

日本電子化辞書研究所における 概念体系

荻野孝野、仲尾由雄、小笠原あゆみ、長澤陽子

(株)日本電子化辞書研究所

当研究所では、日英各20万語の基本語と各10万語の専門用語に関する、単語辞書と概念辞書を開発している。概念辞書には、それらの語彙の語義に相当する「概念」全て(異なりで40万概念)が収録され、「概念」を分類・体系化した概念体系と、「概念」間に成立しうる関係を記述した概念記述の二種の情報が付与されている。

概念体系は、「概念」を単位として分類しているので、従来の語彙分類に見られるような、多義性によって複数の分類カテゴリに対応してしまう、という繁雑さが回避されている。

ここでは、概念体系の概要を明らかにするとともに、大規模な概念体系の開発手順、開発段階ででてきた諸問題とその扱いなどについて、報告する。

THE EDR CONCEPT CLASSIFICATION DICTIONARY

Takano Ogino, Yoshio Nakao, Ayumi Ogasawara, Yoko Nagasawa

Japan Electronic Dictionary Research Institute, Ltd. (EDR)

EDR has been developing a Concept Dictionary, which contains 400,000 concepts listed in Japanese and English word Dictionaries of 300,000 words each. The Concepts Classification Dictionary which is a part of it, classifies all the concepts in it by their meaning.

A multiple-meaning word can be categorized into several word classification groups. Concept classification reduces the number of items that are otherwise categorized multiply.

This paper summarizes the contents of the Concept Classification Dictionary, and reports on the classification procedure and some interesting topics discovered in the course of project.

1. はじめに

日本電子化辞書研究所（略称 EDR）では、1986年より9年間のプロジェクトとして、大規模な電子化辞書の開発に取り組んできた。EDRで開発している電子化辞書は、単語辞書、概念辞書、対訳辞書、共起辞書の4種類の辞書と、EDRコーパスから構成されている。それぞれの辞書には、形態素・構文レベルの情報、意味レベルの情報、語用（言い回し）レベルの情報等が収められている。これらの辞書は、自然言語処理の様々な用途に応じて必要な部分を組み合わせ使用することができるように設計されている。

ここの報告は、上記の辞書のうち、概念辞書に含まれる、概念体系に関するものである。

EDR電子化辞書における概念体系とは、以下のようなものである。

- (1) 単語辞書に登録された全ての語の語義に相当する概念を収録し、概念間の階層関係を記述したものである。
- (2) 概念を単位として分類を行ったものである。
- (3) 分類された類似概念のグループの全てを、階層構造に配置したものである。

2. 概念体系の概要

2.1 EDR電子化辞書における概念体系の位置づけ

EDR単語辞書には、一般の辞書の語義に相当するものとして、概念が、見出しに対応して設定されている。自然言語の文が自立語を主要な単位として構成されるように、EDRの意味表現形式（概念関係表現）では、この概念が、文の意味内容を表現する概念関係表現の主要な単位となる。

概念辞書は、単語辞書で設定された全ての概念について、概念間関係として、関係を構成する二つの概念と関係子から

【概念1】 【概念2】 関係子

の内部形式で表現したデータを蓄積している。このデータは、関係子の違いにより、概念体系と概念記述との2種類のデータとして区別される。agent（行為者）、object（対象）といった文における役割に相当する関係子によって、通常の文における概念間の結び付きの妥当性を示したものが概念記述であり、概念の上位-下位関係を規定する関係子 kind-of によって、概念間の階層関係を示したものが概念体系である。本稿では、このうち、概念体系に関する部分について報告する。

概念体系は、概念の類似度を判断する基礎データであり、また、概念記述のデータを補完する機能を持つ。概念体系上で二つの概念が近い位置にあるということは、それらの概念が類似している、あるいは、ある意味で等価であると解釈できる。ある概念間の関係の妥当性を判断する際に、当該概念間に概念記述データが存在しない場合には、概念体系により類似の概念を求めて、類似概念間の概念記述を当該概念間の概念記述データとして援用することが可能である。

