

## 生成語彙による合金研究論文からの知識抽出

中挾 知延子  
東洋大学国際地域学部

星本 健一  
科学技術庁金属材料技術研究所

本稿では、Pustejovsky の生成語彙の考えに基づいた記述方法による、論文文章の意味表現を試みる。論文の意味表現として、複数の観点から語の意味を記述することや、意味中心的に記述する方法が必要であるとの考えから研究をすすめている。対象とする論文の分野は、共同研究者の専門分野である合金研究に設定している。意味表現を行う目的は、それらに適切な演算を行って、論文間の内容の差異や矛盾点を抽出したり、新たに入力した論文と既存の論文との知識構造を統合することである。意味表現は、論文文章の特徴に適合する形で生成語彙の記述方法に多少変更を加えている。

## Knowledge Extraction from Technical Papers of Metallic Materials based on the Generative Lexicon

Chieko Nakabasami

Department of Regional Development Studies, Toyo University

Kenichi Hoshimoto

National Research Institute for Metals

In this paper we try to describe semantic representations of technical papers based on Pustejovsky's THE GENERATIVE LEXICON. We claim that it is necessary for semantic representations of technical papers to describe a word's meaning from various points of view and to put emphasis on the semantic aspect of words. Technical papers concerned with metallic materials are focused on because this research is proceeded in cooperation with an expert on the metal engineering. Our purpose for making semantic representations is (1) to detect differences and inconsistencies among papers by applying appropriate mechanisms to them, (2) to integrate new contents of another paper into existing contents of papers. The semantic representations of technical papers based on the generative lexicon are modified in order to adjust to characteristics of papers.

## 1. はじめに

我々は、文章の意味解析を行うためには、文章中に現れる語の意味を詳細に記述し、それらの記述から適切な演算を行って文章全体の意味をつかむことが必要であると考えている。意味記述については、従来さまざまな方法が提案されている [1,2]。それらの多くは、統語構造 (syntactic structure) と意味構造 (semantic structure) との統合をねらいとしたものである。これらの記述は同義な文章間の語の統語的なふるまいを説明することや、統語解析における多義性解消などに用いられる場合がほとんどであり、文章そのものの意味解析への適用はまだ実現されていない。

本研究では、意味中心の考え方に沿った語彙意味論の立場から、Pustejovsky が提案する生成語彙 [3] の考えを採用入れて論文の意味記述を行う。語彙意味論の立場としては、他に Jackendoff の語彙概念構造 [4] がある。

本研究の目的は、論文に記述されている内容を知識構造として抽出し、複数の論文間での内容の差異や矛盾点の抽出を実現すること、さらに新規の文章が入力として加えられた時にその内容を既存の知識構造に適切に統合することである。そのための第一段階として、共同研究者の専門分野である合金研究についての論文を対象にして意味記述の構築をすすめている。

論文文章を対象にした理由としては、

- 1) 出現する語は専門用語が多く、用語についての多義性が一般の文章に比

べて少ない。

- 2) 文章構造が明確であり、類似のテーマについて書かれた論文の間では、内容の対応が取りやすい。

などのことから、意味記述の手がかりを得るための第1段階として適していると考えられるからである。論文文章の意味表現には、生成語彙の意味表現手法をそのまま持ち込むのではなく、論文の内容上の特徴をつかみ易いように多少変更している。

論文の解析に関する研究に [5] がある。[5]でも論文間類似性の解析を行っているが、類似度の説明として項目の抽出にとどまっており、「どのような点のどのような内容で異なっているのか」といった内容の抽出はなされていない。また、論文間の参照情報を考慮して参照の目的を解析することで論文間の関係を明らかにしようとする研究 [6] があるが、論文間の関係の内容に立ち入った解析は行われていない。

本稿では合金研究で行われる実験の一つである tensile test (引張り試験) について、専門知識を整理し意味記述を行う。論文は英文で書かれたものを採用している。以下各章では、生成語彙を論文の意味記述へ適用するにあたっての有効性について説明し、作成した論文文章の意味記述の内容について述べる。これ以降何も断らない限り、論文とは合金研究に関する論文を指す。

