。〈文法範疇〉は、〈文型情報〉と同様、文法形式によってまとめられるカテゴリカルな意味と表現の類別を示す。述語にはテンス・モード、そのうち動詞にはさらにヴォイス・アスペクトが認められる。日本語ではこれらの文法範疇は、主に接辞の付加によって形態的に示される。ただし、それぞれの形態素は幾つかの異なる機能を併せ持つ場合があるため、各形態素の接続の可否と接続する場合のカテゴリカルな意味を記述するのが効率的である。〈語構造〉は、その語が複数の形態素からなる場合、その内部構造を表示する。

〈語義ラベル〉は、語義を識別するための標識である。例えば、表層語に語義番号を添字したものなどである。〈語義説明〉は、自然言語による語義の説明文、定義文である。通常の辞書の語義文や用語辞典の用語説明が対応する。人間に語義の内容を正確かつ容易に想起させる役割をもつが、使用言語に適切な制限を加え、定義の明確化やコンピュータ処理を容易にする工夫をほどこすこともある。使用語彙セットを制限したり、説明文に表現構造を明示する等である。〈意味関係情報〉は、その語の類義語、反義語、上位語等を列挙することにより、語義を相対的に表現する。これだけでもある程度までの意味処理が可能となる。〈対意関係情報〉は、その語の語義と、より深い意味記述のレベルの項目、例えば、語概念辞書項目との対応を示す。〈構成関係情報〉は、その語を含む上位の構成単位、すなわち文表層辞書項目との対応を示す。通常の辞書の用例文の列挙に対応する。〈対訳関係情報〉は、その語と、その対訳となる他言語の語（句、文）との対応関係を示す。

〈環境情報〉は、その語が使用される分野、頻度、使用者層、品格等の位相、和語、漢語等の語種など、語の運用に関わる情報を示す。

### 3.3 文表層辞書

文表層辞書は、句や文を対象として表層情報を定義する。一般化、体系化されたコーパスがこれにあたる。図3.4に文表層辞書項目を示す。



===== [文表層辞書項目] =====

<表記情報>

<文表記> : 文の表記

<構成情報>

<要素カテゴリ>

<文タイプ> : 句、文等の統語範疇

<構成構造>

<形態素列> : 文構成要素への分割

<構文木> : 文の統語構造の表示

<対応情報>

<意義情報>

<文義ラベル> : 文義を識別するための標識

<文義説明> : 文義の自然言語による説明文

<意味関係情報> : 同義(パラフレーズ)文等

<構成関係情報> : 要素として含む文章表層辞書項目

<対意関係情報> : 文概念への対応情報

<対訳関係情報> : パラレルコーパスに対応

<環境情報> : 出典情報等

図 3.4 文表層辞書項目

＜文表記＞は、文を構成する文字列を定義する。意味記述の観点からは、複文や重文を文章のレベルと考える立場があるが、ここでは、形態的な文という区別を判断基準にする。

＜文タイプ＞は、その文自身の統語範疇を定義する。句、文等の範疇名である。＜形態素列＞は、その文を構成する各形態素への分割を示す。＜構文木＞は、その文の統語構造を示す。

＜文義ラベル＞は、＜語義ラベル＞と同様に文義の識別のための標識である。＜文義ラベル＞は、文の表層表記そのものである。多義の場合は、文義を識別するための番号を付記する。＜文義説明＞は、文の意味を曖昧性のない自然言語文で説明する。例えば、その文のパラフレーズを、可読性の高い標準的な文、すなわち、適切な語彙を用いた平易な文構造の文で定義する。＜意味関係情報＞は、同義文や反義文等の列挙である。これにより、この表層文に対しては意味解析等の重い処理を行うことなく目的の出力を得ることができる。＜構成関係情報＞は、その文が現れる文脈、つまりその文を含む上位の構造単位である文章表層辞書項目が示す。＜対意関係情報＞は、その文の文義と、より深い意味記述のレベルの項目、例えば、文概念辞書項目との対応を示す。＜対訳関係情報＞は、他言語の訳文対応を与える。つまりパラレルコーパスの機能である。単語同士の対応などのアラインメント情報も含まれる。

＜環境情報＞は、その文の出典情報などである。

### 3.4 文章表層辞書・文書表層辞書

文章および文書表層辞書は、一連の文の連鎖を対象として表層情報を定義する。ここでは、文書とは単独でまとまりのある体系的な情報を表現するもので、論文、記事等さまざまな形態をとるものであり、文章は、文書を構成する段落、段落列等を指す。いずれも、その辞書項目は、図3.2の基本構造に納まるものである。

文章表層辞書では、＜要素カテゴリ＞に起承転結のような文書上での役割を定義し、

＜構成構造＞に文章論的修辭構造を定義する。一方、文書表層辞書では、＜要素カテゴリ＞に評論、報道記事といった文書の分野、タイプを定義し、＜構成構造＞に SGML 等のマークアップ言語による文章間構造を定義する。＜対応情報＞は、文章、文書ともほぼ共通した内容を定義する。＜意義情報＞は、その文章もしくは文書をパラフレーズした文章であり、＜意味関係情報＞は、要約文章（文書）等であり、＜対意関係情報＞は、文章（文書）概念辞書項目への対応情報であり、＜対訳関係情報＞は、アラインメント情報を含む他言語による翻訳文章（文書）への対応である。文章表層辞書では、＜構成関係情報＞にその文章の出典となる文書表層辞書項目が示される。ただし、文書表層辞書ではそれ以上の記述の単位を仮定せず、定義しない。

### 3.5 EDR 電子化辞書における実現

表層辞書の情報構造の実現事例として EDR 電子化辞書の表層辞書対応部分を説明する。これによって情報構造の妥当性を立証するとともに、情報構造の具体的な実現の方法が明らかにされることになる。辞書仕様と数値データの両面から実現の有り様を説明する。

#### 3.5.1 辞書仕様

EDR 電子化辞書では、単語辞書、対訳辞書、共起辞書、コーパス、テキストベースの5種の辞書によって表層辞書を部分的に実現する。言語の種類は、日本語と英語の2種である。

語表層辞書を基本的実現したものが単語辞書と対訳辞書である。2つの辞書への分離は現実的要請による。単語辞書は、語表層辞書の主要部分を担い、その仕様をほぼ実現している。対訳辞書は、文字通り語表層辞書の＜対訳関係情報＞を与える。対訳辞書では、訳語の他に源言語との意味的關係を示す訳語種別等の情報も定義されている。こ



の有り様を、日本語単語辞書の辞書レコードの仕様（付.3 参照）と日英対訳辞書の辞書レコードの仕様（付.6 参照）の一部を図 3.5 に示すことによって表わす。EDR 電子化辞書では、辞書項目に対応するものを辞書レコードと呼ぶ。

表層辞書項目における〈表記情報〉、〈構成情報〉、〈対応情報〉、〈環境情報〉は、EDR 辞書レコードでは、それぞれ、〈見出し情報〉、〈文法情報〉、〈意味情報〉、〈運用・その他情報〉に対応している。このうち、〈対応情報〉については〈対意関係情報〉に対応する部分が〈意味情報〉として、〈対訳関係情報〉に対応する部分が〈対訳情報〉として各々実現されている。〈意味情報〉は、概念レベルの語辞書である EDR 概念辞書と共有している情報であり、単語表層辞書項目と単語概念辞書項目との直接の対応付けを行うことにより、辞書間の連続性を保証するものである。一方、〈意義情報〉、すなわち語義は、利用者の混乱を避けるため提供されていないが、やはり、〈意味情報〉と同じ形式で実現されている。語に対して定義された語義は、概念レベルの情報記述が目的とする分解能にあわせて、同義のものが一つに統合されひとつの概念に近似される。〈意味情報〉はこの近似の結果である。単語辞書および対訳辞書は、語表層辞書項目のうち、当面必要なものはすべて実現されているが、〈代表表記〉と〈異表記リスト〉の区別、〈語形成〉、〈意味関係情報〉、〈構成関係情報〉の拡充については今後に委ねられている。

文表層辞書の基本部分の実現となるのが、共起辞書とコーパスである。ただし、共起辞書は、現実的要請から、コーパスの情報の一部である語の共起に関する情報のみ抽出しコンパクトにまとめたものである。EDR コーパスの辞書レコード（付.10 参照）を図 3.6 に示す。文表層辞書項目における〈文表記〉は EDR コーパスでは〈文〉で、〈構成構造〉は〈構成要素情報〉〈形態素情報〉〈構文情報〉で、〈環境情報〉は〈出典情報〉で、それぞれ実現されている。〈構文情報〉は依存文法に基づいたものである。〈対応情報〉は基本的に実現の対象にはしていない。ただし、〈構成関係情報〉については〈テキスト番号〉という EDR テキストベースへのリンクで暗黙に表現されている。なお、EDR コーパスにおける〈意味情報〉は、概念レベルの情報であり本研究の〈対

===== [日本語単語辞書レコード] =====

|              |                        |
|--------------|------------------------|
| <レコード番号>     | : レコードタイプと識別番号         |
| <見出し情報>      |                        |
| <単語見出し>      | : 活用語尾つき見出し表記とその読み     |
| <不変化部-連接属性対> | : 構成語の不変化部とその連接属性の対の列  |
| <かな表記>       | : 不変化部のカタカナによる表記       |
| <発音>         | : 不変化部の発音とアクセント(単一語のみ) |
| <文法情報>       |                        |
| <品詞>         | : 品詞                   |
| <構文木>        | : 構成語の係受け関係(慣用句のみ)     |
| <活用情報>       |                        |
| <活用形情報>      | : 活用形(不規則活用語のみ)        |
| <活用型情報>      | : 活用型(規則活用語のみ)         |
| <表層格情報>      | : 表層格(用言, 述語句のみ)       |
| <相情報>        | : 相情報(動詞のみ)            |
| <機能語情報>      | : 機能語情報(機能語のみ)         |
| <意味情報>       |                        |
| <概念識別子>      | : 概念の同一性を示す番号          |
| <概念見出し>      | : 概念を代表する単語見出し(内容語のみ)  |
| <日本語概念見出し>   | : 概念を代表する日本語単語見出し      |
| <英語概念見出し>    | : 概念を代表する英語単語見出し       |
| <概念説明>       | : 概念の文章による説明(内容語のみ)    |
| <日本語概念説明>    | : 概念の日本語による説明          |
| <英語概念説明>     | : 概念の英語による説明           |
| <運用・その他情報>   |                        |
| <用法>         | : 語の用法                 |
| <頻度>         | : 出現頻度                 |
| <管理情報>       |                        |
| <管理履歴レコード>   | : 更新日付等の管理情報           |

===== [日英対訳辞書レコード] =====

|        |         |
|--------|---------|
| <対訳情報> |         |
| <訳語種別> | : 訳語の種類 |
| <訳語表記> | : 訳語の表記 |
| <訳語品詞> | : 訳語の品詞 |

図 3.5 EDR 電子化辞書の語表層辞書対応部分

===== [日本語コーパスレコード] =====

|            |                          |
|------------|--------------------------|
| <レコード番号>   | : レコードタイプと識別番号           |
| <文情報>      |                          |
| <テキスト番号>   | : 文管理番号                  |
| <出典情報>     | : 出典テキスト名                |
| <文>        | : 用例文の表記                 |
| <構成要素情報>   |                          |
| <構成要素列>    | : 形態素、構文、意味情報が共有する文構成要素列 |
| <形態素情報>    |                          |
| <形態素列>     | : 文の形態素分割                |
| <構文情報>     |                          |
| <構文木>      | : 文の構文構造                 |
| <管理情報>     |                          |
| <管理履歴レコード> | : 更新日付等の管理情報             |

図 3.6 EDR 電子化辞書の文表層辞書対応部分



応情報>の範囲ではない。EDR コーパスは、単言語コーパスである。今後は<対訳関係情報>の付加も望まれる。また、もともとコーパス開発は主要目標でなかったため、収集したテキストの出典に偏りがあるなどの課題が残されている。

文章・文書表層辞書に対応するものとしてテキストベースがある。ただし、現在実現されているのは<表記情報>と簡単に構成情報>のみである。

### 3.5.2 数値データ

語表層辞書に相当する EDR 電子化辞書の数値データのうち、単語辞書および対訳辞書に関するものを表 3.1 と表 3.2 に示す。表 3.1 は日本語基本語、表 3.2 は英語基本語にかかわるものである。ただし、仕様上対象としていない項目（機能語など実体概念を持たない語に対する平均概念数等）は、“-”で示す。

表の構成比とは、単語数等が全体数に占める割合を百分率で示したものである。日本語基本語に関しては、一般の冊子体辞書の見出し語の品詞分布とは若干の違いが見られ、名詞に比べ動詞や形容動詞の比率が高い。一般に大規模の冊子体辞書では、一般語の他に百科語や固有名詞等が多く含まれるため、名詞の構成比が 80%を超える事も多い。EDR 電子化辞書では固有名詞の登録を最小限に留めていることと、品詞認定に対する姿勢の違いが主たる要因である。固有名詞は多種多数になるが、一般的なもの拡充は容易であり個別のものは利用者に任せるべきであるとの判断からである。品詞認定に関しては、通常名詞としてのみ分類される語も、用言化（サ変動詞化）して用いられるものについては用言としての分類も行っている。これは用言としての振る舞いも記述しておく必要があるからである。このぶん用言の構成比率が高くなっている。平均語義数、平均概念数とは、各々単語あたりの語義数、概念数であり、意味分割の細かさ、粒度を表す指標となる。平均語義数と平均概念数との比は、語義として意味分割したものが語概念としてどの程度同一の概念として近似されているかを示す。全体的に英語より日本語の方が統合（同一化）が進んでいる事が読み取れる。英語の語義に対しては同義性の

| 品詞類   | 単語数     | 構成比(%) | 平均<br>語義数 | 平均<br>概念数 | 語義あたり<br>訳語数 | 単語あたり<br>訳語数 |
|-------|---------|--------|-----------|-----------|--------------|--------------|
| 名詞類   | 189,042 | 73.2   | 1.46      | 0.81      | 1.96         | 1.06         |
| 動詞類   | 44,556  | 17.3   | 1.73      | 0.69      | 3.09         | 1.50         |
| 形容詞類  | 2,161   | 0.8    | 1.73      | 0.73      | 3.70         | 1.59         |
| 形容動詞類 | 7,937   | 3.1    | 1.52      | 0.65      | 4.61         | 1.46         |
| 副詞類   | 4,634   | 1.8    | 1.60      | 0.65      | 3.46         | 1.76         |
| 連体詞類  | 368     | 0.1    | 1.49      | 0.74      | 3.70         | 1.46         |
| 接統詞類  | 378     | 0.2    | 1.15      | --        | --           | --           |
| 接頭語類  | 349     | 0.1    | 1.65      | 1.16      | 3.18         | 2.47         |
| 接尾語類  | 1,719   | 0.7    | 1.38      | 0.87      | 2.18         | 1.27         |
| 語尾類   | 91      | 0.0    | 1.36      | --        | --           | --           |
| 構文要素類 | 5,783   | 2.2    | 1.26      | 0.89      | --           | --           |
| 助詞類   | 189     | 0.1    | 1.01      | --        | --           | --           |
| 助動詞類  | 228     | 0.1    | 1.07      | --        | --           | --           |
| 補助用言類 | 252     | 0.1    | 1.50      | 0.59      | 2.64         | 1.44         |
| 感動詞類  | 574     | 0.2    | 1.44      | 0.77      | 2.50         | 1.23         |
| その他   | 61      | 0.0    | 1.00      | 0.80      | 4.26         | 3.84         |
| 総計    | 258,322 | 100.0  | 1.51      | 0.78      | 2.27         | 1.03         |

表 3.1 EDR 日本語単語辞書・日英対訳辞書諸元

| 品詞類   | 単語数     | 構成比(%) | 平均<br>語義数 | 平均<br>概念数 | 語義あたり<br>訳語数 | 単語あたり<br>訳語数 |
|-------|---------|--------|-----------|-----------|--------------|--------------|
| 名詞類   | 116,896 | 61.3   | 1.56      | 1.23      | 1.35         | 1.25         |
| 代名詞類  | 174     | 0.1    | 1.37      | 0.59      | 1.05         | 0.96         |
| 動詞類   | 14,573  | 7.6    | 3.29      | 1.97      | 1.70         | 2.77         |
| 形容詞類  | 38,168  | 20.0   | 1.52      | 1.30      | 1.52         | 1.36         |
| 副詞類   | 7,846   | 4.1    | 1.28      | 1.10      | 1.61         | 1.13         |
| 前置詞類  | 221     | 0.1    | 1.00      | --        | --           | --           |
| 決定詞類  | 30      | 0.0    | 1.10      | --        | --           | --           |
| 助動詞類  | 96      | 0.1    | 1.32      | --        | --           | --           |
| 間投詞類  | 380     | 0.2    | 1.19      | 0.86      | 1.62         | 0.91         |
| 接続詞類  | 19      | 0.0    | 1.00      | --        | --           | --           |
| 語尾類   | 173     | 0.1    | 1.00      | --        | --           | --           |
| 構文要素類 | 12,148  | 6.4    | 1.19      | 0.39      | 0.78         | 0.35         |
| その他   | 118     | 0.1    | 1.00      | 0.41      | 0.68         | 0.68         |
| 総計    | 190,842 | 100.0  | 1.65      | 1.24      | 1.42         | 1.30         |

表 3.2 EDR 英語単語辞書・英日対訳辞書諸元



判断が難しく近似の度合いに差が生じたものと思われる。

語義あたり訳語数は、訳語の延べ総数を語義数で割った値であり、単語の語義に対して平均いくつの訳語が与えられているのかを示す。単語あたり訳語数は、訳語の異なり数を単語数で割った値であり、訳語を、対訳言語を意味表現言語とした時の意味記述とすると、その時の意味分割の粒度を表わす指標となる。語義あたり訳語数を日本語と英語で比較してみると、全般に日英対訳の方が訳語数が多い。これは訳語を記述していく上での言語による違いを反映していると考えられる。日本語は造語能力が高いため、英日対訳における日本語による訳語は複合語表現によって適切な一語で表わすことが比較的容易である。一方、日英対訳では英語で同義性を満たすような訳語を記述しようとすると複雑な句や文で表現せざるを得なくなり、これを避けるために類義語の列挙による記述が多くなる。このことは、単語あたりの訳語数で比較した時に英日対訳の方が訳語の数が多いことにも反映されている。

平均語義数、単語あたり訳語数、平均概念数を比較してみると、日本語、英語ともにこの順番で小さくなる。これは、訳語の異なりを意味の粒度の基準として見た時に、語義はより細かい粒度で、概念はより粗い粒度で記述されていることを表わしている。これは EDR 電子化辞書の一つの特徴といえる。

文表層辞書に相当する EDR 電子化辞書の数値データとして、EDR コーパスに関するものを日本語について表 3.3、英語について表 3.4 に示す。表の品詞は形態素に対するものである。複合語に対してはそれを構成する形態素の品詞を集計してある。出典別にみた場合の構成は、日本語では、新聞 43%、雑誌 32%、教科書 14%、事典 11%であり、英語では、新聞 65%、用例集 35%である。形態素総数で見た品詞の構成比は書き言葉の文構造を反映した比率となっている。特に、日本語では、話し言葉に比較すると副詞、感動詞の比率が小さい。形態素の異なり数で見た場合は、日本語、英語ともに名詞の比率が非常に大きく、オープンクラスの単語であることが反映されている。

| 品詞類   | 形態素数(総数)  | 構成比(%) | 形態素数<br>(異なり数) | 構成比(%) |
|-------|-----------|--------|----------------|--------|
| 名詞類   | 1,375,378 | 26.1   | 110,912        | 81.3   |
| 動詞類   | 622,125   | 11.8   | 14,638         | 10.7   |
| 形容詞類  | 58,742    | 1.1    | 1,204          | 0.9    |
| 形容動詞類 | 61,192    | 1.2    | 3,796          | 2.8    |
| 副詞類   | 74,332    | 1.4    | 2,934          | 2.1    |
| 連体詞類  | 40,271    | 0.8    | 247            | 0.2    |
| 接統詞類  | 23,562    | 0.4    | 247            | 0.2    |
| 接頭語類  | 21,063    | 0.4    | 318            | 0.2    |
| 接尾語類  | 122,954   | 2.3    | 1,330          | 1.0    |
| 語尾類   | 631,304   | 12.0   | 155            | 0.1    |
| 助詞類   | 1,402,757 | 26.7   | 171            | 0.1    |
| 助動詞類  | 319,852   | 6.1    | 203            | 0.1    |
| 感動詞類  | 356       | 0.0    | 105            | 0.1    |
| その他   | 508,333   | 9.7    | 226            | 0.2    |
| 総計    | 5,262,221 | 100.0  | 136,486        | 100.0  |

表 3.3 EDR 日本語コーパス諸元

| 品詞類  | 形態素数(総数)  | 構成比(%) | 形態素数<br>(異なり数) | 構成比(%) |
|------|-----------|--------|----------------|--------|
| 名詞類  | 660,027   | 23.6   | 59,091         | 72.8   |
| 代名詞類 | 185,208   | 6.6    | 391            | 0.5    |
| 動詞類  | 369,583   | 13.2   | 8,338          | 10.3   |
| 形容詞類 | 188,115   | 6.7    | 9,217          | 11.4   |
| 副詞類  | 130,193   | 4.7    | 2,251          | 2.8    |
| 前置詞類 | 289,064   | 10.3   | 273            | 0.3    |
| 決定詞類 | 204,199   | 7.3    | 36             | 0.0    |
| 助動詞類 | 65,213    | 2.3    | 109            | 0.1    |
| 間投詞類 | 1,856     | 0.1    | 158            | 0.2    |
| 接統詞類 | 88,709    | 3.2    | 148            | 0.2    |
| 接辞類  | 198       | 0.0    | 53             | 0.1    |
| 接尾類  | 336,489   | 12.0   | 192            | 0.2    |
| その他  | 274,759   | 9.8    | 939            | 1.2    |
| 総計   | 2,793,613 | 100.0  | 81,196         | 100.0  |

