

## 単語間情報とランダム性を利用した 創造性を有する概念ネットワークの自動生成

松本芳樹 小磯拓也 乾伸雄 小谷善行

東京農工大学 工学教育部

「コンピュータに創造性を持たせるにはランダム性が一つの要素である」という仮説に基づき、概念ネットワークを自動生成する。「創造性」を、ある程度の制限を与えた上での概念の組み合わせの中から作られる新たな概念関係と定義する。本研究において、ランダム性を次の3つの処理で適用する：(1)どのような意味で単語を用いるか、(2)どの二つの概念を関連付けるか、(3)どのような関係を概念間に割り付けるか。ランダム性だけで構築された概念ネットワークは、意味や正当性が欠如したものとなるため、単語間情報による制限を加える。単語間情報として、本研究では特に名詞間関係辞書とパターンによる制約を用いる。名詞間関係辞書は複合語や名詞句「A の B」表現についての関連研究に基づき作成した。概念ネットワークは、IPAL 辞書より得られる意味素性、EDR 辞書より得られる概念関係子を用いた三つ組み表現を構造の最小単位として構成される。

### Generating Creative Conceptual Networks by using Words and Randomness.

Yoshiki MATSUMOTO Takuya KOISO Nobuo INUI Yoshiyuki KOTANI

The Graduate School of Technology, Tokyo University of Agriculture and Technology

This paper describes a method of generating creative conceptual networks by using information of words and randomness. We propose a hypothesis that randomness is a factor which gives creativity to computers. The creativity is a conceptual relation randomly generated from various combinations of concepts under conditions that man understands them. A conceptual network contains triplets (two concepts and one relation). Our system randomly generates a triplet which is constrained by linguistic semantic knowledge. In this paper, we describe the generation method, dictionaries used and the reconstruction method of conceptual networks. Experimental results showed that triplets for verb-verb relations and noun-verb relation are feasible in the interest and the adequateness, and conceptual networks made by our system are interesting one.

#### 1. はじめに

自然言語処理の分野をはじめとして、文や事象の関係を表現する方法として、概念ネットワークが用いられる。これは事象とその間の関係を抽象的に表現したものである。過去の研究では文を解析し、その構造をネットワークとして表現する意味解析の研究や各事象の間の関係を解明する関連付けの研究が数多く行われている[4][5]。本研究では概念ネットワークを新たな意味の生成のために用いる。意味解析の分野では、与えられた現象や表現を解析した結果として、概念ネットワークを生成する。これに対し、本研究では、意味生成の方法として、ランダムに選択された概念や関係に制約や情報を適用することで、概念ネットワークを生成し、新たな現象や関係コンピュータによって作り出すことを目的とする。特に「創造性」に着目し、コンピュータが生成した概念ネットワークを評価する。このような研究は過去には見られず、コンピュータによる創造性の発現メカニズムの研究に寄与したいと考えている。

本研究で生成された概念ネットワークは、人間が文章生成をする際の参考として用いられることを期待している。ランダムに生成される概念ネットワークは次の要素によって制約される。

- 名詞間関係辞書(本研究で構築)
- 用言が持つ格フレームと任意格の制限
- 概念ネットワークのパターンによる制約(本研究で定義)

本研究では、ランダム性は創造性を概念ネットワークに持たせるために導入され、以下の場面で適用される。

- 事象(単語)の意味選択：各単語をどういう意味で扱うかの選択
- 関係の選択：事象(単語)間をどのような関係で結びつけるかの選択
- 関係付ける事象(単語)の選択：どの事象とどの事象を関係付けるかの選択

これらの場面でランダム性は、人間のひらめきや創造性に相当するメカニズムとして機能すると考える。しかし、ランダム性だけで生成された概念ネットワークは人間から見て正当性のないものとなる可能性が高い。これは生成される概念ネットワークを理解する機構が存在しないためであるが、意味的な側面よりこの問題を解決するために単語間情報によって生成される概念ネットワークを制約する。本研究では単語間情報とランダム性を利用して概念ネットワークを生成する手法を提案する。