以上で述べた、単語辞書と概念体系、概念記述の関係を、文例にそって見てみよう。

「先生が走る」の「先生」と「走る」は、それぞれ、単語辞書によると、[#教師という職業の人]、[#人や動物が走ること]等に該当する。この概念を使って、この文の意味内容を表現すると、「[#人や動物が走ること-agent→#教師という職業の人]となる。概念辞書上でこの表現の妥当性を調べてみると、例2.1に示すように概念辞書には[#人や動物が走ること]と[#教師という職業の人]との概念記述データはない。そこで、概念体系を使って、それぞれの上位概念に関する概念記述を調べていくと、概念体系上の中間ノードの間に「*主体の移動 -agent→ *人間」という概念記述が得られる。これを援用することで、[#人や動物が走ること]と[#教師という職業の人]がagentで関係づけられたこの表現は妥当であると判断できる。

例2.1 先生が走る。

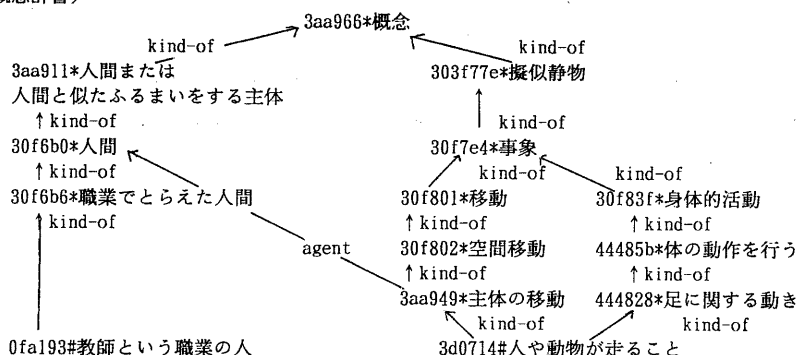
◇（概念関係表現）

走る [#人や動物が走ること] -agent→先生 [#教師という職業の人]

◇（単語辞書）

先生	0fa193	#教師という職業の人
:		
走る	3d0714	#人や動物が走ること
	103ac7	#感情などが瞬間的に現れる
:		

◇（概念辞書）



2. 2 中間ノードとインスタンス

(1) 中間ノード

概念体系において、概念グループを統括する概念分類項目を本稿では中間ノードとよぶ。EDRの概念体系では、約5000の中間ノードがある。これらの中間ノードを示す概念には、単語辞書にリンクされている概念(ここでは単語起源の概念とよぶ)と、中間ノードとして新しく追加した、単語辞書へのリンクがない概念とがある。

後者は、[*始点から終点への移動]のように、概念分類の観点を示したような中間ノードにみられる。

資料1は、体系の上位部分に位置する中間ノードと、下位の方に向かって展開していく中間ノードの数を示したものである。紙面の都合上、中間ノードのすべてを列挙することはできないが、体系の上位に位置する中間ノードについてふれておく。

概念体系の上位の中間ノードには、次のようなものがある。

人間または人間と似たふるまいをする主体→[*人間]、[*生命体]、[*自立活動体]
 存在を具体的に確認できるもの→[*静物]
 存在を具体的に確認できないもの→[*抽象的なもの概念]、[*時間]、[*位置]、[*動き]、
 [*状態]、[*性質]、[*関係]

そのうち、[*動き]は、

主体の意志のかかわりによって起きる[動き]→[*活動]

主体の働きかけや、自然の動きによって起きる、客体側からみた[動き]→[*現象]

に、分ける。従って、例2.2、例2.3に示すように、「倒す」と「倒れる」は、別扱いでそれぞれ異なる中間ノードに配置される。

例 2.2 444c0d [*物を倒す]¹⁾ <²⁾ 444caf [*物に力を加える行為] <30f87e [*対物活動] <30f83e [*活動]

例 2.3 3cf455 [*立っているものが倒れる] <3f985d [*具体物の動き] <30f7ea[*静物に関する現象]<30f7e5[*現象]

(2) インスタンス

中間ノードの下位には、更に下位概念を持つ中間ノードと、それ以上、下位概念を持たないインスタンスの両方が存在する。例2.4は、資料2に示した、中間ノードから下位として展開される概念の一部であるが、中間ノードとインスタンスが混在していることがわかる。

例 2.4 r444b90[*品物でとらえた販売] ----中間ノード
 3c3f99[*品物を売る]>↳3d0613[*物を発売する] ----インスタンス
 ↳3cfd6c[*品物をうりさばく] ----インスタンス