表 3.4 EDR 英語コーパス諸元



## 第4章

### 概念レベルの情報構造

概念レベル(図4.1)のサブ辞書の情報構造[横井ほか 96b]を詳細化する。語概念辞書、文概念辞書<sup>1</sup>、文章概念辞書、文書概念辞書の情報構造である。ただし、文章概念、文書概念については、大枠としての説明にとどめる。また、説明は日本語と英語を対象に進める。

2つの側面から概念という言葉を用いる。一つは、言語によって表現される情報や知識(対象知識、世界知識)を概念のレベルで扱うということである。すなわち、知識のレベルで扱うという程には深い意味の扱いには立ち入らないということである。もう一つは、個別の意味の表出を包括した共通となる意味内容を扱うということである。言語表現は、具体的な文脈の中でさまざまな意味と対応付けられるが、それらを包括し一般化した形、すなわち、概念化された形の意味を基本的な対象とする。

表層レベルでは、語、句、文、文章、文書という言語表現の構成単位にそって言語知識が記述される。表層に近いということから、概念レベルにおいてもこの構成単位に沿うことになる。したがって、概念レベルは、表層レベルに対応した情報構造を持つ。

深層レベルの中で概念レベルをひとつの基準となるものとして設定する理由は以下の4つである。

1)言語学においても、自然言語処理においても、類似の考え方に基づく研究がなされ、

---

<sup>1</sup> 句概念辞書は、語概念辞書と文概念辞書の両者の性質を持つ。ここでは説明を省略する。

多くの蓄積がある。フィルモアの格文法理論やシャンクの概念依存関係等から始まった手法の延長上に概念レベルの考え方は位置している。

- 2) 言葉の意味を表層に近い浅いレベルで扱うため一般性が高い。また、表層表現の多義性を概念レベルの記述の単位として整理することで、表層レベルの記述よりより正規化された形式で言語知識を体系化することが出来る。概念レベルにおける正規化の度合は初段階のものであるが、このレベルを用いる意味処理だけでも、有用な自然言語処理機能を実現することができる。例えば、いろいろな曖昧性解消機能などである。
- 3) より深い意味の表現形式のレベルに対し、表層レベルからの変換プロセスの適切な中間ステップとなる。概念レベルを経由することによって他の深層レベルの開発が容易になる。
- 4) 表層に近いということから、表層レベルを利用して本格的な開発が可能である。

概念レベルを含む深層レベルの言語知識の役割は大きく2つ考えられる。ひとつは、表層レベルを含めて、表層表現からそのレベルの意味表現にあるいはその逆に、適切にかつ効率良く変換処理を行うための知識を与えることである。もうひとつは、そのレベルの意味表現の比較処理をして、ある意味的關係が成り立つか否か、あるいはどのような意味的關係が成り立つのか、あるいは成り立つ度合いはどのようなものか等々を判断処理する知識を与えることである。ここで意味的關係とは、同義、類義、反義、上位、下位等を指す。そして、どのくらいの複雑なものに対し、また、どのくらいの範囲のものに対してこのような知識を与えうるかによって意味表現形式の利用価値が決まることになる。概念レベルは表層表現に密に対応したレベルであるので、言葉で陽に表現される部分を扱う。会話のようにその意味が状況に強く依存する場合については、別の仕組みが必要である。概念レベルは、あまり複雑でないものに対し、そして、かなり広範囲のものに対し、意味処理のための知識を与える。概念レベルは、読み手に対して特別な了解事項を要求しない記述的な文書、すなわち、事実や自己の見解などを広く伝える

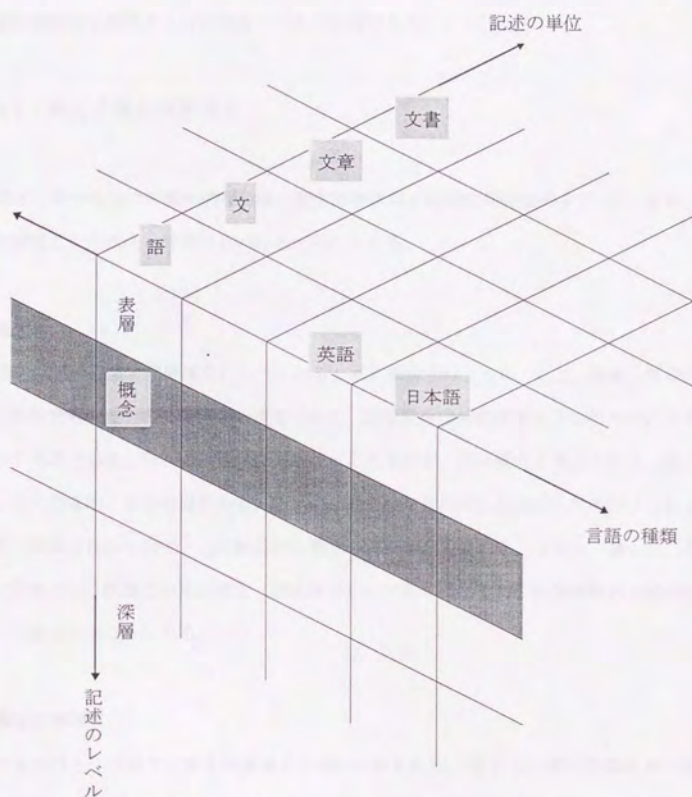


図 4.1 全体構造と概念レベル



ために書かれた記事や論文等を扱う基礎的な言語知識を与える。このように限定することによって、各種の曖昧性の解消機能の実現や検索、翻訳、要約等に対するより高度な文書処理機能を実現するのが概念レベルの役割である。

#### 4.1 概念と概念辞書項目

概念と称するものの基本的な定義と概念辞書項目の概略仕様を説明する。続く各節で構成単位ごとの概念と辞書項目の詳細化が行われる。

概念：

概念は実体概念と関係概念に分けられる。実体概念とは、もの、こと、事象、事象列など実体を有するものに対応する概念である。関係概念は実体概念どうしのかかわり方を表す概念である。ひとつの概念を実体概念と見るのか、関係概念と見るのかは、観点によって恣意的となる側面がある。ここでの概念は表層表記に直截的に対応付けられる状態で設定されるもので、この観点から概念の種別が決定される。さらに、語、文、文章、文書という表層での記述単位（構成単位）に対応して、それぞれ実体概念と関係概念が定義されることになる。

概念辞書項目：

ひとつひとつの概念に概念辞書項目が対応付けられる。図4.2に概念辞書項目の基本構造を示す。概念を識別するための情報と基本的な定義情報が<識別情報>に記述される。構成要素となる概念からその概念がどのように構成されているのか等の情報が<概念構成情報>に記述される。概念は他の概念との意味的な関係を列挙することによってさらに詳細な定義がなされる。<概念関係情報>には主要な3つの意味的な関係情報があげられている。構成関係と対訳関係は2章で述べた辞書間関係である。辞書内関係である階層関係は属種関係とも呼ばれるもので概念の上位一下位関係を表す。階層関係

| ===== [概念辞書項目] ===== |                          |
|----------------------|--------------------------|
| <識別情報>               | : 概念の識別や基本的な定義           |
| <概念構成情報>             | : 概念の構成構造                |
| .....                |                          |
| <概念関係情報>             | : 他の概念辞書項目との意味的な関係       |
| <階層関係情報>             | : 概念辞書項目間の階層関係           |
| <構成関係情報>             | : 上位、下位の構成単位の概念辞書項目との関係  |
| <対意関係情報>             | : 上層、下層の記述レベルの概念辞書項目との関係 |
| <対訳関係情報>             | : 他言語の概念辞書項目との関係         |
| .....                |                          |
| =====                |                          |

図 4.2 概念辞書項目

は概念レベルの意味処理において判断の基礎となる知識である。ただし、現在のところ明確な意味付けを持つのは語概念と文概念に対してである。

このような概念レベルの情報構造によって、語概念から構成関係をたどることによって、要素的な概念から大きな構成単位の問題がどのように構成されていくのかを段階的に記述することが出来る。逆に、文書概念から概念構成情報をたどることによってマクロな構造を把握することが出来る。そして、各々の構成単位に対して類似性を判断する知識を階層関係によって記述することで、辞書には記述しきれない多様な言語表現に対して、辞書に記述された知識を援用する手段を与えることが出来る。つまり、各々の構成単位で記述された階層関係の知識と、構成関係情報・概念構成情報とを合わせて用いることで、構成単位自身の類似性、その構成単位が置かれた文脈の類似性、その構成単位の構造や構成する下位の要素の類似性というように、多数の観点から言語知識を利用する手段が与えられる。このようにして、広範囲の処理に利用できる言語知識を統一的に整理することが出来る。

## 4.2 語概念辞書

表層レベルで語と称されているものが表す語概念と、その内容記述を行う語概念辞書項目について述べる。

語概念：

名詞、動詞、形容詞、副詞等の概念語（内容語）と通常呼ばれているものによって表されるものが実体概念である。一方、助詞、前置詞等の関係語と通常呼ばれているものによって表されるものが関係概念である。関係概念は語の位置等によって表される場合もある。この場合の関係概念は、対応する表層表現を持たぬことになる。実体概念についても、また関係概念についても対応する品詞は言語学的な観点ばかりではなく自然言語処理の観点からも実際の扱いがなされる。すなわち、通常の語と同じような働きを



示す慣用的に用いられる表現などは相当語として扱い、それらの働きに応じてひとつの概念を対応させる。

語は実際の文、文章の中で、ひとつの表層表記のままで実に多種多様な意味を表現する。電子化辞書としては、意味の記述単位である概念として何をもってくるのか、いかなる観点に立つか、いかなる粒度のものとして見るか等についての明確な指針が必要である。指針を以下の3点に整理する。まず、実体概念を中心にしてである。

#### 1)概念化されたもの

文脈上では語はほとんど個別の対象を表すが、それらのすべてに共通の性質をまとめて概念化した内容を対象とする。いうなれば、クラスを定義するものとして概念を扱う。

#### 2)慣用化されたもの

文脈上で語はさまざまなレトリカルな意味に対応付けられることがある。この場合で慣用化し、定着している場合は、派生義を原義とは独立した概念として扱う。

#### 3)一体化されたもの

多くの属性、あるいは多くの要素概念を含む一体化されたものとして扱い、属性（要素概念）への分解は別途適切な手段を講ずる。文脈上では、それぞれの属性（要素概念）が個別に対応付けられる場合が多くある。しかし、個別の属性（要素概念）に分解する作業をこの段階では行わない。

関係概念は、実体概念ほどバラエティに富むものではないが、色々な詳細化のレベルが設定できる。しかし、まずは粗い近似のレベルで電子化辞書の開発を行い、利用経験を蓄積する中で詳細化を進めるのが妥当である。

次に、以上の指針に従い、概念の定義を実質的なものにするため表層語と語概念との対応付けを行う。表層レベルにおいて、各単語に語義が対応付けられている。多義語に

は複数の語義が対応している。まずこの語義を上記の指針に従って整理する。単語ごとに整理された語義を、異なる単語の間で比較して、同義とみなされる語義どうしを統合化する。このようにして得られたものを語概念とする。

#### 語概念辞書項目：

実体概念に対する語概念辞書項目を図4.3に示す。

＜識別情報＞は辞書を利用する人間ないし計算機が語概念を識別するための情報である。＜概念ラベル＞は語概念を識別するための標識であり、概念内容を容易に想起できるように工夫されたものである。対応する表層語を利用するために語義ラベル（表層語＋語義番号）を用いる。複数の語義が統合化されている場合には、語義ラベルの集合が代表語義ラベルを用いる。＜概念説明＞は自然言語文による概念の説明である。形式等は語義説明文に準ずる。概念ラベルを補足して、概念内容を正確かつ容易に人間に理解させるのが役割である。

＜概念構成情報＞はこの語概念を要素概念に分解した時の構造である。どのようなものに分解するかについては、さまざまな立場からさまざまな提案がなされている。意味素性構造、コーパスや定義文における他の単語との関連を表すベクトル等がある。

＜概念関係情報＞は辞書内関係と辞書間関係にかかわる情報の記述である。辞書内関係には階層（属種）関係、全体－部分関係、反義関係等の意味的關係が含まれる。これらによって同一要素辞書内の語概念が関係付けられる。これらの関係の中で最も重要なものが階層関係である。その語概念の上位となる語概念（複数の可能性あり）と下位となる語概念が対応付けられる。この上位、下位の判断を行う観点を何にするかによって異なった階層関係が定義される。語概念の階層関係は、（ある発話状況で）語概念が指し示しうる内容の集合の包含関係であるが、下位概念の指示内容を包含する上位概念は複数考えられる場合が多い。自然言語処理の立場から見ると、述語概念の選択制約の処理などのために、このような階層関係のうち、文概念の中での共起の仕方と強く結びついた観点によるものが重要となる。語概念Aを含む文概念を列挙し、各文概念中の語概

=====【語概念辞書項目】=====

<識別情報>

<概念ラベル> : 語概念の識別標識

<概念説明> : 自然言語文等による概念の説明

<概念構成情報> : 要素概念による語概念の構成構造

<意味素性構造> : 意味素性による記述

..... : 色々な方式による語概念の構造の記述

<概念関係情報>

<階層関係情報> : 語概念の属種関係情報

<上位概念> : 直上位概念の列挙

<下位概念> : 直下位概念の列挙

<全体-部分関係情報> : 全体となる概念と部分となる概念への関係情報

<反義関係情報> : 反義となる概念への関係情報

.....

<構成関係情報> : この語概念を要素とする文概念への対応情報

<対意関係情報> : 表層やより深層の語辞書項目への対応情報

<対訳関係情報>

<言語 1 対訳> : 言語 1 の語概念との対応情報

<言語 2 対訳> : 言語 2 の語概念との対応情報

.....

図 4.3 語概念辞書項目



念Aを語概念Bで置き換えた時、すべてのものが文概念として成り立つ時、語概念Aを語概念Bの上位概念とするという視点である<sup>1</sup>。階層関係によって語概念の体系が形作られることになる。なお、この体系の上位には、体系構成上の必要性から、対応する表層語を持たない仮想された語概念が配置される。

辞書間関係に基づく＜構成関係情報＞は、この語概念を要素として含む文概念（隣接する文概念辞書内の文概念辞書項目）すべてへの対応付けである。＜対意関係情報＞は、語概念項目から語表層辞書項目への対応関係を与えるもので、語概念からそれを表しうる表層の語すべてへの対応付けを行うものである。また、より深層での意味記述がなされている場合は、その辞書項目への対応付けもなされる。＜対訳関係情報＞は、他のすべての言語の語概念で同義となるものへの対応付けである。同義となるものが無い場合は、一番近い上位概念への対応付けがなされる。

辞書項目内の各サブ項目も互いに関係し合う。例えば、＜概念説明＞、＜概念構成情報＞、＜階層関係情報＞は語概念の内部構造の詳細化という共通の性質を持つ。このような性質は、電子化辞書の開発工程で有効利用できる。

### 4.3 文概念辞書

表層レベルで文と称されるものが表す文概念とその内容記述を行う文概念辞書項目を説明する。文は語に比べ格段と多様になる。対象とする文の種類によって文辞書の性格が異なってくる。自然言語処理の立場からは単文を網羅し複文は文章辞書の中で扱うという考え方もある。

文概念：

文の表す事象が実体概念である。文の並び具合、接続詞等で関係概念が表される。文

<sup>1</sup> これは、観念の定義であり、実際の階層関係の判断プロセスにこのままのものが用いられる訳ではない。

の実体概念を定義していくにあたっては、文概念の内部構造をどのようなものとして見るのが重要な判断となる。自然言語処理の立場からは、次のような見方が妥当である。すなわち、その事象に何がどのようにかかわったのかを示す中核となる命題に、相、時制、極性、モダリティ等の属性を付加した構造を持つというものである[中右 94]。そこで文（実体）概念は、命題、相、時制、極性、モダリティ等の属性を付加した構造を持つ、あるいはそのような要素概念から構成される概念として定義されることになる。そして命題部分は深層格パターンの考え方にならない、述語概念（語実体概念）を中心にそれを修飾する語の実体概念と関係概念によって構成される。

表層文と文概念の対応付けは単純である。すなわち、表層レベルで表層文に割り振られた文義（文義ラベル）をそのまま文概念（文概念ラベル）に対応付ける。同義となる文概念どうしの対応付けは、陽に記述される。対応付けの仕方を知ることが重要な情報となるからである。

#### 文概念辞書項目：

文概念辞書項目を図4.4に示す。＜識別情報＞は辞書を利用する時に文概念を同定するための情報である。＜文概念ラベル＞はほぼ対応する表層文そのままを識別標識としたものである。＜文概念説明＞は表層文を可読性の高い、ないしは構造の明確な文や文章へ言い換え、意味を明確に把握できるようにしたものである。＜文概念構造＞は文概念を命題、相、時制、極性、モダリティ等の要素概念の合成構造として表現したものである。命題を表現する語概念は文という文脈では、通常はあるインスタンスを表す。ただし粗い近似として語概念のまま用いた命題の表現も利用される。この場合、命題は格パターンに対応する。

＜同義関係情報＞は、同義となる文概念の対応付けである。対応付けは文概念構造の内部構造を対応付ける仕方で行われる。＜階層関係情報＞は、文概念の上位－下位の関係である。この関係も観点によって多数のものが考えられる。その中で、表層に近く、したがって、安定した判断ができ、また、有用なものとして、次のような観点がある。

===== [文概念辞書項目] =====

<識別情報>

- <文概念ラベル> : 表層文を用いた文概念の識別標識  
<文概念説明> : 文、文章による説明やパラフレーズ

<概念構成情報>

- <文概念構造> : 語概念等を要素概念とした文概念の構造記述

.....

<概念関係情報>

- <同義関係情報> : 同義となる文概念への対応情報  
<階層関係情報> : 上位・下位となる文概念への対応情報

.....

- <構成関係情報> : 文章概念への対応情報  
<対意関係情報> : 表層やより深層の文辞書項目への対応情報  
<対訳関係情報> : 他言語の同義となる文概念との対応情報

<言語 1 対訳>

<言語 2 対訳>

.....