## 2. 本研究における概念ネットワークの定義

図1に概念ネットワークとその表層表現例を示す。これは[5]で示された概念ネットワークを参考に作成したものである。

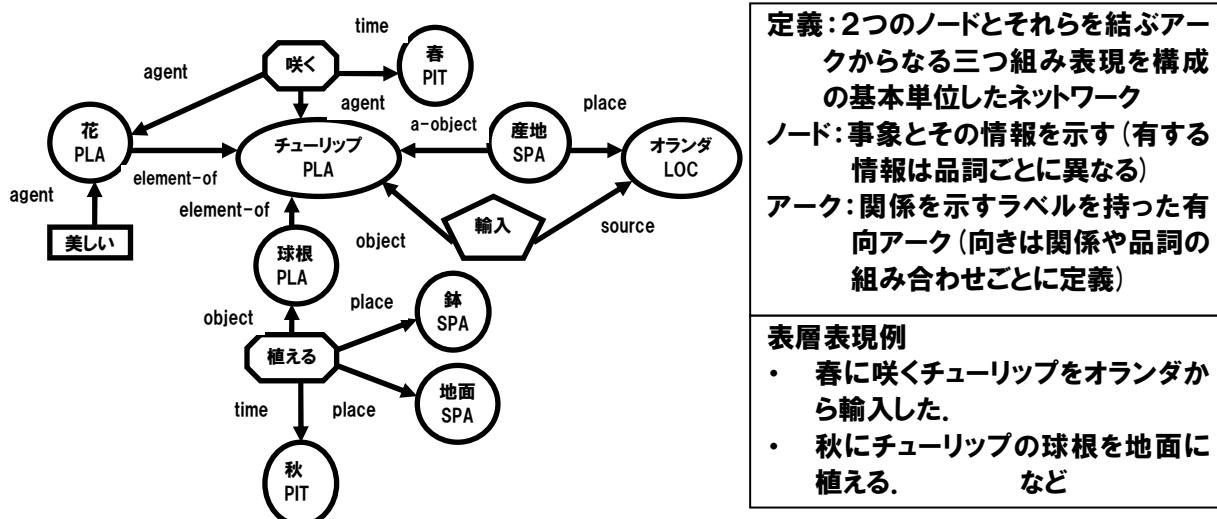


図1 本研究で生成する概念ネットワークの例

## 3. 概念ネットワーク生成で用いる情報

本研究では概念を単語によって表現し、ノードに記述される要素となる。表1に概念で表現される情報の概要を示す。

表1 品詞の利用情報

品詞	生成時に利用する情報
名詞	表層表現, 品詞分類, 意味素性(概念)
動詞・サ変動詞・形容詞	表層表現, 品詞分類, 格フレーム(必須格), 格の出現確率(任意格)

表1で格の出現確率は意味素性ごとにコーパスから得たものを利用し、任意格の推定や未知語の格推定に利用する。また、表層格として扱う格を以下に示す。

動詞・サ変動詞が持つ格(任意格含む): ガ・ニ・ヲ・デ・ヨリ・カラ・マデ・ヘ・ト(9種)

形容詞が持つ格(任意格含む): ガ・ニ・ヲ(3種)

#### 4. ランダム性の利用と役割

本研究では、ランダム性を用いて創造性を持つ概念ネットワークの生成を行う。これは次のような仮説に基づいている。

**仮説：**コンピュータに創造性を持たせるにはランダム性が一つの要素である  
また、本研究では創造性を次のように考える。

##### 創造性

- 表層と程度(頻度)にとらわれない関係構築で面白みを含んでいるもの
- 既存のデータ(コーパス)から得られにくい新たな表現が可能であるもの
- 主観的・状況的なものであり、絶対的なものではない

出現頻度(コーパスから取得される種々の統計的情報)の重要性を高く扱う場合はコーパスに基づいた概念ネットワークの生成、低く扱った場合はランダム性に基づいた概念ネットワークの生成と考えることができ、これらは下記の長所短所が考えられる。両者は背反するものであるが、これらをバランス良く扱うことで創造性を持った概念ネットワークが生成できると考える。そのためのランダム性であり、またこれによりコーパスから得られない既存以外の組み合わせも生成できると考える。