2. 3 数量的概要

表1 EDR概念体系の数量的概要

中間ノード数	5, 300
分類対象となった概念数	400, 000
体系の深さ(最大)	16

3. 概念体系作成の基本方針

(1) 体系の深さ

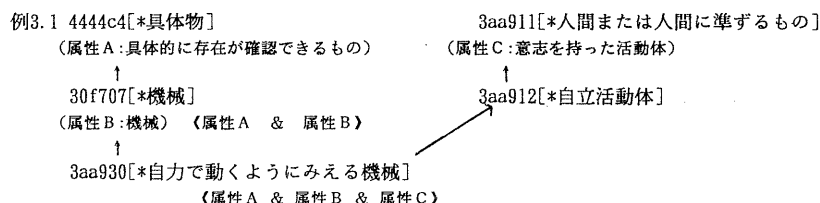
基本的に体系の深さは限定しない(資料1)。

(2) 下位ノードの数

下位に配置される中間ノード数の限定はしない(資料1)。

(3) 中間ノードの配置

中間ノードは、まとめられた概念グループの概念上の共通性に着目して築きあげていくものであり、下位にいくほど共通属性の限定が増え、上にいくほど共通属性の限定が減る。



¹⁾ 本稿では、概念を以下の形式で示す。

概念識別子[概念]

概念識別子は16進数、概念のうち中間ノードに相当するものを[*概念]、インスタンスに相当するものは[#概念]で示す。

²⁾ 例中の「概念A>概念B」は、[概念A]が[概念B]の上位概念であることを示す。

(4)多重リンク

単語一つに一つ概念とは限らず、多くの単語は複数の概念を持っている。いわゆる単語の多義である。EDRの単語辞書では、多義に応じて分解し、概念を設定するのが基本であるから、多義は概念体系の多重リンクではなく、例3.2に示すように単語辞書の概念の区別で扱う。

例3.2 概念レベルで分割---インスタンスの多重リンクは不要

殺す 0e3935[#人を殺す]<30f85e[*殺す]
3cf7a1[#感情を抑える]<3d0232[*我慢する]

ただし、多義としては扱いたいものについては、概念体系上での多重リンクを認めている。例3.3は、単語レベルの概念が複合的な概念で、二つの中間ノードにリンクしている例である。例3.4は、中間ノード自体が二つの上位の中間ノードにリンクしている例である。「歩く」、「走る」などは、[*身体的活動]にも[*主体の移動]にもなる。同じ中間ノードでも別の観点からみれば、他の中間ノードとの間にも上下関係が成り立つ場合の扱いである。

例3.3 売約

104996[#売約約束をする]< |
r30f82a[*売る]
L3cf0d5[*約束する]

例3.4 歩く

3cefe6[*足を使って歩く]< |
r444828[*足に関する動き]<30f83f[*身体的活動]
L3aa949[*主体の移動]<30f802[*空間移動]

(5)複合語の扱い

複合語の構成部分の意味の重みによって扱いを分けている。

(ア)複合語の前構成部分と後構成部分がほぼ同じ重みである場合には、多重リンクで扱う。

例3.5 r30f88f[*切る]
[#切り倒す]< |
L444c0d[*物を倒す]

(イ)前構成部分あるいは後構成部分が付属的で、重みのある方の構成部分の概念が複合語全体の概念を担う場合は、主たる概念のみ考慮する。前後に接辞がついているタイプなどが、これに該当する。

例3.6 [#かっ倒す]<444c0d[*物を倒す]

4. 開発手順

4.1 概念分類と概念体系

概念体系の開発は、「概念的に近似している概念を集める概念分類」と「まとめたグループどうしを概念上の上下関係で配置する体系化作業」からなる。この二つの作業は相互に関連するものであり、

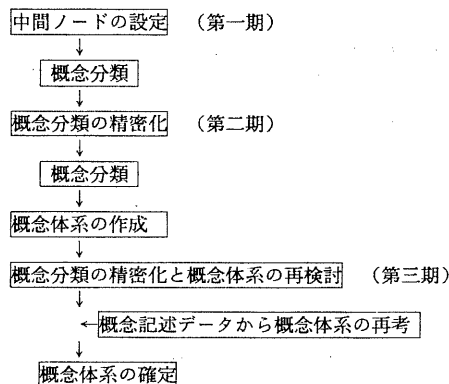
(1)概念分類から概念体系に進むボトムアップ型

(2)恣意的に作成した概念体系から概念分類に進むトップダウン型の両方向からのアプローチがある(文献4)。

4.2 EDRにおける開発順序

EDRの概念体系開発は、図1に示すようにおおまかに分けて3段階の過程があり、開発段階により、アプローチの仕方が異なる。第一期は中間ノードをあらかじめ用意した概念分類で、トップダウンに分類した。第二期は、第一期で過度に多くの概念が集まってしまった項目を対象に、小人数で分析し、ボトムアップで中間ノードを設定した後、多人数で残りの概念をトップダウンに分類した。第三期は既に分類された概念の細分ということで小人数でボトムアップに近いやり方で進めた。第一、二期は、ノード設定者と分類作業者は別であるが、第三期はノード設定と分類作業は同じ作業者にやって行われた。