=====

図 4.4 文概念辞書項目



すなわち、文概念は、命題、相、時制、極性、モダリティ等の意味要素（概念属性）からなり立っているとする観点に基づくものである。それらの意味要素が特性付けられる文概念ほど下位にするという観点である。この観点は＜文概念構造＞に直接的に対応する。個々の文概念についての記述がなくとも文概念の妥当性を推論により判定することを考えると、命題、相、時制、極性、モダリティという要素概念を特性付けるという観点から文概念の階層を構成することが重要となる。すなわち、命題に関しては、詳細に規定される程下位になるという観点である。

＜構成関係情報＞はこの文概念を要素として含む文章概念への対応付けである。＜対意関係情報＞は、この文概念と表層文との内部構造を含めた対応付けである。また、より深層での意味記述がなされている場合は、それへの対応付けもなされる。＜対訳関係情報＞は他のすべての言語の文概念で同義となるものへの文概念構造の上での対応付けである。

#### 4.4 文章概念辞書・文書概念辞書

文章と文書に関する概念辞書は、大枠としては本研究の情報構造に納まるものの細部にわたっては、まだ、つめるべき課題が多い。ここでは概要だけの説明にとどめる。ここでいう文章とは、複文、段落、段落列を含む。文書とは、それ単独でまとまりのある体系的な情報を表現するもので、論文、記事などさまざまな形態をとる。

文章における実体概念はその文章が表現する事象列のことである。関係概念とは、文書の論理的な構造、すなわち、談話や論理の構成を表現するものである。文章の概念構造は、文の実体概念を関係概念によって関係付けたものである。文章概念は文概念が成立する環境として機能するものであり、自然言語処理の立場からは、文概念の解釈の妥当性を判定する機能が求められる。ただし、現実世界に対する精密な知識に踏み込まずに判断するものであるから、妥当性も文章の表面から読み取れる談話や論理の流れとの整合性に基づくものである。

同様に文書の概念構造は、文章の実体概念と関係概念によって文書構造を表現したものである。概念レベルにおいて語概念から概念構成情報をたどることによって、ミクロなものから大きな構成単位の問題がどう構成されていくのかを把握することができる。逆に、文書概念から概念構成情報をみることによってマクロな構造を把握することができる。この両面からの概念構造の把握という仕組みは非常に重要である。

#### 4.5 EDR 電子化辞書における実現

概念レベルの情報構造の実現事例として EDR 電子化辞書の概念辞書対応部分を説明する。これによって情報構造の妥当性を実証するとともに、情報構造の具体的な実現の方法が明らかにされることになる。辞書仕様と数値データの両面から実現の有り様を説明する。

##### 4.5.1 辞書仕様

情報構造は、電子化辞書が全体として持つべき情報の種類と相互の関係という立場で仕様を検討したものである。一方、EDR 電子化辞書は辞書を利用者に提供するに際しての現実的な要請を加味して仕様が決定されている。そのために生じた構造上のずれを補正しつつ説明する。また、EDR 電子化辞書は日本語と英語に対し、日常一般に用いられる基本語と情報処理用語を対象にしているが、ここでは基本語に由来する部分に全体を代表させる。

語概念辞書の基本部分の実現となるのが EDR 電子化辞書の概念見出し辞書と概念体系辞書である。文概念辞書の部分的な実現になるのが概念記述辞書、共起辞書、EDR コーパスである。なお、EDR 電子化辞書では、概念見出し辞書、概念体系辞書、概念記述辞書を併せたものを概念辞書とよんでいる。これは利用者の便宜を考慮してそのようにまとめたものである。

辞書項目に対応するものが辞書レコードである。図4.5に概念見出し辞書と概念体系辞書の辞書レコードの構造(付.5 参照)を示す。概念見出し辞書レコードの内容が語概念辞書の<識別情報>に対応する。<階層関係情報>を上位概念と下位概念の対に分解したものが概念体系レコードである。これは管理の便宜上の処置である。階層関係によって構成される全体の構造物を概念体系と呼んでいる。その上位層部を図4.6(a)<sup>1</sup>に、体系をたどった例を図4.6(b)に示す。(a)の概念名で単語以外の形式のものは、対応する表層語を持っていない。<構成関係情報>は検索によって得られるため陽には示されていない。<対訳関係情報>には、日本語由来の概念と英語由来の概念を一緒にして、同義や階層関係の処理を行うことによって対応している。

文概念辞書に対応する部分を含む辞書レコードを図4.7に示す。EDR コーパスでは、一文が一文義、一文概念である。EDR コーパスレコード(付.10 参照)には表層レベルの情報も併せて記述されている。したがって<概念ラベル>は省略されている。<意味情報>が文概念辞書の<文概念構造>に対応する。これは、命題を文中の自立語の概念と概念関係子のネットワークで表現し、相、時制、極性、モダリティ(発話の意図、話者の視点、文体等)を補助的な構造として付加し表現したもので機械翻訳における中間言語[国際情報化協力センター 93]としても使用できるものである。文概念の命題部分に対し、階層関係の上位にあるものが概念記述辞書レコード(付.5 参照)の<記述タブル>である。これは、述語となる語概念ともう一つの語概念とを関係概念<sup>2</sup>で結んだという最簡命題概念である。最簡命題概念は、最も要素的な文概念(命題)であるが、依存関係の強い述語概念と修飾概念の組み合わせは、その述語が取りうる全体の格パターンと同等の制約として働きうるものである。この最簡命題概念には、コーパスの文概念に結びつけられたものと、概念体系の上位層で作成した、表層表現を持たぬものとに分けられる。この区分が<記述区分>に記されている。表層表現との対応付けを陽に記述

<sup>1</sup> もう一段下位の概念を含めたものが付.5の「概念体系の主要部」に記述されている。

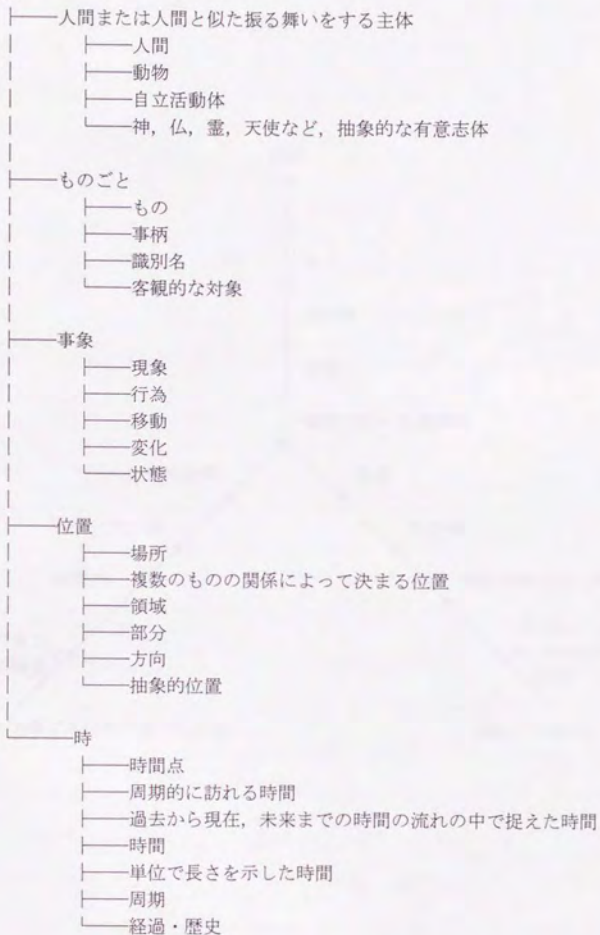
<sup>2</sup> 関係概念を表わすものを関係子と呼ぶ。用いられたものが付.5の「関係子一覧」に列挙されている。



|                           |                 |
|---------------------------|-----------------|
| ===== [概念見出し辞書レコード] ===== |                 |
| <レコード番号>                  | : レコードタイプと識別番号  |
| <概念識別子>                   | : 概念を一意に特定する整数  |
| <概念見出し>                   | : 概念に意味を表現する単語  |
| <概念説明>                    | : 概念の内容を表現する説明文 |
| <英語概念説明>                  | : 英文による説明       |
| <日本語概念説明>                 | : 和文による説明       |
| <管理情報>                    | : レコードの管理履歴情報   |
| ===== [概念体系辞書レコード] =====  |                 |
| <レコード番号>                  | : レコードタイプと識別番号  |
| <上位概念識別子>                 | : 上位となる概念の識別子   |
| <下位概念識別子>                 | : 下位となる概念の識別子   |
| <管理情報>                    | : レコードの管理履歴情報   |
| =====                     |                 |

図 4.5 EDR 電子化辞書の語概念辞書対応部分

概念



(a) 基本語概念体系の上位層部



(b) 概念系列の例

図 4.6 基本語概念体系



===== [EDRコーパスレコード] =====

|          |                     |
|----------|---------------------|
| <レコード番号> | : レコードタイプと識別番号      |
| <文情報>    | : 文識別子と文見出し         |
| <構成要素情報> | : 文を構成する形態素列        |
| <形態素情報>  | : 形態素、複合語の構成情報      |
| <構文情報>   | : 構文木               |
| <意味情報>   | : 文概念の構成構造 (中間言語表現) |
| <管理情報>   | : レコードの管理履歴情報       |

===== [共起辞書レコード] =====

|             |                        |
|-------------|------------------------|
| <レコード番号>    | : レコードタイプと識別番号         |
| <見出し情報>     | : 共起句見出し               |
| <共起句構成要素情報> | : 共起句 (最簡文) を構成する形態素列  |
| <構文情報>      | : 共起句 (最簡文) の構文構造      |
| <意味情報>      | : 共起句 (最簡文) 概念の構成構造    |
| <共起状況情報>    | : コーパスにおける出現頻度と出現部分の周辺 |
| <管理情報>      | : レコードの管理履歴情報          |

===== [概念記述辞書レコード] =====

|          |                |
|----------|----------------|
| <レコード番号> | : レコードタイプと識別番号 |
| <記述区分>   | : 最簡命題概念の区分    |
| <記述タブル>  | : 最簡命題概念の構成情報  |
| <管理情報>   | : レコードの管理履歴情報  |

図 4.7 EDR 電子化辞書の文概念辞書対応部分

したのが共起辞書レコード（付.8 参照）である。ただし、これも最簡文、最簡命題に対するものである。最簡命題は文概念体系上、上位層に位置する。そのために捨象される部分が多くなりすぎるきらいがある。そこで中間層に位置するものとして、日本語の基本動詞に対し設けたものが日本語動詞共起パターン副辞書に記述されている。

#### 4.5.2 数値データ

語概念辞書の<階層関係情報>に相当する EDR 電子化辞書の数値データとして、表 4.1 に概念体系の規模に関するデータを、表 4.2 に概念体系の全体構造に関するデータを示す。

表 4.1 で、親概念数とはある概念の直接の上位として関係付けられた概念の異なり数である。親概念数が 2 以上となっている概念は、2 つ以上の上位概念を持っており、いわゆる多重継承の構造に対応するものである。概念体系における多重継承は排他的でない複数の観点からある概念が多重に分類されている場合に生ずる。表 4.1 は約 5 % 程度の概念しか多重に分類されていないことを示している。このことは、ひとつには 38 万概念という膨大な量の概念を分類していくための工数上の制約から、第一版としては基本的な部分のみに対応したからである[Nakao and Ogino 94]。また、もうひとつは語概念のもつ性質のうち文脈によらないもののみを考慮したことによる。すなわち、文脈毎に指し示すものによって真偽が異なってくるような性質では分類はしない。例えば（単なる）「皿」という語の概念について「一般に丸いものである」（形の観点）とか「瀬戸物であることが多い」（材質の観点）などの性質では分類せず、「形のある具体的な物質で出来たもの」として分類する。これは強い連関はあるが実際の真偽は文脈に依存する性質は、概念体系という静的な枠に収めきれものではなく、事例などから動的に推定するなどの別のメカニズムで取り扱うべきであるとの考えからである。EDR 電子化辞書では、このような考えに対しては、概念記述や共起辞書、あるいはコーパスが情報を提供することになる。

表4.2は概念体系の階層における概念の分布の様子を示している。表4.2では、類似の概念を下位概念としてまとめている概念、すなわち、概念項目<sup>1</sup>を、その子孫概念<sup>2</sup>の多い順に7グループに分け、体系の位置の目安とした<sup>3</sup>。そして、グループ毎に、含まれる概念項目の直接の下位概念（子概念）の異なり数を集計した。このように分けた概念項目のグループは、原則として多くの子孫概念を持つ概念項目のグループ程体系の上層に位置しているので、ある概念がいずれのグループの概念項目の子概念となっているのかは、その概念の弁別性のひとつの指標となる[Nakao and Ogino 94]。概念体系の上層程概念の属性に対する規定が緩く、概念体系による概念の弁別性が小さくなるからである。例えば、「上層1000項目レベル」の概念項目の子概念は、概念数全体の1000分の1オーダーの数の概念群と姉妹関係<sup>4</sup>にあることになり、1000個レベルの異なりを弁別できる程度の情報が与えられていると考えられる。EDR 電子化辞書の基本語概念体系の約半数の概念は、少なくとも1000項目レベルの弁別性を持っている。50項目レベルより小さい弁別性しか持っていない約4%の概念の多くは、「物事」に対応する概念等で、指し示す内容自体が広くそれ以上の分類が困難なものである。100項目-1000項目レベルの弁別性を与えられている概念には、分類の観点を増やせばさらに詳細な分類が可能なものもある。しかしながら、実際に行うには、現在の数千オーダーの分類項目を数万オーダーに増やすことになり。この場合、作業者が分類項目全体を把握することが困難になるので、分類作業の方法を大きく見直す必要がある。

表4.3に文辞書の<階層関係情報>における最簡命題概念に相当する<記述タブ

<sup>1</sup> 通常のシソーラスにおける分類カテゴリの機能を持つ概念。表層レベルの単語に対応を持つものと持たないものがある。前者は下位概念を持つという点以外には概念項目でない概念との区別は無い。後者は、概念体系を構成するために導入された特別な概念である。

<sup>2</sup> 概念体系でその概念の下位に位置している全概念。すなわち、直接の下位概念、下位概念の下位概念等々をすべて含む。

<sup>3</sup> 多重継承を持つ概念体系を正規のN進木に近似し、上層-下層の軸に沿ってゾーンを設定する操作に対応。体系が深さが一定のN進木の形をしている場合は、子孫概念の数が大きい概念ほど体系のルートに近い。

<sup>4</sup> 同じ概念を直接の上位とする概念同士、すなわち、姉妹概念同士の間の関係だけでなく、姉妹概念の下位概念となるもの同士の関係も含める。



>の規模を示す数値データをあげる。表中、「記述タプル」とは、概念記述辞書で記述された概念関係の数であり、「事例数」とは、日本語・英語コーパスのそれぞれに現れた概念関係の数である。「構成比」は、辞書に記述されている8つの関係の合計数を基準とした百分率で示されている。「重複度」は、コーパス中の関係事例の延べ数と異なり数との比である。事例数には、固有名詞に対応する概念など、EDR 概念辞書には登録されていない概念に関する事例も含まれている。「その他」の関係とは、名詞と名詞の修飾関係に対応する関係や、述語概念と時間を指示する概念との関係、述語概念どうしの時間的な前後関係など、格パターン以外の関係である。なお、コーパスには、実際の文では省略されている要素に対応する概念の関係や文の構造に関する関係なども含まれているが、表4.3にはそのような関係は含めてはいない。

| 親概念数 | 概念数     | (比率)     |
|------|---------|----------|
| 1    | 360,454 | (94.6%)  |
| 2    | 19,073  | (5.0%)   |
| それ以上 | 1,608   | (0.4%)   |
| 合計   | 381,135 | (100.0%) |

表 4.1 基本語概念体系における一概念当たりの親概念数

| 概念項目のグループ |              |    | 子概念数 (比率) |         |
|-----------|--------------|----|-----------|---------|
| 上位        | 1- 10        | 項目 | 4,435     | (1.1%)  |
| 上位        | 11- 50       | 項目 | 11,326    | (2.9%)  |
| 上位        | 51- 100      | 項目 | 17,197    | (4.3%)  |
| 上位        | 101- 500     | 項目 | 105,892   | (26.8%) |
| 上位        | 501- 1,000   | 項目 | 61,440    | (15.5%) |
| 上位        | 1,001- 5,000 | 項目 | 167,733   | (42.4%) |
| 最下位       |              |    | 27,879    | (7.0%)  |

表 4.2 基本語概念体系における概念の分布



| 関係名       | 記述タブル   |       | 日本語事例数 (25万文中) |        |           |      | 英語事例数 (20万文中) |        |         |      |
|-----------|---------|-------|----------------|--------|-----------|------|---------------|--------|---------|------|
|           | 異なり     | 構成比   | 延べ             | 構成比    | 異なり       | 重複度  | 延べ            | 構成比    | 異なり     | 重複度  |
| agent     | 40,221  | 7.9%  | 96,816         | 10.5%  | 79,416    | 1.22 | 118,808       | 20.2%  | 74,248  | 1.60 |
| object    | 282,800 | 55.8% | 514,056        | 55.8%  | 388,957   | 1.32 | 285,263       | 48.4%  | 232,038 | 1.23 |
| a-object  | 46,161  | 9.1%  | 73,602         | 8.0%   | 66,036    | 1.11 | 64,298        | 10.9%  | 57,231  | 1.12 |
| goal      | 46,138  | 9.1%  | 99,105         | 10.8%  | 74,459    | 1.33 | 43,453        | 7.4%   | 37,699  | 1.15 |
| implement | 20,043  | 4.0%  | 25,983         | 2.8%   | 24,020    | 1.08 | 13,492        | 2.3%   | 12,894  | 1.05 |
| cause     | 9,623   | 1.9%  | 9,547          | 1.0%   | 9,067     | 1.05 | 5,743         | 1.0%   | 5,468   | 1.05 |
| place     | 26,039  | 5.1%  | 49,519         | 5.4%   | 42,833    | 1.16 | 29,053        | 4.9%   | 25,381  | 1.14 |
| scene     | 36,181  | 7.1%  | 53,071         | 5.8%   | 47,946    | 1.11 | 29,407        | 5.0%   | 27,823  | 1.06 |
| 小計        | 507,206 | 100%  | 921,699        | 100%   | 732,734   | 1.26 | 589,517       | 100%   | 472,782 | 1.25 |
| その他       | —       | —     | 1,095,061      | 118.8% | 850,091   | 1.29 | 577,275       | 97.9%  | 477,993 | 1.21 |
| 総計        | 507,206 | 100%  | 2,016,760      | 218.8% | 1,582,825 | 1.27 | 1,166,792     | 197.9% | 950,775 | 1.23 |

表 4.3 概念記述の規模(主要関係子)

## 第5章

### 情報構造の妥当性

言語知識（電子化辞書）の情報構造が、2.1節で述べた要求仕様を満たし、諸技術の動向に合致する妥当なものであることを説明する。妥当性を、自然言語処理にとっての有用性と柔軟性と汎用性、世界知識（世界知識ベース）実現への有力なアプローチであること、開発プロセスから見た現実可能性の3つの観点から説明する。

#### 5.1 自然言語処理からの検討

自然言語処理における課題を問題解決の立場から体系付けることによって、電子化辞書とのかかわり方が明確に整理される。課題は、次の様な問題に整理される。

判定問題：言語情報がある条件に合うか否かを判定する。あるいは、合う度合いを求める。条件としては、形態的、構文的、意味的に正しいか（ほぼ正しいか）否か、可読性の良さ等。

（例）スペルチェック、スタイルチェック、可読性判定。

比較問題：言語情報同士がある条件に合うか否かを比較する。あるいは、比較した結果の度合いを求める。条件としては同一であるか否か、類似しているか否か等である。

（例）キーワード検索（語と語）、フルテキスト検索（語と文書）、類似語検索（語と文書）、類似文検索（文と文書）、連想検索。

単純変換問題：言語情報を同種（同系統）のメディアの言語情報に等価変換する。

（例）かな漢字変換、点字変換、解析・生成。

翻訳変換問題：言語情報を別種（別系統）のメディアの言語情報に等価変換する。

（例）機械翻訳（対外国語）、自然言語インタフェース（対形式言語）、自然言語理解（対形式言語）。

要約変換問題：言語情報を同種（同系統）のメディアの言語情報に縮約変換する。

（例）キーワード抽出、カテゴリ抽出、知識フレーム抽出、アブストラクト生成。

問題解決の過程は分割と合成の操作（divide and conquer）による基本手続からなる。重要なことは、互いに独立で適切な副問題にどのように分割するかということである。自然言語処理における問題では、複雑に絡み合う副問題に対処しなければならない。したがって、工学的な観点からはいかに独立したものに近似するかが要点となる。副問題への対処法から問題解決の基本手続は、規則ベースに基づく手続と事例ベースに基づく手続に分類される。実際には、2種類の手続が適切に組合されて基本手続が構成される。規則ベースの手続は、副問題への分割という方針に直接的にそった手続である。そのため、場合によっては、副問題間の依存性（非独立性）の処理のために手続全体が煩雑なものになる。事例ベースの手続は、問題とその解に関する事例から類推するという手続であり、副問題間の依存性への別の対処法である。

言語知識の情報構造が以上のような自然言語処理における問題解決の構造に適切に対応付けられることを説明する。副問題への分割は、基本的に言語表現の表層的な構成構造に沿ってなされる。すなわち、文書、文章、文、句、語という構成単位と相互の構成関係が基準となる。言語知識の情報構造の記述単位に沿った構造がこの基準に対応し、構成関係にかかわる辞書項目の記述内容が構造と要素に関する詳細な知識を与えてくれる。記述の単位にそって各サブ辞書が基本的に同じ構造をしていること、すなわち、同種の知識をすべてが持っているということは、規則ベースの手続と事例ベースの手続



のいずれにも適切に対応しうることを保証している。検索や要約の問題に関しては、辞書内関係に関する記述内容が対応し、翻訳の課題に関しては対訳関係の記述内容が対応する。意味処理への本格的な取り組みに対しては、概念や、さらには、より深層の記述レベルのサブ辞書が知識を与えてくれる。すべての記述単位に対応した深層の記述レベルが設定されているということは、表層の知識と深層の知識の選択や組み合わせを問題ごとに適切に行えることを保証している。深層の知識を利用することによって、問題解決の手続きは、曖昧度の低い、より簡潔なものになる。ただし、処理の負荷が増すこと、近似の粗さが精度の低下に結びつく場合があることなどが生ずる。

妥当性をより具体的に検証するために、言語知識の利用事例を挙げる。この事例は EDR 電子化辞書を用いた日英機械翻訳システム[Yasuhara 93][Komatsu et al. 94]である。このシステムは実験システムとして研究開発された段階のものである。したがって、翻訳率のような数値データを挙げて検証できるものではなく、システムの方式の斬新さで有用性を訴えようと言うものである。

このシステムは、中間言語方式事例ベース機械翻訳システムと呼ばれるものである。翻訳の仕組みを言語知識の情報構造に照らして説明すると次のようになる。すなわち、まず原文を、原言語の文表層辞書と文概念辞書の対応を事例ベースとして用いて、文概念の表現（中間言語表現）に変換する。類似性の判定には語概念辞書の階層関係情報を用いる。次に、中間言語表現を文概念辞書の対訳関係情報から抽出した規則を適用して訳言語の中間言語表現に変換する。最後に訳言語の文概念辞書と文表層辞書の対応を利用して訳文を生成する。EDR 電子化辞書を利用する実際のシステムでは、文表層辞書と文概念辞書の対応は日本語コーパス、英語コーパス内の情報を用いて実現している。

通常、事例ベース翻訳といわれる方式では、表層レベルでの対訳事例を用いる。言うなれば、これはトランスファ方式事例ベース機械翻訳ともいうべきものである。当然のことながら、本言語知識はこの方式にも対応できる。両事例ベース翻訳それぞれの長所・短所はほぼそのまま両方式の長所・短所を反映するものと思われる。したがって、いずれがと言うよりも、新たな可能性を示すことによって中間言語方式事例ベース機械

翻訳は、本研究の情報構造自身が多くの可能性を包含していることを実証したことになる。

## 5.2 世界知識の情報構造からの検討

言語知識の情報構造が世界知識の情報構造のプロトタイプとなっていることを説明する。記述単位の軸に沿って見た時、言語データとしては語から文書へ、世界知識としては文書から語へ向かって言語知識に対応付けられる。