出現頻度の重要性	長所	短所
高い	関係が構築しやすい 堅実な表現が可能	面白み・目新しさが無い
低い	創造性・面白み・ 目新しさを含む	理解が困難であったり、矛盾した関係が存在する。

#### 5. 概念関係子と意味素性の利用

本研究ではノード間の関係を EDR 電子辞書[1]で定義された概念関係子(25種+4種)を用いて表現する。本研究の性質のため一部変更・追加を行っている。主な変更点として、概念関係子の解釈の変更、アークの向きとの関連化、関係表現の不足を補うための関係子の追加などである。名詞の意味記述のために、本研究では IPAL 辞書[2]で定義された意味素性(第1階層6種、第2階層52種)を用いる。名詞である単語は1つ以上の意味素性を有している。本節では概念ネットワークの生成方法について述べる。

##### 5. 1. ノードおよび関係生成

概念ネットワークは複数の単語がノードにより構成される。これらのノードおよびノード間の関係は以下の手順で生成される。

- 1) 概念ネットワークが存在しない場合(未生成の場合)は単語1および単語2を辞書からランダムで選択
- 2) すでに概念ネットワークが存在する場合は単語1を辞書から、単語2を概念ネットワークの中からランダムで選択
- 3) 単語1と単語2の間の関係を品詞の組み合わせごとに生成する。個々の関係の生成は次節から述べる手法により行われる。

##### 5. 2. 名詞間の関係生成

- 1) 名詞である単語1・2をそれぞれノード1・2とし、可能な意味素性を抽出する
- 2) 各ノードの意味素性をランダムで一つに決定
- 3) ノード1・2の意味素性の組み合わせに対応した概念関係子の候補を名詞間関係辞書から取得
- 4) 取得した概念関係子の候補の中からランダムで概念関係子を決定
- 5) 決定した概念関係子に基づいて、アークの向きを決定

##### 5. 3. 名詞-用言間の関係生成

- 1) ノード1を名詞、ノード2を用言とし、ノード1の意味素性を抽出、ノード2からは格フ

- レームを抽出
- 2) ノード1が持つ意味素性の中からランダムで意味素性を決定
  - 3) ノード1の意味素性がノード2の格フレーム内の格と一致するかチェックする
  - 4) 一致したら格フレームで規定された概念関係子により、名詞・用言間を関係付ける。一文一格の原則から使用された格を格フレームから消去
  - 5) 一致する格が存在しないならば、ノード1の持つ意味素性をランダムに選び直し、3の処理に戻る
  - 6) 名詞・用言間の関係が生成されたら終了(この時点で生成されれば必須格に割り付けられたことになる)
  - 7) 名詞・用言間の関係が生成されなければ、以下の手順で任意格として関係生成を行う
  - 8) ノード1の持つ意味素性からランダムで意味素性を決定
  - 9) 名詞の意味素性が時間であるかのチェック
  - 10) 時間の意味素性を持つ場合は、時間としての関係生成を行う(表2で示したいずれかの関係を生成する)
  - 11) 時間の意味素性でないならば、意味素性ごとに得られたコーパスからの関係構築頻度を用い、確率的に概念関係子を決定
  - 12) コーパスから得られた頻度が0ならば、選択した意味素性の第1階層に意味素性を変更し、11の処理を再度行う
  - 13) 概念関係子が決定し、関係が構築できたら終了
  - 14) 第1階層の意味素性でも関係が構築できない場合は、ノード間の関係を生成せずに終了

**表2 時間としての関係を示す概念関係子([1]より)**

概念関係子	関係	表層表現例
time	事象の起こる時間	明日行く
time-from	事象の始まる時間	今日から始まる
time-to	事象の終わる時間	来年まで延びる

#### 5. 4. 用言・用言間の関係生成

- 1) 単語1・2をそれぞれノード1・2に当てはめる
- 2) ノード1・2の組み合わせによって、概念関係子を候補(表3参照)の中からランダムで選択し、終了

**表3 用言間で構築できる概念関係子の候補**

品詞の組み合わせ	概念関係子の候補
動詞-動詞, サ変-動詞, サ変-サ変	cooccurrence, purpose, sequence, cause
形容詞-動詞, 形容詞-サ変, 形容詞-形容詞	cooccurrence, purpose, sequence, cause, manner, condition