図1：概念体系の開発段階



○第一期：概念分類

大、中、小分類からなるノードを順次分類者に提示し、分類作業を進めた。まず、20程度の大分類ノードに全概念を分類した後、その中間ノードを更に細分した80程度の中分類ノードに分類した。その後、この中分類ノードを更に細分した660程度の小分類ノードに同様に分類した。この方法では、1回に示すノードは直下に展開するノードのみで、作業者は1回の判断で10個前後の枠のいずれかを選択すればよく、効率的に作業を進めることができた。

○第二期：概念分類の精密化

第一期で細分化したノードのいずれにもあてはまらず、中分類ノードで分類が止まったままの概念を分類するノードを追加した。追加したノードには、既にあるノードを統合するタイプのもものと細分するタイプのもものがあつた。

○第三期：概念分類の精密化と概念体系

第一期、第二期でできた概念分類を元に、分類の精密化と概念体系の構築を行った。作業は、第一、二期と異なり、中間ノード作成者と概念分類をする者が同じで、概念辞書に直接アクセスし、オンラインで修正を進めたものである。

上記のような開発段階を踏んで、概念体系が細分化され、中間ノード数と深さは表2に示すように変化していった。

表2 開発段階における中間ノード数と深さの変化

開発段階	中間ノードの深さ	中間ノード数
1	5 (一定)	770
2	5 (一定)	950
3	最大16 (任意)	5300

5. 本概念体系の特色 --- 既存資料と比較して

既存の意味分類資料の代表的なものとして、国立国語研究所の分類語彙表(文獻5)と角川類語新辞典(文獻6)とがある。これらの資料とEDRの概念体系とを比較参照してみると、それぞれに作成の背景があり、以下のような点で違いが見られる。(1)概念を対象として分類した体系である。

分類語彙表は、単語に複数の意味がある場合は、単語が複数のノードに配置される。分類は、単語の持つ意味別におこなったものと思われるが、書名の一部に含まれる「語彙」に示されるように、エントリーはあくまでも単語見出しで表現されている。使用者は、配置された単語見出しの分類項目から、その単語のどの意味で分類されたものかを推測することになる。

(2)分類対象が大量である。

表3に示すように、EDRの単語辞書自体が大規模なため、多くの概念を対象とした分類となっている。

表3 既存体系との数量的な比較

資料名	中間ノード数	中間ノードの深さ	分類対象数
類語新辞典	約2800	4	約6万語
分類語彙表	約800	6	約3万6千語
EDR概念体系	約5300	最大16 (任意)	約30万語

(3)文中の概念的ふるまいを考慮したグループ化である。

概念体系による概念記述の省力化という、EDRにおける概念体系の役割を考慮し、体系の上位の方の中間ノードについては、文中の概念的ふるまいを考慮した中間ノードの設定となっている。

類語新辞典は、文中の概念の働きとは関係なく、例5.1に示すように、連想される語や関連語を同じグループにするので、概念記述の観点で同じふるまいをするグループとして用いるには不適切な部分がある。類語新辞典では、「降る」も「雨天」も同じ「雨が降ること」のグループであるが、EDRの概念体系では、「雨が降る」は[*現象]に、「雨天」は[*状態]になる。

例5.1 類語新辞典

023c[降雨—雨が降る事] > 降る、降雨、雨天、雨がち、雨あがり

EDRの概念体系

降る:0f08a9[*雨が降ること] < 3f9692[*天候の変化] < 30f7e9[*気象現象] < 30f7e5[*現象]

雨天:3c2c4b[*#雨の降る空模様(雨天)] < 30f951[*天候] < 3f98a9[*気象の値]... < 3aa963[*状態]

「分類語彙表」は、大分類を「体、用、相」で分けてあり、これは文中の概念の働きを考える手掛かりにはなるが、「体、用、相」がそれぞれ「名詞、動詞、形容詞」という品詞上の分類に対応しているため、例5.2に見られるように、概念で同じグループに入るものが必ずしも品詞グループに対応しない場合もあり、大分類項目では別グループになる。ただ、この場合も、「分類語彙表」は、少数点以下のけたで対応はとってあるので、小数点以下だけでみれば同じグループとみることできる。