自然言語処理の立場からは、語の方からのサブ辞書を重要視する。それは、言語現象を安定して網羅的に捕捉することを重視するからである。世界知識にアプローチしようという立場からは、文書のサブ辞書から言語知識を見ることになる。脈絡のある知識を実際に利用できる形態にまとめたものが文書であるからである。知識の種類によって文書の形式は様々となるが、語のサブ辞書に向かうにしたがって知識の種類にはよらないものになる。

世界知識のプロトタイプとしての言語知識は、標準となる知識事例の集積である文書辞書と、使用されることが予想される語を網羅した語辞書、この2つからなる情報構造が初期状態である。この状態から文章辞書や文辞書の必要な部分が補充され、全体として深層レベルの辞書が充実されるにつれ、世界知識としての情報構造を完備していくことになる。

世界知識を利用する最も一般的な知識処理のタスクは当然のことながら新しい知識を作成するというタスクである。すなわち、その分野の知識として必要な制約条件を満たす文書を作成するというタスクである。知識処理におけるこのようなタスクの支援機能の実現は自然言語処理における文書作成支援機能の実現と大きく重なる。

知識をどう使うのかというメタな知識の扱いについては、ここでは省略する。言語知識においては、通常はメタな知識は対象にしない。解析や変換の処理手続そのものの記述に直接かわる知識ではなく、処理アルゴリズムの実現に間接的に利用できる事例的

な知識にしろるのが通例である。世界知識においても、これと同様な考え方を採用することが知識の汎用性を確保するための現実的な方策である。

### 5.3 開発工程からの検討

言語知識の情報構造にそれを実現するための適切な開発プロセスが対応付けられることを説明する。サブ辞書が開発プロセスの基本単位、基本モジュールに対応する。どの程度安定した内容の記述作業が行えるのかという安定性、既存の素材からどの程度安易に情報が得られるのかという容易性、言語現象をどのくらい捕捉できるのかという網羅性、この3つの性質の度合いからどのサブ辞書から着手するのか、辞書項目のどのサブ項目から記述していくのが決定される。一般的に、サブ辞書に関しては記述レベルの表層に近いほど安定性が高く、記述単位から見ると語と文書の容易性が高く、記述単位が小さいほど網羅性が達成しやすい。言語の種類に関しては、母国語の周辺になるほど作業者の確保や既存の蓄積の利用が容易になる。

実際の開発プロセスは、言語直観に基づく作業者と辞書項目間の制約条件をチェックする支援システムの協調作業となる。最近、コーパスからの言語知識の獲得に関する研究が活発化してきている[宇津呂、松本 95][Zernik 91]。このような研究を本論文に述べた情報構造にそって進めることによって支援システムに高度な機能を実現することができる。



## 第6章

### 他の事例との比較

情報構造の妥当性の検討をさらに進めるために、国内外の他の代表的な事例との比較検討を行う。表層レベル、概念レベル、知識レベルとの関連と言う3点で、比較を行い、本研究の意義と妥当性を明らかにする。ただし、検討は概観の範囲にとどめる。

#### 表層レベル

事例としては辞書として分類されるものとコーパスとして分類されるものがある。辞書として分類される事例として、欧州の *Acquilex*、*Multilex*、*Genelex*、米国の *Complex*、日本の *IPAL* 辞書、*ALT* 辞書、*JICST* 対訳辞書を採り上げる。

*Acquilex*[Briscoe et al. 90][Briscoe 91]は、機械可読辞書(MRD)からの半自動的な語彙知識の獲得を目的としている。最終的に獲得される語彙知識項目は、統語情報と意味情報が同じ枠組み(タイプつき素性構造)で記述され、単一化ベースの自然言語処理で直接利用できるものである。語彙知識項目は、MRD(正確にはその内容を分析し構造化したデータベース)の持つ情報を予め定義したタイプシステムを通じて展開することにより得られる。主な関心は、既存のMRDに内包された知識をいかに形式化し利用するかにあり、本研究や他の事例のように、MRDが内包しない知識をも対象として電子化辞書に要請される情報構造を検討し、設計開発を行っていく立場とは異なる。

*Multilex*[Serasset 94]は、多言語間の辞書データベースの基準を設け、実験することを目的としている。*Multilex* データベースは、言語ごとにひとつの単語辞書と言語対ご

とにふたつの対訳辞書から構成される辞書の集合体である。Multilex で提案されている単語辞書は、Lexical Units(LU)と呼ばれる項目から構成され、LU は、語表層辞書項目と同様に一つの意味に対応するように設定される。各 LU は、表記・音韻・形態的単位(GPMU)、統語情報、意味情報、相互参照(同義語、反義語などへのリンク情報)、変換(対訳辞書項目へのリンク情報)、管理レコードを持ち、本研究の語表層辞書の情報構造にもっとも近い構造をもっている。

Genelex[Normier and Nossin 91]は、共通の辞書仕様のもとで、様々なヨーロッパ言語に対する汎用辞書を開発することを目的としている。特定の言語処理応用システムは想定しない。辞書項目の単位は、形態単位と呼ばれ、表記が同じでも品詞が異なれば分割される。各形態単位は、1 つ以上の統語単位と意味へのリンクを持つ。汎用性を重視しあまり複雑な情報構造はもたないが、そのかわり大規模な実現を見ている。

Comlex[Grishman et al. 94]は、大規模テキスト処理研究のための共有資源となることを目的とし、商用を含めて広く一般に提供されているものである。既存の MRD を使ってひな型を作りその後人手作業を加える戦略により、開発効率が非常に高い。Comlex の特徴は、意味の領域には踏み込まず、ユーザに共通の理解がある構文構造の記述、特に、下位範疇化素性の定義に的を絞っていることである。ただし、意味情報については、WordNet[Miller et al. 93]との統合を狙っている。また、収録対象は、オープンクラスの単語、すなわち、名詞、動詞、形容詞、副詞に焦点をあて、他の品詞についてはあまり重要視していないことも特徴である。word structure と呼ばれる辞書項目は、見出し表記、品詞、語形、テンス、単複、下位範疇化素性、文法属性、変更履歴からなり、本研究の語表層辞書項目の部分的実現に留まっただけのもの、すばやい提供を可能としたことにその意義がある。

IPAL 辞書[情報処理振興事業協会 87][情報処理振興事業協会 90]は、語数を制限し基本的な語についてのみではあるが、非常に綿密な分析と記述を行っている。特に、文型情報と文法範疇の内容は充実している。

ALT 辞書(機械翻訳システム ALT 用辞書)[池原ほか 93]は、自然言語処理、特に機械

翻訳において、表層に近いレベルの意味解析を実用規模で実現することを目的として開発されている。ALT 辞書では、表現の仕方（表層構造）を意味の一部ととらえており、これを端的に表す構文意味辞書と呼ばれる用言の辞書は、語に対して定義されるものの、慣用文型を含む豊富な文型パターンに対し、文の意味、本研究でいう文義を分節している点で、思想的には文表層辞書と文概念辞書を一体化したものを部分的に実現していると考えられる。

JICST 対訳辞書[芦崎 94]は、日本語科学技術論文の抄録を翻訳するために利用されている。目的を特化したことにより、必要な記述項目を限定し、語表層辞書の実現を行った事例といえる。

コーパスとして分類される事例として、米国の Penn Treebank と英国の BNC をとりあげる。Penn Treebank[Marcus et al. 93]は、書き言葉、話し言葉の両方を含み、系統的に収集されたブラウンコーパス以外に、新聞、官公庁資料、マニュアル等様々な文章を収集している。付加されたタグは、品詞と構文構造が中心であり、文表層辞書の部分的実現事例といえる。

BNC[BNC 92]は、辞書の編纂、言語学研究、言語処理など幅広い利用を目的として収集されている。1億語の収集目標に対し、90%が書き言葉であり、10%が話し言葉である。書き言葉に関しては現代英国英語の代表的なサンプルが幅広い分野から取り入れられるように配慮されている。コーパス内のテキスト部分では、表題、段落、脚注等を表すのに SGML が使用されている。またソフトウェアで自動生成されるタグも SGML で記述されている。また、テキストの出典にかかわる書誌的事項には、執筆者の情報を始めとして充実した内容が記載されている。これは文章・文書表層辞書の部分的実現事例である。

#### 概念レベル

事例としては辞書として分類されるものとシソーラスとして分類されるものがある。



また、仕様の策定に重点を置き実現の方は実験段階にとどまっているものと、本格的な実現を達成したものがある。

辞書に分類され仕様策定に重点を置いた事例に欧州の Acquilex と Multilex、日本の IPAL 辞書がある。Acquilex[Briscoe et al. 90][Briscoe 91]においては、辞書項目は HPSG 流のグラフユニフィケーションに基づく表現言語で記述されている。タイプシステムと素性構造の意味構造部分に語概念や文概念を扱う仕組みが設けられている。しかし、辞書としては野心的な形式化を進めたために、本格的な実現を進めるのが難しいものになっている。また、そのような形式化が自然言語処理にもたらす効用の議論も十分とはいえない。本研究の情報構造くらいから段階を追うのが妥当であろう。Multilex[Serasset 94]においては、意味情報部分の仕様に関して語概念辞書の仕様の基部分に対応するものが検討されている。あくまでも仕様レベルの議論である。

IPAL 辞書[情報処理振興事業協会 87][情報処理振興事業協会 90]においては、名詞概念の分類に関して基本動詞辞書で 18 種類、基本形容詞辞書で 41 種類の意味素性を用いている。意味素性の数は、記述可能な述語の文型パターン・選択制限パターン（本研究の用語でいうと命題のパターン）の種類を決定する。実際の自然言語処理では、何を目的とするかで区別すべきパターンの精密さは異なるが、例えば日本語の英訳語を確定したいというプロセスへ適用するためには、このレベルの分解能ではまだ不足であることが報告されている。また、自然言語処理に適用するためには、述語の選択制限パターンとともに、フィルターとなる名詞についての語概念の階層関係などの情報が必要である。IPAL 辞書ではこの部分について、名詞辞書の開発も行われている。

自然言語処理、特に機械翻訳のための実用規模の実現をはたしているのが ALT 辞書（機械翻訳システム ALT 用辞書）[池原ほか 93]である。3000 の分類項目によって意味属性、すなわち語概念属性の体系が作られ、単語意味辞書の中で各語（語概念）にこれらの属性が割つけられている。これによって多義性の解消を十分に行える分解能を実現している。構文意味辞書の中に文型という形式で文に対する文概念の扱いが記述されている。

シソーラスとして分類される事例として WordNet、ロジェ・シソーラス、分類語彙表がある。WordNet [Miller et al. 93]は心理言語学の研究のために実現されたもので品詞ごとに意味的な関係が詳しく記述されている。語概念辞書の仕様にほぼそのまま対応する構造をもっている。自然言語処理のためには表層レベルへの対応付けが必要であるが、その作業も計画されている。ロジェ・シソーラス[Lloyd 82]は、人間が作文する際に、適切な表現を選び出すことができるようにするためのものである。したがって類義のものを品詞の区別なく集めるという語概念辞書の階層関係とは観点が異なる。その点を除けば、長い歴史に裏打ちされた蓄積として利用できる。分類語彙表[国立国語研究所 64]は語彙を意味の世界でとらえようという試みである。多義語の扱いが徹底していない面があるが、やはり大いに利用できる蓄積である。

#### 知識レベルとの関連

世界知識への結び付きと言う観点から言語知識の情報構造の妥当性や位置付けを明らかにするために、EDR 電子化辞書[日本電子化辞書研究所 95]、WordNet [Miller et al. 93]、CYC[Guha and Lenat 94]の3つの比較検討を行う。

EDR 電子化辞書、WordNet、CYCの3者ともに、知識（常識）を考える上では、自然言語が最も重要な鍵となるという認識では一致している。ただし、対象とした知識（自然言語自身に関する知識、自然言語が記述するものの知識）のレベルに関しては3者は異なる所に重点を置いている。知識のレベルを表層レベル、概念レベル、知識レベルに分けると3者の重点の置き方は次のようになる。

EDR 電子化辞書：表層レベルと概念レベルを対象にしている。ただし、概念間の意味的關係としては、同義、階層、包含（Entailment）を対象にしている。

WordNet：概念レベルが対象である。意味的關係としては、同義、階層、包含に加えて、反義や全体-部分などを対象にする。

CYC: 知識レベルを対象にする。

これらの知識によって実現を目指す機能には次のようなものが考えられる。

- (a) 文が構文的に正しいか否かを判定する機能（構文処理の機能）。
- (b) 文を構成する各語がどの語義（概念）や深層格に対応しているかを判定する機能（意味処理の機能）。
- (c) 省略の処理、相互参照の処理等によって文章中の不完全な文を文として完結したものに仕上げる機能（文脈処理の機能）。
- (d) 任意の単文で表現される基本命題の真偽を判定する機能。
- (e) 任意の複文で表現される複合命題の真偽を判定する機能。

EDR 電子化辞書においては(a)と(b)の機能の実現を目指したものである。ただし、ともに完了したわけではない。(a)に関しては語彙の拡充や精度向上の努力が今後も続けられねばならない。(b)に関しては継続的なさらなる改良・拡張が必要である。また、全体的に英語にかかわる部分の品質が日本語にかかわる部分に比べ低いと言う事情がある。(c)の実現のためには、WordNet が対象にしたように色々な意味関係による概念の定義が必要である。CYC には、(d)、(e)に対する試みが含まれているが、体系的な世界知識の集積とはなっていない。(a)、(b)、(c)から(d)、(e)への結びつきを明らかにし、(d)、(e)のレベルの知識を実現する方法を明らかにすることによって、世界知識や常識が解明されていくことになる。以上の3つの比較から、この試みのためには言語知識の情報構造が確実な枠組みとなることが明らかであろう。



## 第7章

### 結論

本論文は、筆者が日本電子化辞書研究所（EDR）において推進した電子化辞書プロジェクトを通じて、あるいは、関連させながら行ってきた研究の成果をまとめたものである。以下にその成果を要約し、最後に今後の課題を述べる。

第1章では、言語知識の研究やその実現形態である電子化辞書の研究・開発の背景や状況を述べ、本研究の位置付けを明らかにした。すなわち、言語知識に関する研究分野の現状と取り組むべき課題を説明し、国内外で行われている諸活動を概観し、それらにおける問題点を明らかにし、本研究がいかなる問題点の解決を試みたかを説明した。また、全体として情報そのものに関する研究、すなわち、情報構造に関する研究の重要性を指摘した。

第2章では、言語知識の情報構造の全体像を把握できるようにするためにその全体構造を明らかにした。基本的な要求仕様と実現事例であるEDR電子化辞書との対応付けも明らかにした。

第3章では、記述レベルに関して表層レベルにおける言語知識の情報構造を明らかにした。語表層辞書、文表層辞書、文章表層辞書、文書表層辞書という基本構造に対し、情報の基本単位となる表層辞書項目の基本仕様を明らかにした。実現事例としてEDR電子化辞書の表層レベル対応部分を辞書仕様と数値データの両面から述べた。

第4章では、記述レベルに関して概念レベルにおける言語知識の情報構造を明らかにした。語概念辞書、文概念辞書、文章概念辞書、文書概念辞書という基本構造に対し、概念の定義と情報の基本単位となる概念辞書項目の基本仕様を明らかにした。EDR電

子化辞書の深層レベル対応部分を実現事例として辞書仕様と数値データの両面から述べた。

第5章では、2章、3章、4章で明らかにした情報構造の妥当性を説明した。自然言語処理、世界知識の情報構造、開発工程の3点から検討を行い、それぞれに適切であることを説明した。

第6章では、国内外の関連する代表的な活動事例との比較を行い、本研究の意義、妥当性を明らかにした。比較は表層レベル、概念レベル、知識レベルとの関連という3点で行った。

なお、続く付録においては、言語知識の情報構造の実現状況をより詳細に説明するため、EDR 電子化辞書の仕様内容を例とともに詳述する。

本研究によって得られた情報構造は、①自然言語処理に対しては、問題解決の構造に適切に対応付けられること、さらに、規則ベースと事例ベースのいずれにも柔軟に対応しうること、②知識処理に対しては、世界知識を文書から語に向かって言語知識に対応付ける事によって、世界知識の情報構造のプロトタイプとなっていること、③開発プロセスに対しては、サブ辞書を基本モジュールにして適切な工程が定義できること、これらによってその妥当性が明らかにされ、その有用性が示されたことになる。

本研究の次のステップとしての今後の課題を情報構造の詳細化、作成支援システムの開発、世界知識（常識）ベースの実現の3点に整理する。本研究の情報構造は概念仕様の段階のものである。仕様の詳細化が必要である。ただし、机上にての研究ではなく、現実の言語知識、言語データの開発と併せて進めねばならない。今後必要とされる言語知識・言語データにとって、現在開発・整備されているものではとうてい十分であるとは言えない。適切な方策を講じ、連携を保ちながら努力を持続していかなければならない。

開発プロセスの洗練化のため、言語知識の自動獲得、獲得支援に関する高度な機能の研究・開発が必要である。言語知識・言語データは、大規模で、高精度で、しかも安価に開発できるものでなくてはならない。そのためにも、高度な支援システムの開発は重要である。かかる機能の実現には、最新の自然言語処理技術が必要であり、それ自身と

しても、新しい興味あるテーマを提供してくれるものである。統計的手法をはじめ、大規模情報処理のための理論・方式・アルゴリズムに関し、新しいアプローチが必要である。

世界知識の知識ベース実現に向け本格的な研究・開発を開始せねばならない。特に、長期的なテーマとしては、人間が持つ常識の体系的な解明と常識ベースの研究開発である。人間の知能解明に向け、ハードウェア面からの脳研究に対し、ソフトウェア面からの常識の研究が必要である。本研究の成果をはじめとし、そのための基本的な道具立ては整ってきたと思われる。

いずれにしろ、ダウンサイジングやパーソナル化の潮流に対し、その対極に情報の大規模化の潮流がある。現状としては、まだ、ネットワークの大規模化の段階にさしかかったところで、情報そのものの大規模化はまだ部分的に実感されるだけである。ただし、将来を見込む研究・開発としては、この潮流に対する本格的な取り組みに向け努力が始まろうとしている。著者が 1993 年に企画し、12 月に東京で開催した Knowledge Building and Knowledge Sharing の国際会議[Fuchi and Yokoi 94]は、第 2 回が 1995 年 4 月にオランダで開催され、第 3 回は 1997 年 4 月にワシントンで開催される予定である。



## 付録

### EDR 電子化辞書における実現仕様

言語知識の情報構造の実現事例として、EDR 電子化辞書の辞書レコードの仕様を説明する。これは本文中で述べた情報構造の実現仕様である。この実現仕様は本研究の一環として試みたものであり、ひとつの典型例を与えるものである。

#### 付.1 仕様の記述形式

仕様の記述のために、直観的な理解のし易さに重点を置いた記述形式（記述法）を定める。この記述形式は形式的な厳密さは欠くが、誤解の生じない伝達が可能であるように配慮されている。

記述の基本となる単位はレコード、あるいは、項目と呼ばれるものである。レコード（項目）は、階層化された多数のフィールドを含むことができる。フィールドとそのサブフィールドとの関係は、タグ（フィールドの指定と名前付け）のインデントで表わされる。フィールドの役割や格納される情報内容の記述が、タグの右側に書かれる。これらの記法は、SGML の記法から一部借用したものである。ただし、SGML そのものではないことに注意されたい。

この記述形式を用いて電子化辞書全体の構成、各サブ辞書の構成、辞書レコードの構成などが記述される。記述形式の基本的な文法（便法）は以下のとおりである。

(1) <aaa>

<bbb>

<ccc>

-----

フィールド<aaa>はサブフィールド<bbb>と<ccc>からなる。

(2) <aaa>...

-----

フィールド<aaa>が任意個連続して設定される。<aaa>がサブフィールドを持つ時は、それを含めて設定がなされる。

(3) <aaa>。。。<aaa>

-----

フィールド<aaa>の下に設定されているサブフィールドが組で任意個連続して設定される。

(4) <aaa>

フィールド<aaa>が任意個連続して設定される。サブフィールドを持たない時の便法である。

(5) <aaa> :bbb

フィールド<aaa>の役割や格納される情報内容についての自然言語文等による bbb という説明。

(6) <aaa> bbb

フィールド<aaa>に格納される情報内容が bbb であることを例示する。(2), (3)の場合については、別途便法を設ける。

(7) <aaa> \*<数字>

フィールド<aaa>に格納される情報内容の拡張 BNF による記述が番号<数字>の脚注にあることを示す。脚注内容の記述についても、(1)から(6)までの便法を流用する。拡張 BNF におけるメタ変数の記法とフィールドのタグの記法とが同じとなるが、前者は BNF、後者は SGML から借用したものであることに注意されたい。

\*拡張 BNF について.....