ここまでで生成された概念ネットワークに対し、正当性を高めるため6章で述べるパターンによる制約を適用する。また、生成手法でランダム性の利用を表4に示す。

**表4 品詞間の関係生成におけるランダム性の利用点**

名詞-名詞	名詞-用言	用言-用言
単語(名詞)の選択 関係付けるノードの選択(概念ネットワーク中から) 意味素性の選択 概念関係子の選択	単語(動詞・サ変動詞・形容詞)の選択 関係付けるノードの選択(概念ネットワーク中から) 意味素性の選択 概念関係子の選択	単語(動詞・サ変動詞・形容詞)の選択 関係付けるノードの選択(概念ネットワーク中から) 概念関係子の選択

#### 5. 5. 任意格の推定

5. 3で述べたように格フレームによる必須格としての関係構築が失敗した場合は任意格とし

て構築を行う。任意格の構築はコーパスからの統計情報に基づいて行う。ここでは 2 つの辞書の情報を統一化し、表層格、意味素性、概念関係子の対応化を行う。辞書(コーパス)から得られる組はそれぞれ以下の通りである。

**EDR 電子辞書から生成される組の例**

{概念関係子, 表層格} : {agent, ガ}, {goal, へ}

**IPAL 辞書から生成される組の例**

{意味素性, 表層格} : {HUM, ガ}, {DIV, ヲ}, {ABS, ト}, {LIN, デ}

辞書内の各用言の単語に対して組を集計する。まず、IPAL 辞書より得られたデータを用いて任意格の推定を行う。概念関係子の構築確率は以下の通りである。

$$P_{pos}(I, R) = \frac{N_{pos}(I, R)}{N_{pos}(I)}$$

$I$ : 任意の意味素性(全58種)

$R$ : 任意の関係(概念関係子, または表層格)

$pos$ : 任意の品詞(動詞, サ変, 形容詞のいずれか)

$P_{pos}(I, R)$ : 品詞 $pos$ において意味素性 $I$ が関係 $R$ を構築する確率

$N_{pos}(I, R)$ : 品詞 $pos$ において意味素性 $I$ が関係 $R$ を構築した数

$N_{pos}(I)$ : 品詞 $pos$ において意味素性 $I$ が関係を構築した総数

実際に扱うデータの例を以下に示す。  $F_{pos}(I)$ を品詞  $pos$  における意味素性  $I$  の関係構築頻度の値とする。

$F_{verb}(\text{意味素性}) = [\text{ガ}, \text{ニ}, \text{ヲ}, \text{デ}, \text{ヨリ}, \text{カラ}, \text{ヘ}, \text{ト}]$ (動詞)

$F_{verb}(\text{HUM}) = [289, 58, 39, 0, 0, 0, 9, 0]$

$F_{verb}(\text{ANI}) = [183, 88, 44, 0, 0, 6, 0, 0]$

次に表層格と概念関係子の対応化を行う。表層格が示す概念関係子は複数存在する。そこで EDR 辞書より得られた組の統計より、表層格が最も表す可能性が高い概念関係子を選択する。

**6. 概念ネットワーク生成への制約と適用パターン**

5 章で生成された概念ネットワークの正当性を高める方法として制約を用いる。制約は生成が終了した概念ネットワークに対して適用され、概念ネットワークを再構築する。パターンに関連した制約の機能は連結・推論・切断・変化の 4 つに分類される。このパターンによる制約は概念ネットワークが再構築され、変化が起こると再帰的に適用される。制約は全部で 18 種(連結・推論 11 種, 切断・変化 7 種)存在し、ネットワークパターンは 6 種定義してある。以下に一部を示す。