例5.2 分類語彙表

同じ<3.112相互・異同…<相の類

違う<2.112異同…<用の類

E D R の概念体系

106644[*同じでないこと]<30f9a5[*相違]…<3aa963[*状態]

1e84b4[*同じであること]<30f9a5[*相違]…<3aa963[*状態]

6. 開発過程で出てきた問題 ---言語の側面から

(1)属性概念で修飾された概念

3(5)複合語の扱い とも重複するが、下記のような例は文中の構文的機能により、[状態]にも[主体]にもなる。

例6.1 美人---3f9619[#美しい女性]

単語見出し「美人」に [#美しい女性]と[#女性が美しいこと]の二つの概念があれば問題ないが、一般的な辞書でも、属性概念で修飾された単語の概念は、属性概念だけの概念を別立てにしていない。本体系では、このような部分を中間ノードの多重リンクで扱った。

例6.2

3f9619[#美しい女性]<3f961a[*容姿の美醜でとらえた人間]< | 30f6b0[*人間]
| 3aa961[*人間の属性]

(2)設定した中間ノードにまたがる概念

「#男女、#売買」のように、実際にたてた中間ノードを越えたような概念が出てくる。文献4では、この関係をあわせよび関係とよんでいる。例6.3に示すように、[#売り買いする][*売る][*買う]は、それぞれ間接的に[*所有権の移動]の下位概念であるが、[#売買]と[#売る]あるいは[#売買]と[#買う]の間に上位関係はつけない。

例6.3 30f82a[*買う]<3f96b4[*所有権の主体への移動]<30f826[*所有権の移動]

30f82a[*売る]<3f96d8[*所有権の主体からの離脱]<30f826[*所有権の移動]

3cfe33[*売り買いする]<0e61d1[*取引]<30f826[*所有権の移動]

7. 今後の課題

E D R の概念を概念体系上に配置するために、概念体系の作成と概念体系の中間ノードに概念を分類する作業を進めてきたが、今後は、中間ノードの全体的な調整を中心に概念体系の改良を進めていきたい。

また、E D R には、具体的な文を概念関係表現に置き換えたデータが蓄積されている。これらの実例に即した概念記述から、概念体系の継承の妥当性を調べ、改良を進めたい。

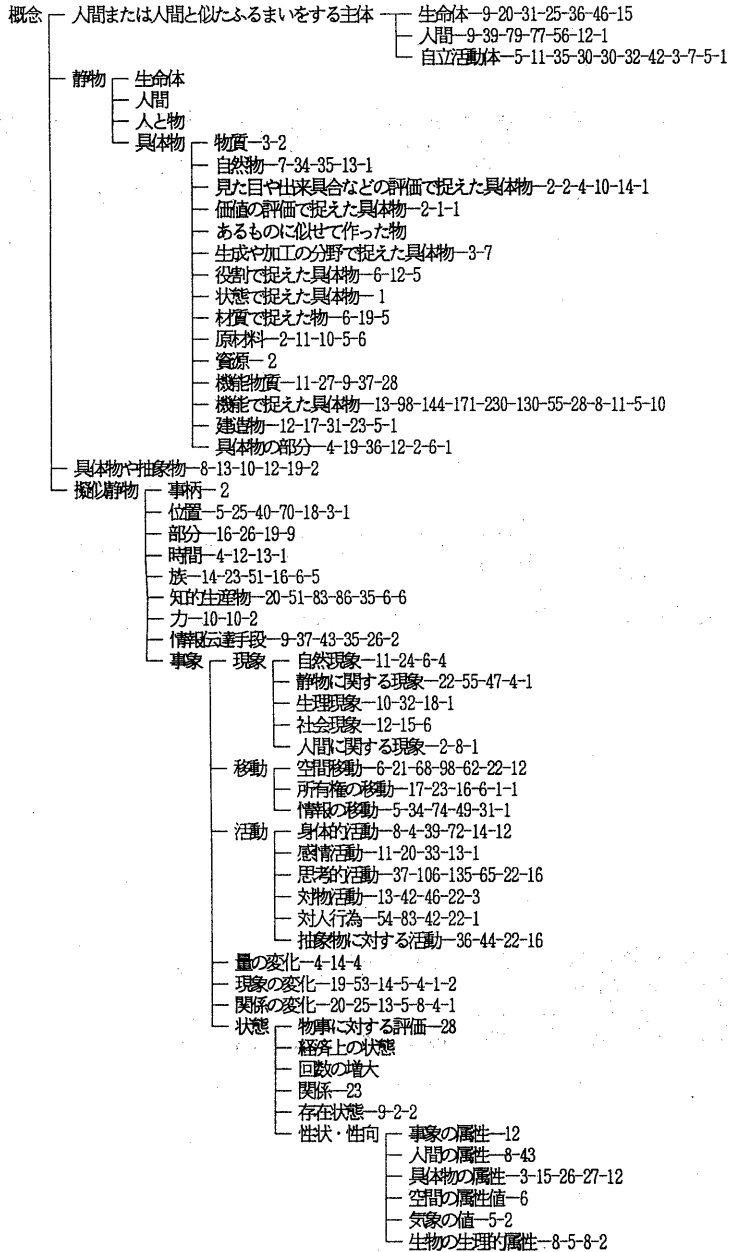
最後に、このような概念体系の研究にかかわる機会と適切な研究指導をいただきました、E D R 所長横井俊夫氏、第一研究室元室長内田裕士氏、現室長杉山健司氏、および、実際の分類作業に従事していただきました方々に、謝意を表します。