メタ変数を"<" ">"で表し、左辺と右辺との境界を"::="、選択の境界を"|"のメタ記号で表す基本的な BNF に対し、次のような拡張を行ったものである。

(1) "。。。"は直前のメタ変数もしくはメタ変数をグループ化したものの一回以上の繰り返しを表わす。

すなわち、

<AAA> ::= <BBB>。。。。

<AAA> ::= <BBB> | <BBB><AAA>

の二つのメタ文は等価である。

(2) "(^)"は繰り返しのセパレータを表すメタ記号である。

(3) "(^)"は、複数の隣接するメタ変数などをグループ化する際に使用するメタ記号である。グループ化したものは一つのメタ変数と同様に扱うことができる。

## 付.2 全体の構成

EDR 電子化辞書の全体構成を示す。専門用語辞書（情報処理）を除く各辞書の辞書レコードの仕様が次節以降に定義される。

### -----[EDR 電子化辞書の構成]-----

#### <日本語単語辞書>

|                  |                      |
|------------------|----------------------|
| <日本語単語辞書レコード>... | :付.3 参照              |
| <連接テーブル>         | :形態素の接続制約の一覧表、形態素解析用 |

#### <英語単語辞書>

|                 |         |
|-----------------|---------|
| <英語単語辞書レコード>... | :付.4 参照 |
|-----------------|---------|

#### <概念辞書>

|                |         |
|----------------|---------|
| <概念見出し辞書>      |         |
| <概念見出しレコード>... | :付.5 参照 |
| <概念体系辞書>       |         |
| <概念体系レコード>...  | :付.5 参照 |
| <概念記述辞書>       |         |
| <概念記述レコード>...  | :付.5 参照 |

#### <日英対訳辞書>

|                 |         |
|-----------------|---------|
| <日英対訳辞書レコード>... | :付.6 参照 |
|-----------------|---------|

#### <英日対訳辞書>

|                 |         |
|-----------------|---------|
| <英日対訳辞書レコード>... | :付.7 参照 |
|-----------------|---------|

#### <日本語共起辞書>

|                         |                     |
|-------------------------|---------------------|
| <日本語共起辞書レコード>...        | :付.8 参照             |
| <日本語動詞共起パターン副辞書>        | :基本動詞に対するより完備した共起辞書 |
| <日本語動詞共起パターン副辞書レコード>... |                     |

#### <英語共起辞書>

|                 |         |
|-----------------|---------|
| <英語共起辞書レコード>... | :付.9 参照 |
|-----------------|---------|

#### <日本語コーパス>

|                  |          |
|------------------|----------|
| <日本語コーパスレコード>... | :付.10 参照 |
|------------------|----------|

#### <英語コーパス>

|                 |          |
|-----------------|----------|
| <英語コーパスレコード>... | :付.11 参照 |
|-----------------|----------|



<専門用語辞書(情報処理)>

:対応する基本語の辞書と同じ構成である

<日本語専門用語単語辞書(情報処理)>

<英語専門用語単語辞書(情報処理)>

<日英専門用語対訳辞書(情報処理)>

<英日専門用語対訳辞書(情報処理)>

<専門用語概念見出し辞書(情報処理)>

<専門用語概念体系辞書(情報処理)>

### 付.3 日本語単語辞書

日本語単語辞書レコードの仕様を記述する。名詞、動詞、述語句に対する例を付す。実際の日本語単語辞書は日本語単語辞書レコードを<単語見出し>の読み順(五十音順)に並べたものとなっている。日常一般の基本語を対象にする。

| =====【日本語単語辞書レコード】===== |                           |
|-------------------------|---------------------------|
| <レコード番号>                | :レコードタイプと識別番号             |
| <見出し情報>                 |                           |
| <単語見出し>                 | :活用語尾つき見出し表記とその読み         |
| <不変化部-連接属性対>            | :構成語の不変化部とその連接属性の対の列      |
| <かな表記>                  | :不変化部のカタカナによる表記           |
| <発音>                    | :カタカナ表記による発音とアクセント(単一語のみ) |
| <文法情報>                  |                           |
| <品詞>                    | :品詞                       |
| <構文木>                   | :構成語の係受け関係(慣用句のみ)         |
| <活用情報>                  |                           |
| <活用形情報>                 | :活用形(不規則活用語のみ)            |
| <活用型情報>                 | :活用型(規則活用語のみ)             |
| <表層格情報>                 | :表層格(用言, 述語句のみ)           |
| <相情報>                   | :相情報(動詞のみ)                |
| <機能語情報>                 | :機能語情報(機能語のみ)             |
| <意味情報>                  |                           |
| <概念識別子>                 | :概念の同一性を示す番号              |
| <概念見出し>                 | :概念を代表する単語見出し(内容語のみ)      |
| <日本語概念見出し>              | :概念を代表する日本語単語見出し          |
| <英語概念見出し>               | :概念を代表する英語単語見出し           |
| <概念説明>                  | :概念の文章による説明(内容語のみ)        |
| <日本語概念説明>               | :概念の日本語による説明              |
| <英語概念説明>                | :概念の英語による説明               |
| <運用・その他情報>              |                           |
| <用法>                    | :語の用法                     |
| <頻度>                    | :出現頻度                     |
| <管理情報>                  |                           |
| <管理履歴レコード>              | :更新日付等の管理情報               |

| -----【日本語単語辞書レコードの例(名詞)】----- |            |
|-------------------------------|------------|
| <レコード番号>                      | JWD0405240 |
| <見出し情報>                       |            |
| <単語見出し>                       | 夕飯【ユウハン】   |

|              |                 |
|--------------|-----------------|
| <不変化部・連接属性対> | 夕飯(普通名詞,普通名詞)   |
| <かな表記>       | ユウハン            |
| <発音>         | ユーハン            |
| <文法情報>       |                 |
| <品詞>         | 普通名詞            |
| <構文木>        |                 |
| <活用情報>       |                 |
| <活用形情報>      |                 |
| <活用型情報>      |                 |
| <表層格情報>      |                 |
| <相情報>        |                 |
| <機能語情報>      |                 |
| <意味情報>       |                 |
| <概念識別子>      | 3c2374          |
| <概念見出し>      |                 |
| <日本語概念見出し>   | 夕食[ユウシヨク]       |
| <英語概念見出し>    | supper          |
| <概念説明>       |                 |
| <日本語概念説明>    | 夕方とする食事         |
| <英語概念説明>     | an evening meal |
| <運用・その他情報>   |                 |
| <用法>         |                 |
| <頻度>         | 8/8             |
| <管理情報>       |                 |
| <管理履歴レコード>   | DATE=93/3/12    |

.....[日本語単語辞書レコードの例(動詞)].....

|              |               |
|--------------|---------------|
| <レコード番号>     | JWD0371001    |
| <見出し情報>      |               |
| <単語見出し>      | 食べる[タベ・ル]     |
| <不変化部・連接属性対> | 食べ(動詞,一段動詞語幹) |
| <かな表記>       | タベ            |
| <発音>         | タベ'           |
| <文法情報>       |               |
| <品詞>         | 動詞            |
| <構文木>        |               |
| <活用情報>       |               |
| <活用形情報>      |               |
| <活用型情報>      | 一段活用          |
| <表層格情報>      | ガ;ヲ;デ         |



|            |                    |
|------------|--------------------|
| <相情報>      | テイル後接で進行相;テクル後接で移動 |
| <機能語情報>    |                    |
| <意味情報>     |                    |
| <概念識別子>    | 3bc6f0             |
| <概念見出し>    |                    |
| <日本語概念見出し> | 食べる[タベ・ル]          |
| <英語概念見出し>  | eat                |
| <概念説明>     |                    |
| <日本語概念説明>  | 食物をとる              |
| <英語概念説明>   | to eat something   |
| <運用・その他情報> |                    |
| <用法>       |                    |
| <頻度>       | 492/531            |
| <管理情報>     |                    |
| <管理履歴レコード> | DATE=89/2/17       |

-----[日本語単語辞書レコードの例(述語句)]-----

|              |  |
|--------------|--|
| <レコード番号>     | JWD0238660   |
| <見出し情報>      |  |
| <単語見出し>      | 足場を固める[アシバヲカタメ・ル]                                      |
| <不変化部・連接属性対> | 足場(普通名詞,普通名詞)/を(体言後接助詞,係助詞後接格助詞)/[(*,*)]/固め(動詞,一段動詞語幹) |
| <かな表記>       | アシバヲカタメ  |
| <発音>         |  |
| <文法情報>       |  |
| <品詞>         | 述語句  |
| <構文木>        | 連用修飾句(普通名詞(足場)/助詞(を))/[連用修飾句(*)]/<br>述語句(動詞(固め))       |
| <活用情報>       |  |
| <活用形情報>      |  |
| <活用型情報>      | 一段活用   |
| <表層格情報>      | ガ;デ;ノ  |
| <相情報>        | テイル後接で進行相;テイル後接で完了相または結果状態相;テクル後接で移動                   |
| <機能語情報>      |  |
| <意味情報>       |  |
| <概念識別子>      | 1ecf38   |
| <概念見出し>      |  |
| <日本語概念見出し>   | 足場を固める[アシバヲカタメ・ル]                                      |
| <英語概念見出し>    |  |

|            |                     |
|------------|---------------------|
| <概念説明>     |                     |
| <日本語概念説明>  | 物事を行う場合の拠り所を確立する    |
| <英語概念説明>   | to obtain a footing |
| <運用・その他情報> |                     |
| <用法>       |                     |
| <頻度>       | 2/2                 |
| <管理情報>     |                     |
| <管理履歴レコード> | DATE=93/10/15       |

---

## 付.4 英語単語辞書

英語単語辞書レコードの仕様を記述する。名詞、動詞、形容詞、副詞、動詞句、句動詞に対する例を付す。実際の英語単語辞書は英語単語辞書レコードを<単語見出し>表記のアルファベット順に並べたものとなっている。日常一般の基本語を対象にする。

| ===== [英語単語辞書レコード] ===== |                              |
|--------------------------|------------------------------|
| <レコード番号>                 | :レコードタイプと識別番号                |
| <見出し情報>                  |                              |
| <単語見出し>                  | :規則変化語は原形語尾つき                |
| <不変化部・連接属性対>             | :構成語の不変化部とその連接属性の対の列         |
| <音節区切り>                  | :音節の区切りに'/'を挿入した文字列(単一語のみ)   |
| <発音>                     | :単語見出しのIPAによる発音とアクセント(単一語のみ) |
| <文法情報>                   |                              |
| <品詞>                     | :品詞                          |
| <構文木>                    | :構成語の構文構造(複合語・慣用句のみ)         |
| <語形・語形変化情報>              |                              |
| <語形情報>                   | :語形                          |
| <語形変化情報>                 | :語形変化パターン                    |
| <文法属性>                   | :(動詞・形容詞・副詞・名詞のみ)            |
| <文型情報>                   | :(動詞・形容詞・名詞のみ)               |
| <機能・位置>                  | :(名詞・形容詞・限定詞・副詞のみ)           |
| <機能語情報>                  | :機能語情報(機能語のみ)                |
| <意味情報>                   |                              |
| <概念識別子>                  | :概念の同一性を示す番号                 |
| <概念見出し>                  | :概念を代表する単語見出し(内容語のみ)         |
| <日本語概念見出し>               | :概念を代表する日本語単語見出し             |
| <英語概念見出し>                | :概念を代表する英語単語見出し              |
| <概念説明>                   | :概念の文章による説明(内容語のみ)           |
| <日本語概念説明>                | :概念の日本語による説明                 |
| <英語概念説明>                 | :概念の英語による説明                  |
| <運用・その他情報>               |                              |
| <用法>                     | :語の用法                        |
| <頻度>                     | :出現頻度                        |
| <管理情報>                   |                              |
| <管理履歴レコード>               | :更新日付等の管理情報                  |
| =====                    |                              |

----- [英語単語辞書レコードの例(名詞)] -----

<レコード番号> EWD1116992



|              |   |
|--------------|---|
| <見出し情報>      |   |
| <単語見出し>      | dictionary  |
| <不変化部・連接属性対> | dictionar(語頭子音名詞,名詞不変化部分(後接 y/ies))   |
| <音節区切り>      | dic/tion/ar/y   |
| <発音>         | :IPA による表記  |
| <文法情報>       |   |
| <品詞>         | 普通名詞  |
| <構文木>        |   |
| <語形・語形変化情報>  |   |
| <語形情報>       | 名詞不変化部分   |
| <語形変化情報>     | y 変化型   |
| <文法属性>       | 可算;中性名詞;必ず冠詞をとる   |
| <文型情報>       |   |
| <機能・位置>      |   |
| <機能語情報>      |   |
| <意味情報>       |   |
| <概念識別子>      | 0f6f4b  |
| <概念見出し>      |   |
| <日本語概念見出し>   | 辞書 [ジショ]  |
| <英語概念見出し>    | dictionary  |
| <概念説明>       |   |
| <日本語概念説明>    | 言葉を類別して集めた書物  |
| <英語概念説明>     | a reference book providing a list of words in some<br>stated order, with their forms, pronunciations,<br>meaning and uses |
| <運用・その他情報>   |   |
| <用法>         |   |
| <頻度>         | 15/73   |
| <管理情報>       |   |
| <管理履歴レコード>   | 2/15/95   |

-----[英語単語辞書レコードの例(動詞)]-----

|              |                            |            |
|--------------|----------------------------|------------|
| <レコード番号>     |                            | EWD1364642 |
| <見出し情報>      |                            |            |
| <単語見出し>      | supply                     |            |
| <不変化部・連接属性対> | suppl(語頭子音動詞,y 変化型動詞不変化部分) |            |
| <音節区切り>      | sup/ply                    |            |
| <発音>         | :IPA による表記                 |            |
| <文法情報>       |                            |            |
| <品詞>         | 動詞                         |            |

|             |   |
|-------------|---|
| <構文木>       |   |
| <語形・語形変化情報> |   |
| <語形情報>      | 動詞不変化部分                                   |
| <語形変化情報>    | y 変化型                                     |
| <文法属性>      |   |
| <文型情報>      | 直接目的語 (DO) を必ずとる; DO=名詞句; 特定前置詞句 (to-)をとる |
| <機能・位置>     |   |
| <機能語情報>     |   |
| <意味情報>      |   |
| <概念識別子>     | 0ec944                                    |
| <概念見出し>     |   |
| <日本語概念見出し>  | 支給する [シキユウ・スル]                            |
| <英語概念見出し>   | supply                                    |
| <概念説明>      |   |
| <日本語概念説明>   | 物をあてがう                                    |
| <英語概念説明>    | to supply goods                           |
| <運用・その他情報>  |   |
| <用法>        |   |
| <頻度>        | 122/234                                   |
| <管理情報>      |   |
| <管理履歴レコード>  | 3/4/93                                    |

-----[英語単語辞書レコードの例(形容詞)]-----

|              |                           |
|--------------|---------------------------|
| <レコード番号>     | EWD1334265                |
| <見出し情報>      |                           |
| <単語見出し>      | short                     |
| <不変化部・連接属性対> | short(語頭子音形容詞, er 変化型形容詞) |
| <音節区切り>      | short                     |
| <発音>         | :IPA による表記                |
| <文法情報>       |                           |
| <品詞>         | 形容詞                       |
| <構文木>        |                           |
| <語形・語形変化情報>  |                           |
| <語形情報>       | 形容詞原級                     |
| <語形変化情報>     | er 変化型形容詞                 |
| <文法属性>       |                           |
| <文型情報>       | 特定前置詞句(of-)をとる            |
| <機能・位置>      |                           |
| <機能語情報>      |                           |

|            |              |
|------------|--------------|
| <意味情報>     |              |
| <概念識別子>    | 10096e       |
| <概念見出し>    |              |
| <日本語概念見出し> | 乏しい [トボシ・イ]  |
| <英語概念見出し>  | scarce       |
| <概念説明>     |              |
| <日本語概念説明>  | 十分でないさま      |
| <英語概念説明>   | insufficient |
| <運用・その他情報> |              |
| <用法>       |              |
| <頻度>       | 45/431       |
| <管理情報>     |              |
| <管理履歴レコード> | 2/13/89      |

---

|                              |  |
|------------------------------|--|
| -----[英語単語辞書レコードの例(副詞)]----- |  |
| <レコード番号>                     | EWD1345229   |
| <見出し情報>                      |  |
| <単語見出し>                      | soon   |
| <不変化部・連接属性対>                 | soon(語頭子音副詞,er 変化型副詞)                                      |
| <音節区切り>                      | soon   |
| <発音>                         | :IPA による表記   |
| <文法情報>                       |  |
| <品詞>                         | 副詞   |
| <構文木>                        |  |
| <語形・語形変化情報>                  |  |
| <語形情報>                       | 副詞原級   |
| <語形変化情報>                     | er 変化型副詞   |
| <文法属性>                       |  |
| <文型情報>                       |  |
| <機能・位置>                      | 中位 (本動詞の直前、または be 動詞の直後) ;後位 (目的語・補語の後、なければ動詞の直後) ;動詞を修飾する |
| <機能語情報>                      |  |
| <意味情報>                       |  |
| <概念識別子>                      | 0ea98d   |
| <概念見出し>                      |  |
| <日本語概念見出し>                   | 近々 [チカチカ]  |
| <英語概念見出し>                    | soon   |
| <概念説明>                       |  |
| <日本語概念説明>                    | やがて  |
| <英語概念説明>                     | of a time period, occurring in the near future             |



<運用・その他情報>

<用法>

<頻度>

209/892

<管理情報>

<管理履歴レコード>

1/21/92

-----[英語単語辞書レコードの例(動詞句)]-----

<レコード番号>

EWD1295652

<見出し情報>

<単語見出し>

put \*one's confidence in \*someone

<不変化部-連接属性対>

put(語頭子音動詞,s-不規則変化型動詞不変化部分)

/\*suf(\*,\*)/(空白,空白)/\*one's(代名詞,人称代名詞所有格)/

(空白,空白)/confidence(語頭子音名詞,名詞単数形)/(空白,

空白)/in(不定冠詞非前接語,前置詞)/(空白,空白)

/\*someone(\*,\*)

<音節区切り>

<発音>

<文法情報>

<品詞>

動詞句

<構文木>

動詞(動詞;動詞原形;s-不規則変化型;主要語(put))[動詞語

尾(\*suf)]/記号( )/普通名詞;直接目的語;不可算(人称代名

詞;所有格(\*one's)/記号( )/普通名詞;名詞単数形;必ず単数

形で;不可算(confidence)/記号( )/前置詞句(前置詞(in)/記

号( )/名詞句(\*someone))

<語形・語形変化情報>

<語形情報>

<語形変化情報>

<文法属性>

<文型情報>

<機能・位置>

<機能語情報>

<意味情報>

<概念識別子>

271f65

<概念見出し>

<日本語概念見出し>

<英語概念見出し>

place \$one's¥ confidence in

<概念説明>

<日本語概念説明>

<英語概念説明>

to trust (a person)

<運用・その他情報>

<用法>

<頻度> 0/0  
 <管理情報>  
 <管理履歴レコード> 7/9/94

---

-----[英語単語辞書レコードの例(句動詞)]-----

<レコード番号> EWD1097635  
 <見出し情報>  
 <単語見出し> cool down  
 <不変化部-連接属性対> cool(語頭子音動詞,s-ed 変化型動詞不変化部分)/<sup>\*</sup>su<sup>f</sup>(<sup>\*</sup>,<sup>\*</sup>)/  
 (空白,空白)/down(不定冠詞非前接語,副詞小辞)  
 <音節区切り>  
 <発音>  
 <文法情報>  
 <品詞> 動詞  
 <構文木> 動詞(動詞;動詞原形;s-ed 変化型動詞;主要語(cool)/[動詞語  
 尾(<sup>\*</sup>su<sup>f</sup>)]/記号( )/副詞小辞(down)  
 <語形・語形変化情報>  
 <語形情報>  
 <語形変化情報>  
 <文法属性>  
 <文型情報> 自動詞  
 <機能・位置>  
 <機能語情報>  
 <意味情報>  
 <概念識別子> 266544  
 <概念見出し>  
 <日本語概念見出し>  
 <英語概念見出し> cool down  
 <概念説明>  
 <日本語概念説明>  
 <英語概念説明> to become cool  
 <運用・その他情報>  
 <用法>  
 <頻度> 0/0  
 <管理情報>  
 <管理履歴レコード> 7/9/94

---

## 付.5 概念辞書

概念辞書は、概念見出し辞書、概念体系辞書、概念記述辞書からなる。概念見出しレコードの定義、概念体系レコードの定義、概念体系の主要部、概念記述レコードの定義、関係子一覧が記述される。

### 概念見出しレコード

| ===== [概念見出しレコード] ===== |                       |
|-------------------------|-----------------------|
| <レコード番号>                | : レコードタイプと識別番号        |
| <概念識別子>                 | : 概念を一意的に特定する 16 進整数  |
| <概念見出し>                 | : 概念の意味内容に近い意味を表現する単語 |
| <英語概念見出し>               | : (英語単語)              |
| <日本語概念見出し>              | : (日本語単語)             |
| <概念説明>                  | : 概念の意味内容を表現する説明文     |
| <英語概念説明>                | : (英文説明)              |
| <日本語概念説明>               | : (日本文説明)             |
| <管理情報>                  | : 更新日付などの管理履歴レコードの列   |
| <管理履歴レコード>              | : 文字列=文字列 の形の属性・属性値ペア |
| =====                   |                       |

| ----- [概念見出しレコードの例] ----- |   |
|---------------------------|---|
| <レコード番号>                  | CPH0314159  |
| <概念識別子>                   | 3d0ecb  |
| <概念見出し>                   |   |
| <英語概念見出し>                 | borrow  |
| <日本語概念見出し>                | 借りる[カリ・ル]   |
| <概念説明>                    |   |
| <英語概念説明>                  | to use a person's property after promising to ... |
| <日本語概念説明>                 | 返す約束で,他人のものを使う                                    |
| <管理情報>                    |   |
| <管理履歴レコード>                | 更新日付="93/04/26"                                   |
| -----                     |   |

### 概念体系レコード

| ===== [概念体系レコード] ===== |                        |
|------------------------|------------------------|
| <レコード番号>               | : レコードタイプと識別番号         |
| <上位概念識別子>              | : 上位概念を表す 16 進整数の概念識別子 |
| <下位概念識別子>              | : 下位概念を表す 16 進整数の概念識別子 |
| <管理情報>                 | : 更新日付などの管理履歴レコードの列    |