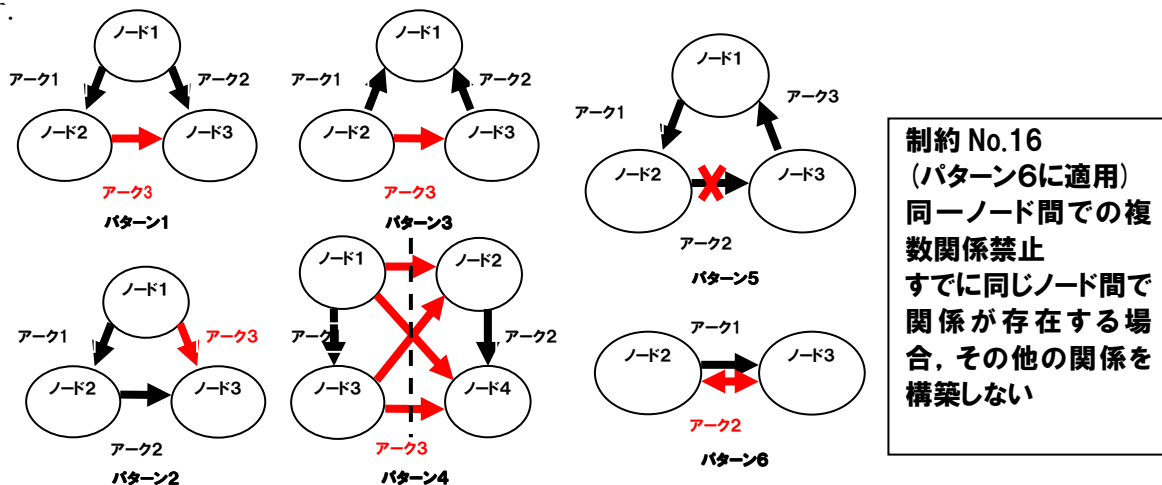


図2 ネットワークパターンと制約の例

## 7. 名詞間関係辞書の構築

本研究において名詞・名詞間を関係付けるために名詞間関係辞書を構築した。これは名詞間に成り立つ関係を示したものである。この辞書は、コーパスの統計からは名詞間の関係を特定できないため必要となる。名詞間関係辞書では、先に述べた概念関係子と意味素性を用いて意味素性ごとに関係が成り立つ関係子の候補が列挙されている。本辞書は以下の3つの手法を組み合わせで作成した。

- 複合名詞の解析手法
- 名詞句「AのB」表現の解析手法
- コーパスからの関係推定(人手)

まず、複合語の生成や解析の手法を元に単語の意味とその間に成り立ちうる関係を決定した。利用した複合語生成の過去研究[3]ではサ変動詞と名詞に関連した生成規則が提案されている。本研究では[3]で提案されている規則を統一して、次のように扱った。「サ変動詞に由来する名詞はACT(動作)の意味素性を持つと考え、これは格フレームと同様の関係を構築可能である」。つまり、ACTの意味素性を持つ名詞は格構造を持ち、用言と同様の関係構築が可能であることを示す。

次に名詞句「AのB」表現の解析手法とその利用法について述べる。ここでは分類の意味に従って概念関係子を当てはめた。本研究では意味素性を扱うので、意味の定義はこれに準ずる。表5で示した関係と意味をそれぞれ本研究では概念関係子と意味素性に置き換えて利用する。

表5 意味的關係の例と規則([4]のものを一部省略)

関係		A	B
AとBが同格	人名を含む同格	固有名詞以外かつ人	固有名詞かつ人
	地名を含む同格	固有名詞以外かつ地名	固有名詞かつ地名
Aが場所を表す		場所	
時間的關係を表す		時間(A, Bのどちらか片方または両方)	
BがAの所属		組織	人
AがBの種類		部分	
AがBの材質		材質	

以上の様ないくつかの過去研究を参考に辞書を作成したが、名詞句・複合語解析や統計などからは得られない(扱われていない)関係に関してはコーパスから人手で抽出した。このように作成された名詞間関係辞書にはすべての意味素性の組み合わせ(58種×58種(第1階層6種, 第2階層52種))に対する概念関係子が記載されている。記載例を以下に示す。

### 名詞間関係辞書の構成(一部)

組み合わせ番号	組み合わせ(第1階層)	意味素性(A)	意味素性(B)	概念関係子候補
100100	ANI:ANI	ANI	ANI	possessor, equal, part-of, sup-sub, element-of
100101	ANI:ANI	ANI	GAT	possessor, equal
100102	ANI:ANI	ANI	AML	possessor, equal
100103	ANI:ANI	ANI	HUM	possessor, equal
100104	ANI:CON	ANI	CON	possessor
...				