文献

- 1)日本電子化辞書研究所TR007(1988)「概念辞書第一版」
- 2)日本電子化辞書研究所TR012(1989)「概念辞書第二版」
- 3)日本電子化辞書研究所TR020(1990)「概念辞書第三版」
- 4)荻野綱男(1993)「現代日本語名詞ソーラスからみた語彙の分類」(平成4年度科学研究費研究成果報告書)
- 5)国立国語研究所(1964)「分類語彙表」(国立国語研究所資料集6) 秀英出版
- 6)大野晋・浜西正人(1981)「角川類語新辞典」角川書店
- 7)日本電子化辞書研究所(1993)「E D R 電子化辞書仕様説明書」

資料 1 概念体系の上位ノードと下位ノードの展開状況

凡例：分類項目→下位 (2) レベルの中間ノード数→3レベルの中間ノード数→...



資料 2 概念体系の細部展開

3f96d8 ● 2 0 2 2 1 0 | 所有権の主体からの離脱
 30f82a ● 2 0 2 2 1 3 | 売る | 売る
 #3c3f99 ● 品物売る
 @444b8f ● 販売のやり方でとらえた販売
 #3d1a80 ● 値段を安くして売る
 @444b90 ● 品物で捉えた販売

概念体系エディタ

辞書出力 統合 編集 Loop Check No Check 上位追加 キャンセル 終了

< >

【3f96d8】 2 0 2 2 1 0 所有 権の主体から	【100c7e】 売買 の取り引きをする				
--	-------------------------	--	--	--	--

【30f82a】 2 0 2 2 1 3 | 売る

1/2
1/114

【104996】 売る 約束をする	【105374】 売る	【3bfcaa】 売春 すること	【3bfcaa】 売春 すること	【3c3f99】 品物 を売る	【3e73ac】 the act of sending products out
----------------------	-------------	---------------------	---------------------	--------------------	--

★

概念体系エディタ

辞書出力 統合 編集 Loop Check No Check 上位追加 キャンセル 終了

< >

【30f82a】 2 0 2 2 1 3 売る					
------------------------------	--	--	--	--	--

★

【3c3f99】 品物売る

1/1
1/4

【444b90】 品物 で捉えた販売	【444b8f】 販売 のやり方からみ た販売	【3d0613】 物を 販売する	【3cfd8c】 品物 を売りさばく		
-----------------------	-------------------------------	---------------------	-----------------------	--	--

●

概念体系エディタ

辞書出力 統合 編集 Loop Check No Check 上位追加 キャンセル 終了

< >

【3c3f99】 品物 を売る					
--------------------	--	--	--	--	--

●

【444b90】 品物で捉えた販売

1/1
13/34

【1ee999】 作家 が自分の著作物 に署名して	【1ea9c2】 塩を 売り歩くこと	【1e9618】 水を 売ること	【1e898c】 家屋 を造作つきで売 ること	【10c3d8】 山を 売ること	【107bff5】 土地 や家屋などをい くつかに分ける
---------------------------------	-----------------------	---------------------	-------------------------------	---------------------	------------------------------------

< >