<管理履歴レコード>

: 文字列=文字列 の形の属性・属性値ペア

-----[概念体系レコードの例]-----

<レコード番号> CPC0271828  
<上位概念識別子> 4445bc:「書いた物」を表す概念識別子  
<下位概念識別子> 4445a0:「書簡」を表す概念識別子  
<管理情報>  
    <管理履歴レコード> 更新日付="92/03/05"

## 概念体系の主要部

最上位の概念である「概念」を除く基本語概念体系の上位3レベル目までの項目一覧である。

### 1 人間または人間と似た振る舞いをする主体 3aa911

#### 1-1 人間 30f6b0

- 1-1-1 職業、肩書、役割で限定した人間 444df2
- 1-1-2 行う行為の内容で捉えた人間 44474a
- 1-1-3 状態や評価で捉えた人間 44483c
- 1-1-4 心身の状態で捉えた人間 3aa919
- 1-1-5 パーソナリティで捉えた人間 3f961d
- 1-1-6 数量で捉えた人間 444752
- 1-1-7 ある個体を基準とした関係で捉えた人間 3aae71
- 1-1-8 人間の指示的な呼称 444846
- 1-1-9 人間の具体的な名前 3f961c
- 1-1-10 民族・種族・人種で捉えた人間 3f960d

#### 1-2 動物 30f6bf

- 1-2-1 種で捉えた動物 30f6c1
- 1-2-2 役割で捉えた動物 3aa91a
- 1-2-3 生息している場所によって捉えた動物 444858
- 1-2-4 食べる物の種類で捉えた動物 444737
- 1-2-5 想像上の動物 444509
- 1-2-6 恐竜など古代に生きた動物 444593

#### 1-3 自立活動体 3aa912

- 1-3-1 組織 30f746
- 1-3-2 多くの人々の集まり 3cfacc
- 1-3-3 民族・種族・人種で捉えた人間 3f960d
- 1-3-4 会議 444614

1-3-5 自動物 3aa930

1-4 神, 仏, 霊, 天使など, 抽象的な有意志体 4444b6

1-4-1 神である有意志体 444530

1-4-2 神話や伝説や人々の心の中でのみ存在する, 信仰対象外の擬似人間や擬似生物

1-4-3 仏およびそれに類する仏教上の信仰対象となる疑似人間 444531 4445ea

1-4-4 天使など, 人からプラスイメージで捉えられる疑似人間 444532

1-4-5 化けもの, 悪魔など人からマイナスイメージで捉えられる疑似人間 444533

1-4-6 宇宙人, 雪男など, 想像上の人間 444534

1-4-7 想像上の動物 444509

1-4-8 人の死後, 心が遊離したとみなされる霊魂 444535

1-4-9 人や物や言葉などに宿るとみなされる霊魂 4445b9

2 ものごと 3d017c

2-1 もの 444d86

2-1-1 具体物 30f6ae

2-1-2 自然現象によってできる物 4444e1

2-1-3 抽象物 444e40

2-1-4 具体的あるいは抽象的存在物 444e3f

2-1-5 この世の中にある有形無形のすべてのもの 0e5c6d

2-2 事柄 444ab5

2-2-1 出来事 30f800

2-2-2 社会や政治の状態を変化させようとする動き 444554

2-2-3 催し 444549

2-2-4 評価でとらえた事柄 444daf

2-2-5 遊びやスポーツ 444d85

2-2-6 政治 444718

2-2-7 経済 10dd19

2-2-8 軍事に関するすべての事 10ba75

2-2-9 組織内の個人に関する事柄 0f837b

2-3 識別名 444daa

2-4 客観的な対象 0e7faa

3 事象 30f7e4

3-1 現象 30f7e5

3-1-1 自然現象 30f7e6

3-1-2 静物に関する現象 30f7ea

3-1-3 生理現象 30f7f4

3-1-4 社会現象 30f7ff

3-1-5 人間に関する現象 3aa947

3-2 行為 30f83e

3-2-1 自身行為 444d96

3-2-2 対象行為 444dd8

3-2-3 身体的活動 30f83f

3-2-4 感情活動 30f863

3-3 移動 30f801

3-3-1 空間移動 30f802

3-3-2 所有権の移動 30f826

3-3-3 情報の移動 30f832

3-4 変化 3f9856

3-5 状態 3aa963

3-5-1 性状・性向 3f9871

3-5-2 物事に対する評価 30f7c8

3-5-3 関係 30f9a4

3-5-4 存在状態 3f98f6

4 位置 30f751

4-1 場所 3aa938

4-1-1 機能で捉えた場所 30f749

4-1-2 形でとらえた場所 4449d9

4-1-3 評価で捉えた場所 444a86

4-1-4 状態で捉えた場所 30f75c

4-1-5 物事が存在する空間 3d002b

4-1-6 抽象的場所 444aa1

4-1-7 地域 3f965a

4-1-8 天体や天球に便宜上設定された架空の線や区域 4449cf

4-1-9 天体 30f742

4-1-10 大空 0ff656

4-1-11 建造物 3aa932

4-1-12 区域施設 3f9635

4-2 複数のものの関係によって決まる位置 30f753

4-2-1 基準となる静物との関係で捉えた場所 3f9655

4-2-2 境 30f754

4-2-3 間 3f9652

4-2-4 指示位置 30f752



4-3 領域 30f767

4-3-1 ある定まった広がり 1e85eb

4-3-2 抽象的領域 444aa0

4-3-3 空間 30f773

4-3-4 スポーツにおける限定された活動領域 444b2a

4-4 部分 3f9651

4-4-1 部分位置 444d63

4-4-2 抽象的部分 444aa5

4-4-3 数量的分割部分 444ble

4-4-4 特定の事柄の部分や側面 444b21

4-5 方向 3f9658

4-5-1 移動方向 3f9656

4-5-2 東西南北 30f75b

4-6 抽象的位置 444a9d

5 時 30f776

5-1 時間点 3f9882

5-1-1 時期 4449df

5-1-2 時代 30f77a

5-1-3 年 4449e4

5-1-4 1年の流れの中の時期 444d6b

5-1-5 1日の流れの中の時 444d68

5-1-6 あることをするのにちょうどよい時機 3ceb7d

5-1-7 期限 30f778

5-1-8 瞬間 444d67

5-1-9 基準時点の前後の時 444da3

5-1-10 期間内の相対的な時間位置 3f988b

5-2 周期的に訪れる時間 444dd2

5-3 過去から現在、未来までの時間の流れの中で捉えた時間 444dd3

5-4 時間 30f77b

5-5 単位で長さを示した時間 444dd4

5-6 周期 4449e2

5-7 経過・歴史 30f7d6

## 概念記述レコード

| ===== [概念記述レコード] ===== |                              |
|------------------------|------------------------------|
| <レコード番号>               | : レコードタイプと識別番号               |
| <記述情報>                 |                              |
| <記述区分>                 | : E 記述/I 記述の別                |
| <記述タブ>                 | : (Cf, R, Ct, P) で表されるタブ     |
| <概念識別子 1>              | : From 概念 Cf を表す概念識別子        |
| <関係子>                  | : 2つの概念の間に成立する関係子 R          |
| <概念識別子 2>              | : To 概念 Ct を表す概念識別子          |
| <真偽値>                  | : 真偽値を表す P の値 (真:P=1, 偽:P=0) |
| <管理情報>                 | : 更新日付などの管理履歴レコードの列          |
| <管理履歴レコード>             | : 文字列=文字列 の形の属性・属性値ペア        |
| =====                  |                              |

| ----- [概念記述レコードの例 (1)] ----- |                       |
|------------------------------|-----------------------|
| <レコード番号>                     | CPT0577216            |
| <記述情報>                       |                       |
| <記述区分>                       | E                     |
| <記述タブ>                       |                       |
| <概念識別子 1>                    | 3d0ecb: 「借りる」を表す概念識別子 |
| <関係子>                        | object                |
| <概念識別子 2>                    | 0e5097: 「本」を表す概念識別子   |
| <真偽値>                        | 1                     |
| <管理情報>                       |                       |
| <管理履歴レコード>                   | 更新日付="92/05/10"       |
| -----                        |                       |

| ----- [概念記述レコードの例 (2)] ----- |                          |
|------------------------------|--------------------------|
| <レコード番号>                     | CPT0123456               |
| <記述情報>                       |                          |
| <記述区分>                       | I                        |
| <記述タブ>                       |                          |
| <概念識別子 1>                    | 30f826: 「所有権の移動」を表す概念識別子 |
| <関係子>                        | object                   |
| <概念識別子 2>                    | 30f6ae: 「静物」を表す概念識別子     |
| <真偽値>                        | 1                        |
| <管理情報>                       |                          |
| <管理履歴レコード>                   | 更新日付="92/05/10"          |
| -----                        |                          |

## 関係子一覧

概念記述レコード、および、共起辞書レコードとコーパスレコードの意味情報の表現に用いられる関係子の一覧である。コーパスレコードの意味情報の表現には、これらの関係子を核にして他の関係や属性を表現するための記述子が多数用いられている。これについては、文献[国際情報化センター93]を参照されたい。以下に、関係子名とその説明(例を含む)の対を列挙する。これらの関係子は、いわゆる深層格等に対応するものばかりではなく、名詞概念間や事象概念間の関係など事実や事象を表現するのに必要な基本的なものをすべて含む。

|           |   |
|-----------|---|
| agent     | <p>有意志動作を引き起こす主体</p> <p>例：父が食べる</p> <p>[ [main 3:食べ:3bc6f0]<br/>[agent 1:父:0e7c00]]</p>                               |
| object    | <p>動作・変化の影響を受ける対象</p> <p>例：りんごを食べる</p> <p>[ [main 3:食べ:3bc6f0]<br/>[object 1:りんご:3bd8db]]</p>                         |
| a-object  | <p>属性をもつ対象</p> <p>例：トマトが赤い</p> <p>[ [main 3:赤:0e29cb]<br/>[a-object 1:トマト:3bc118]]</p>                                |
| implement | <p>有意志動作における道具・手段</p> <p>例：ナイフで切る</p> <p>[ [main 3:切:0ecff7]<br/>[implement 1:ナイフ:3c4e7d]]</p>                        |
| material  | <p>材料または構成要素</p> <p>例：牛乳からバターを作る</p> <p>[ [main 5:作:0fe812]<br/>[object 3:バター:3be1c7]<br/>[material 1:牛乳:3c03b7]]</p> |
| source    | <p>事象の主体または対象の最初の位置</p> <p>例：京都から来る</p> <p>[ [main 3:来る:3d144c]<br/>[source 1:京都:0ecb69]]</p>                         |
| goal      | <p>事象の主体または対象の最後の位置</p> <p>例：東京に行く</p> <p>[ [main 3:行:1e84a2]<br/>[goal 1:東京:0fee1]]</p>                              |
| place     | <p>事象の成立する場所</p> <p>例：部屋で遊ぶ</p> <p>[ [main 3:遊:3cf67f]]</p>   |



|           |   |
|-----------|---|
| scene     | <p>[place 1:部屋:1080e6]]</p> <p>事象の成立する場面</p> <p>例：ドラマで演じる</p> <p>[ [main 3:演じ:3cf94e]</p> <p>[scene 1:ドラマ:1013ed]]</p>  |
| basis     | <p>比較の基準</p> <p>例：バラはチューリップより美しい</p> <p>[ [main @1::c#more]</p> <p>[object [ [main 5:美し:1e84c3]</p> <p>[a-object 1:バラ:0f6013]]]</p> <p>[basis [ [main @2:美し:1e84c3]</p> <p>[a-object 3:チューリップ:3c2801]]]</p> |
| manner    | <p>動作・変化のやり方</p> <p>例1：ゆっくり話す</p> <p>[ [main 2:話:3ce6b9]</p> <p>[manner 1:ゆっくり:0f81ac]]</p> <p>例2：3時間見る</p> <p>[ [main 3:見:1e8643]</p> <p>[manner [ [main 2:時間:0f6fe4]</p> <p>[number 1:3:"=N 3"]]]]</p>    |
| time      | <p>事象の起こる時間</p> <p>例：8時に起きる</p> <p>[ [main 4:起き:3cfbd5]</p> <p>[time [ [main 2:時:0f6f06]</p> <p>[modifier 1:8:"=N 8"]]]]</p>  |
| time-from | <p>事象の始まる時間</p> <p>例：9時から働く</p> <p>[ [main 4:働:0e2799]</p> <p>[time-from [ [main 2:時:0f6f06]</p> <p>[modifier 1:9:"=N 9"]]]]</p>  |
| time-to   | <p>事象の終わる時間</p> <p>例：9時まで働く</p> <p>[ [main 4:働:0e2799]</p> <p>[time-to [ [main 2:時:0f6f06]</p> <p>[modifier 1:9:"=N 9"]]]]</p>  |
| quantity  | <p>物・動作・変化の量</p> <p>例1：3kgのりんご</p> <p>[ [main 4:りんご:3bd8db]</p> <p>[quantity [ [main 2:k g:3c0285]</p> <p>[number 1:3:"=N 3"]]]]</p> <p>例2：3kg痩せる</p> <p>[ [main 3:痩せ:3c049e]</p>                           |

|           |  |
|-----------|--|
|           | [quantity [ [main 2:k g:3c0285]<br>[number 1:3:"=N 3"]]]   |
| modifier  | 修飾関係<br>例1:机の上の本<br>[ [main 5:本:0e5097]<br>[modifier [ [main 3:上:0e5797]<br>[modifier 1:机:3d05cf]]]]   |
| number    | 数<br>例:3 k g<br>[ [main 2:k g:3c0285]<br>[number 1:3:"=N 3"]]  |
| and       | 概念間の連結関係<br>例1:ローマとナポリに行く<br>[ [main 5:行:1e84a2]<br>[goal [ [main 3:ナポリ:1efc5a]<br>[and 1:ローマ:10e979]<br>[attribute focus]]]]]<br>例2:山は美しく、水は澄んでいる<br>[ [main [ [main 8:澄:0f8f10]<br>[a-object 6:水:3bd634]]]<br>[and [ [main 3:美し:1e84c3]<br>[a-object 1:山:3ce994]]]]] |
| or        | 概念間の選択関係<br>例1:ローマかナポリに行く<br>[ [main 5:行:1e84a2]<br>[goal [ [main 3:ナポリ:1efc5a]<br>[or 1:ローマ:10e979]<br>[attribute focus]]]]]<br>例2:学校に行くか図書館に行く<br>[ [main [ [main 8:行:0f8f10]<br>[goal 6:図書館:100648]]]<br>[or [ [main 3:行:0f8f10]<br>[goal 1:学校:3cf8b1]]]]]          |
| condition | 事象・事実の条件関係<br>例:雨が降ったので家に帰った<br>[ [main [ [main 9:帰:0e8e45]<br>[goal 7:家:0e5cdb]]]<br>[condition 1:雨:3bba1f]]]   |
| purpose   | 目的<br>例:映画を見に行く<br>[ [main 5:行:1e84a2]<br>[purpose [ [main 3:見:1e8646]   |

|              |  |
|--------------|--|
| cooccurrence | <p>[object 1:映画:3be65c]]]]</p> <p>事象・事実の同時関係</p> <p>例：家に帰る間中泣いていた</p> <p>[ [main 7:泣:0f4cf1]</p> <p>[cooccurrence [ [main 3:帰:0e8e45]</p> <p>[goal 1:家:0e5cdb]]]]</p>    |
| sequence     | <p>事象・事実の時間的前後関係</p> <p>例：図書館へ行って本を借りた</p> <p>[ [main 8:借り:0e97a9]</p> <p>[object 6:本:0e5097]</p> <p>[sequence [ [main 3:行:0f8f10]</p> <p>[goal 1:図書館:100648]]]]</p>     |
| possessor    | <p>所有関係</p> <p>例：父の本</p> <p>[ [main 3:本:0e5097]</p> <p>[possessor 1:父:0e7c00]]</p>   |
| beneficiary  | <p>利益・不利益の移動先 [受益[者]と被害[者]の両方を含む]</p> <p>例：父に買ってあげる</p> <p>[ [main 3:買:1e84f1]</p> <p>[beneficiary 1:父:0e7c00]]</p>  |
| unit         | <p>単位</p> <p>例：1 ダース当り 5 0 0 円</p> <p>[ [main 5:円:0e6912]</p> <p>[number 4:5 0 0 : "= N 5 0 0"]</p> <p>[unit [ [main 2:ダース:3bf083]</p> <p>[number 1:1 : "= N 1 "]]]]</p> |
| from-to      | <p>範囲</p> <p>例：大阪から東京までの都市</p> <p>[ [main 6:都市:3cfc38]</p> <p>[modifier [ [main 3:東京:0ffee3]</p> <p>[from-to 1:大阪:0e7107]]]]</p>   |



## 付.6 日英対訳辞書

日英対訳辞書レコードの様態を記述する。名詞と動詞に対する例を付す。実際の日英対訳辞書は日英対訳辞書レコードを<単語見出し>の読み順(五十音順)に並べたものとなっている。<単語見出し>と<意味情報>の組みによって仕様上は日本語単語辞書レコードに対応付けられる。

| ===== [日英対訳辞書レコード] ===== |                      |
|--------------------------|----------------------|
| <レコード番号>                 | :レコードタイプと識別番号        |
| <見出し情報>                  |                      |
| <単語見出し>                  | :活用語尾つき見出し表記とその読み    |
| <文法情報>                   |                      |
| <品詞>                     | :品詞                  |
| <意味情報>                   |                      |
| <概念識別子>                  | :概念の同一性を示す番号         |
| <概念見出し>                  | :概念を代表する単語見出し(内容語のみ) |
| <日本語概念見出し>               | :概念を代表する日本語単語見出し     |
| <英語概念見出し>                | :概念を代表する英語単語見出し      |
| <概念説明>                   | :概念の文章による説明(内容語のみ)   |
| <日本語概念説明>                |                      |
| <英語概念説明>                 |                      |
| <対訳情報>                   |                      |
| <訳語情報>。。。                | :訳語の表記と文法的情報         |
| <訳語種別>                   | :訳語の種類               |
| <訳語表記>                   | :訳語の表記               |
| <訳語品詞>                   | :訳語の品詞               |
| <管理情報>                   |                      |
| <管理履歴レコード>               | :更新日付等の管理情報          |
| =====                    |                      |

| ----- [日英対訳辞書レコードの例(名詞)] ----- |            |
|--------------------------------|------------|
| <レコード番号>                       | JEB0185228 |
| <見出し情報>                        |            |
| <単語見出し>                        |            |
| <漢字見出し>                        | 弁天         |
| <かな見出し>                        | ペンテン       |
| <文法情報>                         |            |
| <品詞>                           | 普通名詞       |
| <意味情報>                         |            |
| <概念識別子>                        | 1084e3     |
| <概念見出し>                        |            |

|            |   |
|------------|---|
| <日本語概念見出し> | 弁天[ベンテン]  |
| <英語概念見出し>  |   |
| <概念説明>     |   |
| <日本語概念説明>  | 禪寺の住職の妻   |
| <英語概念説明>   | wife of the chief priest of a Zen Buddhist temple |
| <対訳情報>     |   |
| <訳語情報>     |   |
| <訳語種別>     | 0   |
| <訳語表記>     | wife of chief zen-priest                          |
| <訳語品詞>     | 名詞及び名詞句、名詞節                                       |
| <管理情報>     |   |
| <管理履歴レコード> | DATE="95/2/15"                                    |

---

-----[日英対訳辞書レコードの例(動詞)]-----

|            |  |
|------------|--|
| <レコード番号>   | JEB0368581                                   |
| <見出し情報>    |  |
| <単語見出し>    |  |
| <漢字見出し>    | さえずる   |
| <かな見出し>    | サエズ・ル  |
| <文法情報>     |  |
| <品詞>       | 動詞   |
| <意味情報>     |  |
| <概念識別子>    | 3bbd74                                       |
| <概念見出し>    |  |
| <日本語概念見出し> | さえずる[サエズ・ル]                                  |
| <英語概念見出し>  |  |
| <概念説明>     |  |
| <日本語概念説明>  | 舞楽において、朗詠する                                  |
| <英語概念説明>   | in Japanese court dance and music, to recite |
| <対訳情報>     |  |
| <訳語情報>     |  |
| <訳語種別>     | 0  |
| <訳語表記>     | recite <bunraku>                             |
| <訳語品詞>     | 動詞及び動詞句                                      |
| <管理情報>     |  |
| <管理履歴レコード> | DATE="95/2/15"                               |

---

-----[日英対訳辞書レコードの例(名詞)]-----

|            |  |
|------------|--|
| <レコード番号>   | JEB0393608   |
| <見出し情報>    |  |
| <単語見出し>    |  |
| <漢字見出し>    | 狂言   |
| <かな見出し>    | キョウゲン  |
| <文法情報>     |  |
| <品詞>       | 普通名詞   |
| <意味情報>     |  |
| <概念識別子>    | 3c0380   |
| <概念見出し>    |  |
| <日本語概念見出し> | 狂言[キョウゲン]  |
| <英語概念見出し>  | kyogen   |
| <概念説明>     |  |
| <日本語概念説明>  | 能楽の合い間に演ずる喜劇   |
| <英語概念説明>   | a comedy which is played during the intermission<br>of a Japanese Noh play |
| <対訳情報>     |  |
| <訳語情報>     |  |
| <訳語種別>     | 1  |
| <訳語表記>     | comic interlude  |
| <訳語品詞>     | 名詞及び名詞句、名詞節  |
| <訳語種別>     | 1  |
| <訳語表記>     | Noh comedy   |
| <訳語品詞>     | 名詞及び名詞句、名詞節  |
| <訳語種別>     | 4  |
| <訳語表記>     | farce presented between Noh plays  |
| <訳語品詞>     | 名詞及び名詞句、名詞節  |
| <訳語種別>     | 3  |
| <訳語表記>     | kyo-gen  |
| <訳語品詞>     | 名詞及び名詞句、名詞節  |
| <管理情報>     |  |
| <管理履歴レコード> | DATE="95/2/15"   |