名詞間関係辞書を使うことで、名詞間の関係を規定できる。その例を以下に示す。ここで概念関係子「place」を選んだ場合、「愛媛のみかん(愛媛←みかん)」のような表層表現ができる。

#### 名詞「愛媛」と名詞「みかん」の場合(意味素性はすでに決まっているとする)

名詞A 愛媛 意味素性 LOC 番号 301  
 名詞B みかん 意味素性 PLA 番号 407  
 組み合わせ番号 301407  
 構築可能な関係 place, source

## 8. 評価実験

実験設定と評価方法は以下の通りである。生成結果例と評価の集計結果は図3・図4と表6に

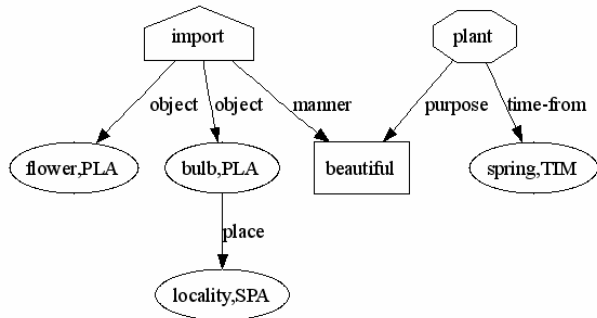
示す。

**実験設定**

- ・ ノードとなる単語を辞書(または概念ネットワーク)からランダム選択
- ・ 10回の実行で1つの概念ネットワークを生成(1回の実行で1つの三つ組み表現を生成)
- ・ 辞書はテーマごとに単語を制限したものを利用

**評価方法**

- ・ アンケート形式
- ・ 比較対象：図1の概念ネットワーク
- ・ 概念ネットワークの基本要素である三つ組み表現と概念ネットワーク全体に対して行う



**生成結果と表層表現例**

産地の球根と花を美しく輸入する  
美しくするために春から植える etc...

図3 生成結果と表層表現例

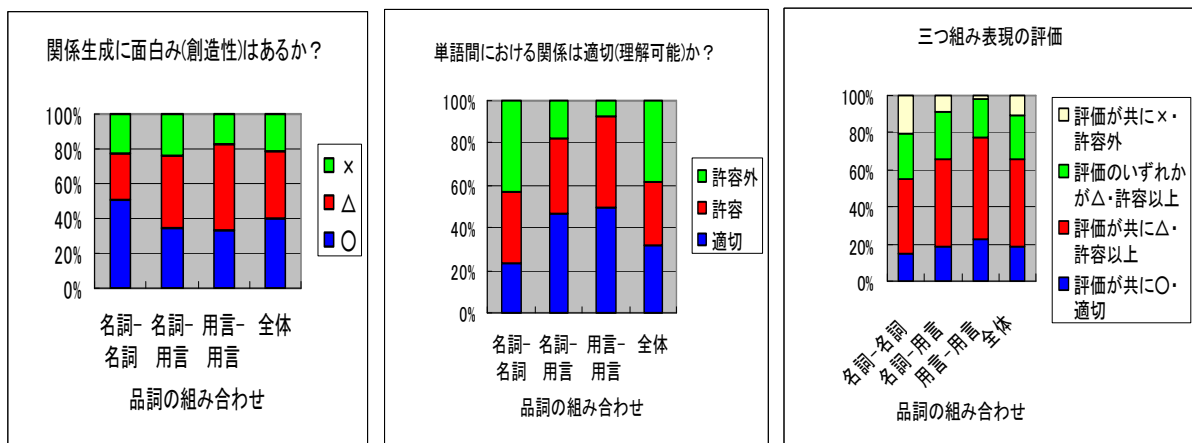


図4 三つ組み表現の評価結果

表6 概念ネットワークの評価結果

評価項目	評価 (得点)	概念ネットワーク(回答数)				合計数
		「例題の単語のみ」	制限辞書「動物」利用	制限辞書「植物」利用	制限辞書「生活」利用	
ネットワークとしての偏りはあるか?	1	0	0	0	0	0
	2	2	1	5	3	11
	3	1	6	6	2	15
	4	3	4	1	5	13
	5	6	0	0	2	8
評価平均		4.1	3.3	2.7	3.5	3.4
ネットワークに面白み(創造性)はあるか?	1	0	0	0	0	0
	2	2	1	3	4	10
	3	5	5	6	6	22
	4	4	4	2	2	12
	5	1	1	1	0	3
評価平均		3.3	3.5	3.1	2.8	3.2

ネットワークが 表す意味の 理解度は？	1	0	1	0	2	3
	2	2	5	5	3	15
	3	4	2	3	2	11
	4	3	3	1	2	8
	5	3	0	3	2	8
評価平均		3.6	2.6	3.2	2.9	3.2

## 10. 考察

名詞-名詞間の三つ組み関係について、面白さ(創造性)に関しては77%が○または△と評価している。面白さ(創造性)は、評価する人の主観を含むものであるが、この結果は関係生成においてランダム性が有効に働いていると評価できる。特に○と評価した結果が50%を超えているのは注目される。これに対し、適切さの評価に関しては57%が適切または許容と言う評価をしている。創造性と理解の容易さのバランスは十分でないと言える。適切さの評価が高い内理由としては、名詞間関係辞書が十分表現されていないことが挙げられ、辞書の改良によって適切な三つ組み関係生成することが可能であると考えている。

名詞-用言間と用言-用言間の生成結果は、面白さについて共に80%程度が○または△と評価している。これは名詞-名詞間と同程度の評価である。さらに適切さの評価は80%以上が適切または許容と評価している。創造性・適切さともに高い評価を得ている。ここでは、必須格の情報及び任意格の推定方法が適切であったとかが得られる。格フレームだけでは既知の関係しか構築できないが、5.5章で述べた任意格推定の統計により、コーパスから得られない名詞と動詞の間で適切な関係生成が実現できた。

表7により、概念ネットワークを生成する際に用いた辞書による評価に大きな違いは生まれなかった。面白みの評価は平均的に3よりやや高い値となっている。これは先の三つ組み関係における面白みの評価よりも低い値となっている。これは、概念ネットワークの一部で面白み(創造性)を表現することと全体で面白みを表現することは別であり、概念ネットワーク全体を調整する制約の導入が必要であることを示唆している。問題となるのは、部分的には理解できても全体として理解しにくい傾向が見られることである。特に、特定の三つ組み関係が理解できない場合、概念ネットワーク全体が意味不明なものになってしまう。

## 11. まとめ

本研究では仮説に基づいてランダム性による創造性を概念ネットワーク生成で表現させた。その手法として、名詞間関係辞書・ネットワークパターンによる制約を定義し、概念ネットワークの生成手法を提案した。三つ組み関係についてはいずれ品詞の組み合わせの生成においても正当性・創造性を含んだ生成を実現した。概念ネットワーク全体では比較対象よりもやや高い評価を得た。概念ネットワークの生成を通して、創造性に関連したランダムの可能性・有効性を導いた。

## 参考文献

- [1]日本電子化辞書研究所：EDR 電子化辞書'95
- [2]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辞書 IPAL(1995)
- [3]李 泰憲，安藤 一秋，大野 将樹，青江 順一：複合語規則を用いたキーワード導出手法，電子情報通信学会誌，2001/5 Vol. J84-D-II No.5
- [4]池原 悟，村上 仁一，宮本 健司：「AのB」型名詞句の日英翻訳規則について，情報処理学会論文誌，2002/7 Vol.43 No.7
- [5]長尾 真：自然言語処理，岩波書店，岩波講座ソフトウェア科学 15
- [6]横山 昌一，加藤 貴子，横重 拓司：格助詞「の」の分類と解析，情報処理学会第1回年次大会，pp.137-140(1995)
- [7]荒木 健司，桃内 佳雄：ヒューリスティックを用いた意味ネットワークの自動生成，情報処理学会，信学技報，NLC94-10，1994-07
- [8]竹内孔一：複合名詞解析のための語彙概念構造付与作業の仕様書 ver1.01,2003