---



## 付.7 英日対訳辞書

英日対訳辞書レコードの様子を記述する。名詞、動詞、形容詞に対する例を付す。実際の英日対訳辞書は英日対訳辞書レコードを<単語見出し>のアルファベット順に並べたものとなっている。<単語見出し>と<意味情報>の組みによって仕様上は英語単語辞書レコードに対応付けられる。

| ===== [英日対訳辞書レコード] ===== |                           |
|--------------------------|---------------------------|
| <レコード番号>                 | :レコードタイプと識別番号             |
| <見出し情報>                  |                           |
| <単語見出し>                  | :規則変化語は原形語尾つき             |
| <文法情報>                   |                           |
| <品詞>                     | :品詞                       |
| <意味情報>                   |                           |
| <概念識別子>                  | :概念の同一性を示す番号              |
| <概念見出し>                  | :概念を代表する単語見出し(内容語のみ)      |
| <英語概念見出し>                | :概念を代表する英語単語見出し           |
| <日本語概念見出し>               | :概念を代表する日本語単語見出し          |
| <概念説明>                   | :概念の文章による説明(内容語のみ)        |
| <英語概念説明>                 | :概念の英語による説明               |
| <日本語概念説明>                | :概念の日本語による説明              |
| <対訳情報>                   |                           |
| <訳語情報>。。                 |                           |
| <訳語種別>                   | :訳語の種類を表す。英日対訳辞書では「同義語」のみ |
| <訳語表記>                   | :英語の見出し語に対する日本語の訳語        |
| <管理情報>                   |                           |
| <管理履歴レコード>               | :更新日付等の管理情報               |
| =====                    |                           |

| ----- [英日対訳辞書レコードの例(名詞)] ----- |            |
|--------------------------------|------------|
| <レコード番号>                       | EJB1083615 |
| <見出し情報>                        |            |
| <単語見出し>                        | claim      |
| <文法情報>                         |            |
| <品詞>                           | 普通名詞       |
| <意味情報>                         |            |
| <概念識別子>                        | 3d01c7     |
| <概念見出し>                        |            |
| <英語概念見出し>                      | claim      |
| <日本語概念見出し>                     | 権利[ケンリ]    |

|            |   |
|------------|---|
| <概念説明>     |   |
| <英語概念説明>   | a lawful power which enables a person to claim his profit |
| <日本語概念説明>  | 法律上で、特定の利益を主張しうる力   |
| <対訳情報>     |   |
| <訳語情報>     |   |
| <訳語種別>     | 0   |
| <訳語表記>     | (当然の)権利   |
| <訳語種別>     | 0   |
| <訳語表記>     | (当然の)資格   |
| <訳語種別>     | 0   |
| <訳語表記>     | 要求権   |
| <管理情報>     |   |
| <管理履歴レコード> | DATE="95/3/10"  |

.....[英日対訳辞書レコードの例(動詞)].....

|            |                            |
|------------|----------------------------|
| <レコード番号>   | EJB1054678                 |
| <見出し情報>    |                            |
| <単語見出し>    | belabor                    |
| <文法情報>     |                            |
| <品詞>       | 動詞                         |
| <意味情報>     |                            |
| <概念識別子>    | 3cecd7                     |
| <概念見出し>    |                            |
| <英語概念見出し>  | blister                    |
| <日本語概念見出し> | 痛罵する[ツウバ・スル]               |
| <概念説明>     |                            |
| <英語概念説明>   | to attack with sharp words |
| <日本語概念説明>  | ひどくののしる                    |
| <対訳情報>     |                            |
| <訳語情報>     |                            |
| <訳語種別>     | 0                          |
| <訳語表記>     | (言葉で)しつこく攻撃する              |
| <訳語種別>     | 0                          |
| <訳語表記>     | (しつこく)攻めたてる                |
| <訳語種別>     | 0                          |

|            |                |
|------------|----------------|
| <訳語表記>     | ののしる           |
| <管理情報>     |                |
| <管理履歴レコード> | DATE="95/3/10" |

-----[英日対訳辞書レコードの例(形容詞)]-----

|            |                                   |
|------------|-----------------------------------|
| <レコード番号>   | EJB1095361                        |
| <見出し情報>    |                                   |
| <単語見出し>    | constant                          |
| <文法情報>     |                                   |
| <品詞>       | 形容詞                               |
| <意味情報>     |                                   |
| <概念識別子>    | 3d0289                            |
| <概念見出し>    |                                   |
| <英語概念見出し>  | constant                          |
| <日本語概念見出し> | 一定する[イッテイ・スル]                     |
| <概念説明>     |                                   |
| <英語概念説明>   | the condition of being unchanging |
| <日本語概念説明>  | 一定して変わりのないこと                      |
| <対訳情報>     |                                   |
| <訳語情報>     |                                   |
| <訳語種別>     | 0                                 |
| <訳語表記>     | 不変の                               |
| <訳語種別>     | 0                                 |
| <訳語表記>     | 一定の                               |
| <訳語種別>     | 0                                 |
| <訳語表記>     | 安定した                              |
| <訳語種別>     | 0                                 |
| <訳語表記>     | 一様の                               |
| <管理情報>     |                                   |
| <管理履歴レコード> | DATE="95/3/10"                    |



## 付.8 日本語共起辞書

日本語共起辞書レコードの仕様を記述する。を格を伴う述語句と連体修飾句に対する例を付す。日本語共起辞書レコードは、日本語コーパスレコードの内容から係り受けを構成している部分、すなわち、共起句を抽出し、それに関連する情報を再構成したものである。日本語共起辞書は日本語共起辞書レコードを<句見出し>の五十音順に並べたものである。

| ===== [日本語共起辞書レコード] ===== |  |
|---------------------------|--|
| <レコード番号>                  | :レコードタイプと識別番号                                    |
| <見出し情報>                   | :共起辞書レコードの見出し                                    |
| <句見出し>                    | *1 :共起句の表記                                       |
| <共起句構成要素情報>               | :共起句を構成する形態素列に関する情報                              |
| <構成要素列>                   | *2 :共起句を構成する個々の形態素                               |
| <構文情報>                    | :共起句の構文構造を示す情報                                   |
| <部分構文木>                   | *3 :共起句の構造を示す構文木                                 |
| <意味情報>                    | :深層の概念関係を示す情報                                    |
| <部分意味フレーム>                | *4 :概念関係を示す意味フレーム情報                              |
| <共起状況情報>                  | :日本語コーパスにおける共起状況を示す情報                            |
| <頻度>                      | :日本語コーパスにおける出現頻度                                 |
| <例文>...                   | :日本語コーパスで出現した文を共起関係に基づき再構成した文                    |
| <管理情報>                    | :辞書開発の管理のための情報                                   |
| <管理履歴レコード>                | :更新日付等の管理情報                                      |
| *1.....                   |  |
| <句見出し>                    | ::=<単語表記 1><共起関係子><単語表記 2>                       |
| <単語表記 1>                  | ::=<単語表記>  |
| <単語表記 2>                  | ::=<単語表記>  |
| .....                     |  |
| <単語表記>                    | :共起句を構成する単語の表記文字列                                |
| <共起関係子>                   | :共起関係子   |
| *2.....                   |  |
| <構成要素列>                   | ::=<構成要素>。。。                                     |
| <構成要素>                    | ::='<要素番号> <形態素> <かな表記> <品詞><br><慣用句フラグ> <概念情報>' |
| <概念情報>                    | ::='   <意味情報>   <補足付き概念説明>                       |
| <意味情報>                    | ::=<概念識別子><概念見出し><概念説明>                          |
| <概念識別子>                   | ::=<16 進整数>                                      |
| <概念見出し>                   | ::=<英語概念見出し> <日本語概念見出し>                          |

|            |  |
|------------|--|
| <概念説明>     | ::=<英語概念説明> <日本語概念説明>                                  |
| <補足付き概念説明> | ::=" <文字列> "   |
| <hr/>      |  |
| <要素番号>     | :文中の出現順に従った通し番号  |
| <形態素>      | :共起句を構成する形態素の表記  |
| <かな表記>     | :共起句を構成する形態素のかな表記                                      |
| <品詞>       | :共起句を構成する形態素の品詞  |
| <慣用句フラグ>   | :形態素の概念との対応関係の種類。(0)通常概念であり直接対応する。(1)慣用句の一部であり直接対応しない。 |
| <日本語概念見出し> | :概念の日本語概念見出し   |
| <英語概念見出し>  | :概念の英語概念見出し  |
| <日本語概念説明>  | :概念の日本語による説明   |
| <英語概念説明>   | :概念の英語による説明  |
| <複合概念説明>   | :複合概念の日本語による説明   |
| <補足付き概念説明> | :単語辞書の単の概念内に適当な概念がない場合に記述される概念の説明                      |

### \*3.....

|         |                                  |
|---------|----------------------------------|
| <部分構文木> | ::=<受け側要素> <関係要素> <係り側要素>        |
| <受け側要素> | ::=<受け側要素番号> / <受け側単語表記>         |
| <関係要素>  | ::=<関係要素番号> / <共起関係子> / <関係単語表記> |
| <係り側要素> | ::=<係り側要素番号> / <係り側単語表記>         |

|           |                    |
|-----------|--------------------|
| <受け側要素番号> | :受け側に対応する要素番号のリスト  |
| <受け側単語表記> | :受け側の単語表記          |
| <関係要素番号>  | :共起関係に対応する要素番号のリスト |
| <共起関係子>   | :共起の状態を示す共起関係子     |
| <関係単語表記>  | :共起関係に対応する単語表記     |
| <係り側要素番号> | :係り側に対応する要素番号のリスト  |
| <係り側単語表記> | :係り側の単語表記          |

### \*4.....

|            |  |
|------------|--|
| <部分意味フレーム> | ::=<受け側概念要素> { ""   <概念関係子> } <係り側概念要素>    |
| <受け側概念要素>  | ::=""   <受け側要素番号> / <受け側概念識別子> / <受け側単語表記> |
| <係り側概念要素>  | ::=""   <係り側要素番号> / <係り側概念識別子> / <係り側単語表記> |

|            |                   |
|------------|-------------------|
| <受け側要素番号>  | :受け側に対応する要素番号のリスト |
| <受け側概念識別子> | :受け側の概念識別子        |
| <受け側単語表記>  | :受け側の単語表記         |
| <概念関係子>    | :深層の概念関係を示す関係子    |





|     |    |    |     |   |  |
|-----|----|----|-----|---|--|
| { 1 | 借り | カリ | 動詞  | 0 | 3cfdb4 borrow 借りる [カリ・ル]<br>"to use a person's property after<br>promising to give it back to the lender"<br>返す約束で、他人のものをを使う } |
| { 2 | た  | タ  | 助動詞 | 0 | " "  |
| { 3 | 本  | ホン | 名詞  | 0 | 0e5097 volume 本 [ホン]<br>"Publications" 書籍 }  |

<構文情報>

<部分構文木>

|         |             |
|---------|-------------|
| <受け側要素> | 1/借り        |
| <関係要素>  | 2/@rentai/た |
| <係り側要素> | 3/本         |

<意味情報>

<部分意味フレーム>

|           |             |
|-----------|-------------|
| <受け側概念要素> | 1/3cfdb4/借り |
| <概念関係子>   |             |
| <係り側概念要素> | 3/0e5097/本  |

<共起状況情報>

|      |                               |
|------|-------------------------------|
| <頻度> | 5;1,562;231                   |
| <例文> | {000600000067-18-15/<借り>た(本)} |

<管理情報>

|            |                |
|------------|----------------|
| <管理履歴レコード> | DATE="95/3/31" |
|------------|----------------|

## 付.9 英語共起辞書

英語共起辞書レコードの仕様は日本語共起辞書レコードの仕様と同じである。述語句の例のみを示す。英語共起辞書レコードは、英語コーパスレコードの内容から係り受けを構成している部分、すなわち、共起句を抽出し、それに関連する情報を再構成したものである。英語共起辞書は英語共起辞書レコードをく句見出し>のアルファベット順に並べたものである。

```

-----[英語共起辞書レコードの例]-----
<レコード番号>                      ECC157145
<見出し情報>
    <句見出し>                      eaten  @d-object  lunch
<共起句構成要素情報>
<要素番号><形態素><原形><品詞>  <慣用句フラグ>  <概念情報>
{ 1      eaten  eat   VERB    0      3bc6f0  eat  食べる [タベ・ル]  "to eat
something"      食物をとる }
{ 2      lunch  lunch  NOUN    0      3bec74  lunch  昼食 [チュウシヨク] "a
meal eaten at noon"  昼の食事 }

<構文情報>
    <部分構文木>
    <受け側要素>                      1/eaten
        <関係要素>                      2/@d-object/"
        <係り側要素>                      2/lunch
<意味情報>
    <部分意味フレーム>
        <受け側概念要素>                1/3bc6f0/eaten
        <概念関係子>                    object
        <係り側概念要素>                1/3bec74/lunch
<共起状況情報>
    <頻度>                              3;2;173;65
    <例文>                              {003000002264/ have you (eaten) <lunch>}
<管理履歴レコード>                    DATE="95/3/31"

```

## 付.10 日本語コーパス

日本語コーパスレコードの仕様を記述する。新聞記事からの一文に対する例を付す。日本語コーパスは日本語コーパスレコードを<文>のJIS漢字コード順に並べたものである。

| =====【日本語コーパスレコード】===== |                                      |
|-------------------------|--------------------------------------|
| <レコード番号>                | :レコードタイプと識別番号                        |
| <文情報>                   |                                      |
| <テキスト番号>                | :文管理番号                               |
| <出典情報>                  | :出典テキスト名                             |
| <文>                     | :用例文の表記                              |
| <構成要素情報>                |                                      |
| <構成要素列>                 | *1                                   |
| <形態素情報>                 |                                      |
| <形態素列>                  | *2                                   |
| <構文情報>                  |                                      |
| <構文木>                   | *3                                   |
| <意味情報>                  |                                      |
| <意味フレーム>                | *4                                   |
| <管理情報>                  |                                      |
| <管理履歴レコード>              | :更新日付等の管理情報                          |
| *1.....                 |                                      |
| <構成要素列>                 | ::=<構成要素>。。。。                        |
| <構成要素>                  | ::=<構成要素番号> <表記><かな表記> <品詞><概念選択>    |
| <概念選択>                  | ::=<概念識別子>   <補足付き概念説明>   <複合語形態素番号> |
| .....                   |                                      |
| <構成要素番号>                | :形態素番号または複合語形態素番号                    |
| <表記>                    | :形態素または複合語の表記                        |
| <かな表記>                  | :表記のカタカナによる読み                        |
| <概念選択>                  |                                      |
| <品詞>                    |                                      |
| <概念識別子>                 | :概念の同一性を示す番号                         |
| <補足付き概念説明>              | :単語辞書の単語の概念内に適当な概念がない場合に記述される概念の説明   |
| <複合語形態素番号>              | :複合語の番号                              |
| *2.....                 |                                      |
| <形態素列>                  | ::=<形態素列要素>。。。/<複合語化列>               |
| <形態素列要素>                | ::=/ <構成要素番号> : <表記>                 |



|           |   |
|-----------|---|
| <複合語化列>   | ::=' <複合語化列要素>。。。 / '   |
| <複合語化列要素> | ::=' <複合語形態素番号> <複合語化>  |
| <複合語化>    | ::=' { <形態素列要素>   <空> }。。。 /  |
| <空>       | ::=' /  |
| *3.....   |   |
| <構文木>     | ::=' <葉>   <中間節点>   |
| <葉>       | ::=' ( <葉識別子> <構成要素番号> <表記> )'  |
| <中間節点>    | ::=' ( <合成関係子> <副節点>。。。 '   ' ( <合成関係子><br><複合語形態素番号> <表記> <副節点>。。。 '  <br>' ( <合成関係子> <数字列> <副節点>。。。 ' ) |
| <副節点>     | ::=' <主節点>   <葉>   <中間節点>   |
| .....     |   |
| <葉>       | :終端節点   |
| <中間節点>    | :非終端節点  |
| <葉識別子>    | :節点が葉であることを示す指標   |
| <合成関係子>   | :合成関係の種類を示す。合成関係には、4種類ある。   |
| <主節点>     | :中心となる下位節点  |
| *4.....   |   |
| <意味フレーム>  | ::=[ <関係スロット>。。。 ]  |
| <関係スロット>  | ::=[ <関係スロット名> <要素概念> ]  <br>[ <関係スロット名> <概念属性子>。。。 ]  <br>[ <関係スロット名> <文属性子>。。。 ]                        |
| <関係スロット名> | ::=' <概念関係子>   main   which   attribute   S-attribute   |
| <要素概念>    | ::=' <要素番号> ; <表記> ; <概念選択>   <概念関係表現>  |
| <要素番号>    | ::=' <構成要素番号>   <追加概念番号>  |
| <概念選択>    | ::=' <概念識別子>   <補足付き概念説明>   <複合語形態素番号>  |
| .....     |   |
| <概念関係子>   | :事象や事実を表すための概念の関係子  |
| <概念属性子>   | :各々の要素につけられた属性子   |
| <文属性子>    | :文につけられた属性子   |
| <追加概念番号>  | :概念を追加した場合の番号   |

|                           |                               |
|---------------------------|-------------------------------|
| .....[日本語コーパスレコードの例]..... |                               |
| <レコード番号>                  | JCO                           |
| <文情報>                     |                               |
| <テキスト番号>                  | 000500017459                  |
| <出典情報>                    | 朝日新聞 8 7 0 3 0 1              |
| <文>                       | 会場は熱気に包まれ、集会后、周辺路上でのデモ行進に移った。 |

<構成要素情報>

| <構成要素番号> | <表記> | <かな表記> | <品詞> | <概念選択>                  |
|----------|------|--------|------|-------------------------|
| 1        | 会場   | カイジョウ  | 名詞   | 3c0841                  |
| 2        | は    | ハ      | 助詞   | 2621d5                  |
| 3        | 熱気   | ネッキ    | 名詞   | 102ab4                  |
| 4        | に    | ニ      | 助詞   | 2621d5                  |
| 5        | 包    | ツツ     | 動詞   | 3ce654                  |
| 6        | ま    | マ      | 語尾   | 2621cb                  |
| 7        | れ    | レ      | 助動詞  | 2621c1                  |
| 8        | 、    | 、      | 記号   | 2621d7                  |
| 9        | 集会   | シュウカイ  | 名詞   | 3cec82                  |
| 10       | 後    | ゴ      | 名詞   | 3d0476                  |
| 11       | 、    | 、      | 記号   | 2621d7                  |
| 12       | 周辺   | シュウヘン  | 名詞   | 3cf780                  |
| 13       | 路上   | ロジョウ   | 名詞   | 10ebf5                  |
| 14       | で    | デ      | 助詞   | 2621d5                  |
| 15       | の    | ノ      | 助詞   | 2621d5                  |
| 16       | デモ   | デモ     | 名詞   | I#1                     |
| 17       | 行進   | コウシン   | 名詞   | I#1                     |
| 18       | に    | ニ      | 助詞   | 2621d5                  |
| 19       | 移    | ウツ     | 動詞   | 0e5eca                  |
| 20       | っ    | ツ      | 語尾   | 2621ce                  |
| 21       | た    | タ      | 助動詞  | 2621c6                  |
| 22       | 。    | 。      | 記号   | 2621d8                  |
| I#1      | デモ行進 | デモコウシン | 名詞   | = Z 何かに反対して行う路上での<br>行進 |

<形態素情報>

/1:会場/2:は/3:熱気/4:に/5:包/6:ま/7:れ/8:、/9:集会/10:後/11:、/12:周辺/13:路上/14:で/15:の/16:  
デモ/17:行進/18:に/19:移/20:っ/21:た/22:。/[I#1/16:デモ/17:行進/]

<構文情報>

(S (t (M (S (t (M (S (t (W 1 "会場"))  
(W 2 "は"))  
(t (M (S (t (W 3 "熱気"))  
(W 4 "に"))  
(t (S (t (W 5 "包"))  
(W 6 "ま")  
(W 7 "れ"))))))))  
(W 8 "、"))  
(t (M (S (t (M (W 9 "集会")  
(t(W 10 "後"))))  
(W 11 "、"))  
(t (M (S (t (M (S (t (M (W 12 "周辺")

```

(t(W 13 "路上"))))
(W 14 "で")
(W 15 "の")
(t (W 16 "デモ行進"))
(t (I 1 "デモ行進"
(W 16 "デモ")
(W 17 "行進"))))
(W 18 "に"))
(t (S (t (W 19 "移")
(W 20 "っ")
(W 21 "た"))))))))
(W 21 "。"))

```

<意味情報>

```

[ [main 18:移:0e5eca]
  [time [ [main 10:後:3d0476]
           [modifier 9:集会:3cec82]]]
  [goal [ [main I#1:デモ行進:"= Z 何かに反対して行う路上での行進"]
           [place [ [main 13:路上:10ebf5]
                     [modifier 12:周辺:3cf780]]]]]
  [and [ [main 5:包:3ce654]
          [object 1:会場:3c0841]
          [implement 3:熱気:102ab4]]]]]

```

---



## 付.1 1 英語コーパス

英語コーパスレコードの仕様は日本語コーパスレコードの仕様と同じである。新聞記事からの一文に対する例のみを示す。英語コーパスは英語コーパスレコードを<文>のJISコード順に並べたものである。

-----[英語コーパスレコードの例]-----

### <文情報>

|          |  |
|----------|--|
| <テキスト番号> | 0020000026cd   |
| <出典情報>   | Japan Times  |
| <文>      | The United States has singled out the Kansai airport project as an example of the closed nature of the Japanese construction market. |

### <構成要素情報>

| <構成要素番号> | <表記>          | <原形表記>        | <品詞> | <概念選択>                           |
|----------|---------------|---------------|------|----------------------------------|
| 1        | the           | the           | ART  | 2dc2f4                           |
| 2        | /             | /             | BLNK | 2dc2ed                           |
| 3        | United/States | United/States | NOUN | "W the United States of America" |
| 4        | /             | /             | BLNK | 2dc2ed                           |
| 5        | has           | have          | AUX  | 2dc2fd                           |
| 6        | /             | /             | BLNK | 2dc2ed                           |
| 7        | singl         | single        | VT   | I#1                              |
| 8        | ed            | ed            | SUF  | 2dc2ed                           |
| 9        | /             | /             | BLNK | 2dc2ed                           |
| 10       | out           | out           | PTCL | I#1                              |
| 11       | /             | /             | BLNK | 2dc2ed                           |
| 12       | the           | the           | ART  | 2dc2f4                           |
| 13       | /             | /             | BLNK | 2dc2ed                           |
| 14       | Kansai        | Kansai        | NOUN | I#2                              |
| 15       | /             | /             | BLNK | 2dc2ed                           |
| 16       | airport       | airport       | NOUN | I#2                              |
| 17       | /             | /             | BLNK | 2dc2ed                           |
| 18       | project       | project       | NOUN | I#2                              |
| 19       | /             | /             | BLNK | 2dc2ed                           |
| 20       | as            | as            | PREP | 2dc2ef                           |
| 21       | /             | /             | BLNK | 2dc2ed                           |
| 22       | an            | an            | ART  | 2dc2f3                           |
| 23       | /             | /             | BLNK | 2dc2ed                           |
| 24       | example       | example       | NOUN | 0bb9e3                           |
| 25       | /             | /             | BLNK | 2dc2ed                           |
| 26       | of            | of            | PREP | 2dc2ef                           |

|     |                            |                            |      |  |
|-----|----------------------------|----------------------------|------|--|
| 27  | /                          | /                          | BLNK | 2dc2ed   |
| 28  | the                        | the                        | ART  | 2dc2f4   |
| 29  | /                          | /                          | BLNK | 2dc2ed   |
| 30  | closed                     | closed                     | ADJ  | 0b3e1a   |
| 31  | /                          | /                          | BLNK | 2dc2ed   |
| 32  | nature                     | nature                     | NOUN | 0ca429   |
| 33  | /                          | /                          | BLNK | 2dc2ed   |
| 34  | of                         | of                         | PREP | 2dc2ef   |
| 35  | /                          | /                          | BLNK | 2dc2ed   |
| 36  | the                        | the                        | ART  | 2dc2f4   |
| 37  | /                          | /                          | BLNK | 2dc2ed   |
| 38  | Japanese                   | Japanese                   | ADJ  | 0a8dcf   |
| 39  | /                          | /                          | BLNK | 2dc2ed   |
| 40  | construction               | construction               | NOUN | 0b5134   |
| 41  | /                          | /                          | BLNK | 2dc2ed   |
| 42  | market                     | market                     | NOUN | 0c7d6d   |
| 43  | .                          | .                          | PUNC | 2dc2e5   |
| I#1 | singl/out                  | single/out                 | VT   | choose, pick, one person or thing<br>from among several for special<br>comment, treatment etc. |
| I#2 | Kansai/airport/<br>project | Kansai/airport/<br>project | NOUN | a project to build an international<br>airport in the Osaka Bay                                |

<形態素情報>

/1:the/2: /3:United States/4: /5:has/6: /7:singl/8:ed/9: /10:out/11: /12:the/13: /14:Kansai/15:  
/16:airport/17: /18:project/19: /20:as/21: /22:an/23: /24:example/25: /26:of/27: /28:the/29:  
/30:closed/31: /32:nature/33: /34:of/35: /36:the/37: /38:Japanese/39: /40:construction/41:  
/42:market/43:/I#1/7:singl/10:out/I#2/14:Kansai/16:airport/18:project/}

<構文情報>

(S (t (M (S (S (t (W 1 "The"))  
(W 2 " "))  
(t (S (t (W 3 "United States"))  
(W 4 " "))))  
(t (M (t (M (t (S (S (t (W 5 "has"))  
(W 6 " ")))  
(t (S (t (W 7 "singled out")  
(I 1 "singled out"  
(S (t (S (t (W 7 "singl"))  
(W 8 "ed"))  
(W 9 " ")))  
(W 10 "out"))  
(W 11 " "))))  
(S (S (t (W 12 "the"))

```

(W 13 " ")
(t (S (t (W 14 "Kansai airport project")
(I 2 "Kansai airport project"
(S (t (W 14 "Kansai"))
(W 15 " ")
(S (t (W 16 "airport"))
(W 17 " ")
(W 18 "project"))))
(W 19 " "))))
(S (S (t (W 20 "as"))
(W 21 " "))
(t (M (t (S (S (t (W 22 "an"))
(W 23 " ")
(t (S (t (W 24 "example))
(W 25 " "))))
(S (S (t (W 26 "of"))
(W 27 " "))
(t (M (t (S (S (t (W 28 "the"))
(W 29 " ")
(t (M (S (t (W 30 "closed"))
(W 31 " ")
(t (S (t (W 32 "nature"))
(W 33 " "))))))
(S (S (t (W 34 "of"))
(W 35 " "))
(t (S (S (t (W 36 "the"))
(W 37 " ")
(t (M (S (t (W 38 "Japanese"))
(W 39 " ")
(t (M (S (t (W 40
"construction"))
(W 41 " ")
(t (W 42
"market"))))))))))))))))
(W 43 " "))

```

<意味情報>

```

[ [main I#1:singled out:"=Z to choose, pick, one person or thing from among several for
special comment, treatment etc."]
[agent 3:United States:"=W the United States of America"]
[object I#2:Kansai airport project:"=Z a project to build an international airport in the
Osaka Bay"]
[modifier [ [main 24:example:0bb9e3]

```



```
[modifier [
  [main 32:nature:0ca429]
  [modifier 30:closed:0b3e1a]
  [modifier [
    [main 42:market:0c7d6d]
    [modifier 38:Japanese:0a8dcf]
    [modifier 40:construction:0b5143]]]]]
```

[attribute already end]]

---

## 参考文献

- [芦崎 94] 芦崎達雄: JICST 機械翻訳システム(1) 機械翻訳辞書の作成, 情報管理, Vol.37, No.1 (1994).
- [BNC 92] BNC: British National Corpus - Fact Sheet(1992).
- [Briscoe et al. 90] Briscoe, T. et al.: ACQUILEX: Acquisition of Lexical Knowledge for Natural Language Processing Systems, *Esprit BRA-3030 Periodic Progress Report*, No.1 (1990).
- [Briscoe 91] Briscoe, T.: Lexical Issues in Natural Language Processing, in E. Klein and F. Vetman (eds.), *Natural Language and Speech*, Springer-Verlag, pp.39-68 (1991).
- [Fuchi and Yokoi 94] Fuchi, K. and Yokoi, T. (Eds.): Knowledge Building and Knowledge Sharing, p. 329, *Ohmsha and IOS Press* (1994).
- [Grishman et al. 94] Grishman, R., Macleod, C., and Meyers, A.: Complex syntax: Building a computational lexicon, *Proc. 15th Int'l Conf. Computational Linguistics (COLING94)* (1994).

[Guha and Lenat 94] Guha, R.V. and Lenat, D.B.: Enabling Agents to Work Together, *CACM*, Vol.37, No.7, pp.123-142 (1994).

[池原ほか 93] 池原悟、宮崎正弘、横尾昭男: 日英機械翻訳のための意味解析用の知識とその分解能, 情報処理学会誌, Vol.34, No.8, pp.1692-1704 (1993).

[情報処理振興事業協会 87] 情報処理振興事業協会技術センター: 計算機用日本語基本動詞辞書 IPAL(Basic Verbs)-解説編-, 61 技-073 (1987).

[情報処理振興事業協会 90] 情報処理振興事業協会技術センター: 計算機用日本語基本形容詞辞書 IPAL (Basic Adjectives) 一解説編一, 2 技-114 (1990).

[国立国語研究所 64] 国立国語研究所: 分類語彙表, 資料集 6 (1964).

[国際情報化協力センター 93] 国際情報化協力センター機械翻訳システム研究所: 中間言語 (最終版), P.52 (1993).

[Lloyd 82] Lloyd, S. M.: Roget's Thesaurus of English words and phrases, *Longman* (1982).

[Marcus et al. 93] Marcus, M., Santorini, B. and Marcinkiewicz, M.: Building a Large Annotated Corpus of English: The Penn Treebank, *Computational Linguistics*, Vol.19, No.2, pp.313-330 (1993).



- [Miller et al. 93] Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K. and Tengi, R.: Five Papers on WordNet, *CSL Report 43, Cognitive Science Laboratory, Princeton University* (1993).
- [Komatsu et al. 94] Komatsu, E., Cui, J. and Yasuhara, H.: English Generation from Interlingua by Example-Based Method, *Proc. 15th Int'l Conf. Computational Linguistics (COLING 94)*, pp. 363-368 (1994).
- [Nakao and Ogino 94] Nakao, Y. and Ogino, T.: EDR Concept Classification and Methodology for its Development, *Proc. 47th FID Conference and Congress*, pp. 122-127 (1994).
- [中右 94] 中右実: 認知意味論の原理、p. 499, 大修館書店 (1994).
- [日本電子化辞書研究所 95] 日本電子化辞書研究所: EDR 電子化辞書仕様説明書 (第二版), 日本電子化辞書研究所, TR-045 (1995).
- [Normier and Nossin 91] Normier, B. and Nossin, M.: GENELEX Project : EUREKA for Linguistic Engineering, *EDR TR-031* (1991).
- [Serasset 94] Serasset, G.: Recent Trends of Electronic Dictionary Research and Development in Europe, *EDR TM-038* (1994).
- [宇津呂、松本 95] 宇津呂武仁、松本裕治: コーパスを用いた言語知識の獲得, 人工知能学会誌, Vol. 10, No. 2, pp. 197-204 (1995).

- [Yasuhara 93] Yasuhara, H.: An Example-Based Multilingual MT System in a Conceptual Language, *Proc. MT Summit IV*, pp.147-155(1993).
- [Yokoi 94] Yokoi, T.: Very Large-Scale Knowledge Bases Embodying Intelligence Space, in K. Fuchi and T. Yokoi (Eds.), *Knowledge Building and Knowledge Sharing*, Ohmsha and IOS Press, pp.9-18 (1994).
- [横井ほか 95] 横井俊夫、安原宏、村木一至、原田千秋、丸山冬樹: 汎電子化辞書: 言語知識のアーキテクチャ, 言語処理学会第1回年次大会論文集, pp.185-188 (1995).
- [Yokoi 95] Yokoi, T.: The Impact of the EDR Electronic Dictionary on Very Large Knowledge Bases, in Mars, N.J.I. (Ed.) *Towards Very Large Knowledge Bases*, Ohmsha and IOS Press, pp. 13-12 (1995).
- [Yokoi and Yasuhara 96] Yokoi, T. and Yasuhara, H.: The Information Structure of Linguistic Knowledge, *Information Systems* (投稿中)(1996).
- [横井ほか 96a] 横井俊夫、木村和広、小泉敦子、三吉秀夫: 表層レベルにおける電子化辞書の情報構造, 情報処理学会論文誌, Vol.37, No.3 (1996).
- [横井ほか 96b] 横井俊夫、仲尾由雄、荻野孝野、田中裕一: 概念レベルにおける電子化辞書の情報構造, 情報処理学会論文誌(投稿中)(1996).
- [Zernik 91] Zernik, U. (ed.): Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon, *Lawrence Erlbaum Associates* (1991).

## 本論文に関する原著論文

### 学術論文（論文誌）

1. Toshio Yokoi and Hiroshi Yasuhara: The Information Structure of a Large-Scale Electronic Dictionary, *Information Systems* (submitted).
2. 横井俊夫、木村和広、小泉敦子、三吉秀夫: 表層レベルにおける電子化辞書の情報構造, 情報処理学会論文誌, Vol.37, No.3, 1996.
3. 横井俊夫、仲尾由雄、荻野孝野、田中裕一: 概念レベルにおける電子化辞書の情報構造, 情報処理学会論文誌（投稿中）。

### 国際会議

1. Toshio Yokoi: Electronic Dictionary : A Language Knowledge Base, *International Forum for Translation Technology*, pp.63-64, 1989.
2. Toshio Yokoi: Large-Scale Knowledge Base for Multilingual Machine Translation, *International Symposium on Multilingual Machine Translation '90 (MMT '90)*, pp.46-50 (guest lecture), 1990.



3. Toshio Yokoi: Large scale electronic dictionary for the basis of knowledge-base, *International Symposium on Mineral Exploration (ISME-AI '90)*, pp.22-28 (special talk), 1990.
4. Toshio Yokoi: Knowledge Archives Towards an International Basis for Knowledge Sharing and Reuse, *10th European Conference on Artificial Intelligence (ECAI '92) Workshop on Knowledge Sharing and Reuse: Ways and Means*, pp.63-65, 1992.
5. Toshio Yokoi: From Electronic Dictionaries to Very Large - Scale Knowledge Bases, *Fourth Machine Translation Summit (MT Summit IV)*, pp.47-62 (invited speech), 1993.
6. Toshio Yokoi: Present Status and Future Plan for the EDR Electronic Dictionary, *Japan-US Workshop on Machine Aided Translation*, pp.109-114, 1993.
7. Toshio Yokoi: Very Large - Scale Knowledge Bases Embodying Intelligence Space, *International Conference on Building and Sharing of Very Large - Scale Knowledge Bases '93 (KB&KS '93)*, pp.11-20 (keynote address), 1993.
8. Toshio Yokoi: Recent Trends in the Development of Very Large - Scale Knowledge Bases, *Japan / Korea Joint Conference on Expert Systems (JKJCES '94)*, pp.1-10 (invited speech), 1994.
9. Toshio Yokoi: The EDR Electronic Dictionary and Its Role in the Information Science, *International Federation for Information and Documentation Conference and Congress (47th FID)*, pp.143-147 (invited lecture), 1994.

10. Toshio Yokoi: The Impact of the EDR Electronic Dictionary on Very Large Knowledge Bases, International Conference on Building and Sharing of Very Large - Scale Knowledge Bases '95 (KB&KS '95), pp.13-21(invited lecture) in *N. J. I. Mars (Ed.) Towards Very Large Knowledge Bases*, IOS Press and Ohmsha, 1995.

#### 研究会、大会、シンポジウム等

1. 横井俊夫: 大規模電子化辞書, 人工知能学会 人工知能基礎論/ヒューマンインターフェースと認知モデル/知識ベースシステム・3 研究会合同研究会, SIG - FAI / HICG / KBS-8901, pp.57-60, 1989.

2. 横井俊夫: 大規模言語データの整備に向けて, 電子情報通信学会第二種研究会「言語獲得と概念形成過程の工学的モデル化時限研究専門委員会」, LA90-5, p.8, 1990.

3. 横井俊夫: 大規模知識ベース研究開発の構想と課題, 情報処理学会: アドバンスト・データベースシステムシンポジウム '90, pp.1-4 (基調講演), 1990.

4. 横井俊夫: 自然言語処理・電子化辞書・大規模知識ベース, 電子情報通信学会・日本ソフトウェア科学会: 自然言語処理における統合シンポジウム, pp.11-16(招待講演), 1991.

5. 横井俊夫: EDR 概念辞書の考え方, 構成, 開発法, 電子情報通信学会第二種研究会「言語獲得と概念形成過程の工学的モデル化時限研究専門委員会」, LA91-6, p.10, 1991.

- 6.横井俊夫: 知識アーカイブプロジェクト—大規模知識獲得をめざして—, 電子情報通信学会第二種研究会「言語と知識の獲得・運用時限研究専門委員会」,LK92-1, p.8,1992.
- 7.横井俊夫: 語彙知識の自己組織化に向けて—EDR 電子化辞書の構造と開発手法—, 情報処理学会情報学基礎研究会, 92-FI-26, pp.1-8,1992.
- 8.横井俊夫: メディアと大規模知識, 情報処理学会メディアと情報処理シンポジウム,情報シンポジウム Vol.92, No.6, pp.3-11,1992.
- 9.横井俊夫: 知識処理と言語処理の融合としての大規模知識ベース, 人工知能学会知識ベースシステム研究会,知識共有/再利用ワークショップ論文集,p.10,1992.
- 10.横井俊夫: なぜ大規模知識ベースが必要か, 情報処理学会知識のリフォーメーションシンポジウム, pp.49-52,1993.
- 11.横井俊夫、安原宏、村木一至、原田千秋、丸山冬樹: 汎電子化辞書: 言語知識のアーキテクチャ, 言語処理学会第1回年次大会, (B3-2),1995.
- 12.横井俊夫、木村和広、小泉敦子、三吉秀夫、川田亮一: 汎電子化辞書: 表層レベルの構造,言語処理学会第1回年次大会,(B3-3),1995.
- 13.横井俊夫、田中裕一、仲尾由雄、荻野孝野、野口喜洋: 汎電子化辞書: 深層レベルの構造,言語処理学会第1回年次大会, (B3-4),1995.



学会誌解説論文

- 1.横井俊夫、相場亮: 制約ロジック・プログラミング — 知識処理への新しいパラダイム —, 情報処理, Vol.30, No.1, pp.29-38, 1989.
- 2.横井俊夫: 電子化辞書から電子化用語辞書へ, 情報の科学と技術, Vol.41, No.4, pp.337-343, 1991.
- 3.横井俊夫: 電子化辞書とテキストデータベース, 日本語学, Vol.10, No.8, pp.78-85, 1991.
- 4.横井俊夫: 語彙知識から世界知識へ — 自然言語処理を拠所とする大規模知識ベースの構想 —, システム制御情報学会誌・システム／制御／情報, Vol.36, No.7, pp.460-465, 1992.
- 5.横井俊夫: 知識処理と自然言語処理の融合としての大規模知識ベース — 電子化辞書から知識アーカイブへ —, 人工知能学会誌, Vol.8, No.3, pp.286-296, 1993.
- 6.横井俊夫: 知の空間を構成する大規模知識ベース — 知識インフラの構築とインテリジェント化の技術 —, 情報処理, Vol.34 No.12, pp.1449-1457, 1993.
- 7.横井俊夫: 大規模知識ベースシステム — 知能システムのユニバーサルアーキテクチャに向けて —, 計測と制御 (計測自動制御学会誌), Vol.33, No.1, pp.52-57, 1994.
- 8.横井俊夫: EDR 電子化辞書, 電子情報通信学会誌, Vol.77, No.12, pp.1228-1231, 1994.

9.Toshio Yokoi: The EDR Electronic Dictionary, *CACM*, Vol.38, No.11, pp.42-44, 1995.

## 図一覧

|                         |    |
|-------------------------|----|
| 2.1 全体構造                | 26 |
| 2.2 EDR 電子化辞書（日本語）による実現 | 29 |
| 2.3 EDR 電子化辞書（英語）による実現  | 30 |
| 3.1 全体構造と表層レベル          | 33 |
| 3.2 表層辞書項目              | 34 |
| 3.3 語表層辞書項目             | 36 |
| 3.4 文表層辞書項目             | 39 |
| 3.5 EDR 電子化辞書の語表層辞書対応部分 | 43 |
| 3.6 EDR 電子化辞書の文表層辞書対応部分 | 44 |
| 4.1 全体構造と概念レベル          | 53 |
| 4.2 概念辞書項目              | 55 |
| 4.3 語概念辞書項目             | 59 |
| 4.4 文概念辞書項目             | 62 |
| 4.5 EDR 電子化辞書の語概念辞書対応部分 | 66 |
| 4.6 基本語概念体系             |    |
| (a) 基本語概念体系の上位層部        | 67 |
| (b) 概念系列の例              | 68 |
| 4.7 EDR 電子化辞書の文概念辞書対応部分 | 69 |

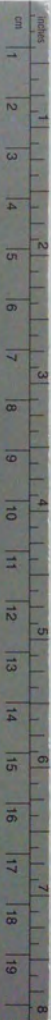


## 表一覧

|   |    |
|---|----|
| 2.1 EDR 電子化辞書の規模（1995 年 4 月第 1 版の基本語対応部分） | 31 |
| 3.1 EDR 日本語単語辞書・日英対訳辞書諸元                  | 46 |
| 3.2 EDR 英語単語辞書・英日対訳辞書諸元                   | 47 |
| 3.3 EDR 日本語コーパス諸元                         | 49 |
| 3.4 EDR 英語コーパス諸元                          | 50 |
| 4.1 基本語概念体系における一概念当たりの親概念数                | 73 |
| 4.2 基本語概念体系における概念の分布                      | 74 |
| 4.3 概念記述の規模（主要関係子）                        | 75 |







# Kodak Color Control Patches

© Kodak, 2007 TM Kodak

Blue Cyan Green Yellow Red Magenta White 3/Color Black



## Kodak Gray Scale



© Kodak, 2007 TM Kodak

A 1 2 3 4 5 6 M 8 9 10 11 12 13 14 15 B 17 18 19