## 2. 生成語彙に基づく意味記述

生成語彙における意味表現の種類と、

<Levels of Representation>	
Argument Structure	統語的に実現される項構造
Event Structure	語が表す事象とその時間的関係
Qualia Structure	種々の観点・役割からの意味内容
Constitutive	語の構成要素
Formal	語の静的な属性, 語が表す対象の客観的な説明
Telic	語が示す概念の目的・機能を示す動作
Agentive	語が示す概念を産み出す動作・原因
Lexical Inheritance Structure	語の概念間の階層構造
<Generative Operations>	
Type Coercion	関数の引数のタイプを予測するタイプへ変更
Selective Binding	引数が自分を支配する述語を選択
Co-Composition	修飾語と被修飾語のタイプが合わない場合, 被修飾語のQualia Structureから修飾語のタイプに合う役割を探し, 修飾する操作

図1 生成語彙の意味構造・生成演算の概要

それらの内容に施される生成的な演算の概略を図1に示す[7].

図1で示した生成語彙の意味記述手法を文章の内容を表現する際に適用する理由として以下のものがあげられる.

(1) 複数の品詞を意味的な側面から統合的に扱える.

語には, 品詞は異なるものの, 基本的な概念は共通している語が複数ある. このような語の集まりに対して, 品詞別に意味構造を用意するのではなく, 単一の意味構造の中に融合させておく方が簡潔に記述できる.

例えば, test という語には, 名詞と動詞の用法がある. どちらの場合にも中心になる概念は同じであるので, 本研究では, test を意味構造のエントリーとして, Event Structure に動詞「試験する」としての test を記述し, Argument Struc-

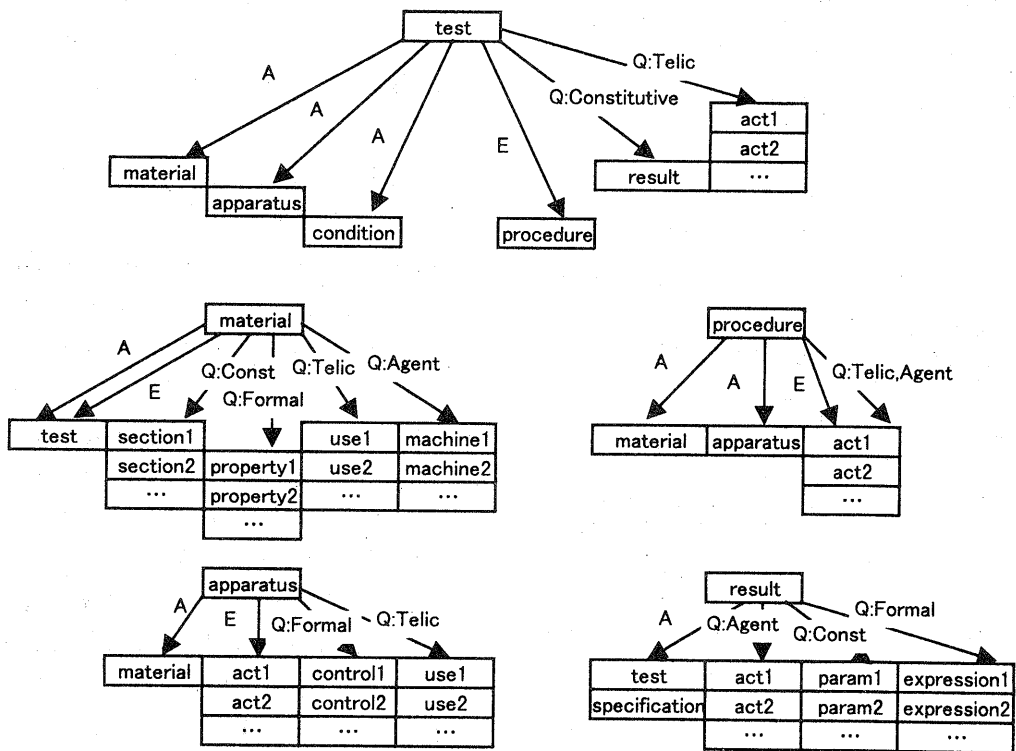
ture に名詞「試験」としての test を記述する. 実際の論文中にも以下に示すように, test という概念がさまざまな品詞や形で現れている.

- Tensile tests were performed between room temperature and 760 C.
  - Most important was the fact that every gauge length was free of extraneous deformation at the start of testing.
  - These specimens were tested in tensile creep under constant stress at temperatures of 800°C and 950°C.
- (2) 文章中で一つの対象についての複数の観点からの内容が, 生成語彙の Qualia Structure における多様な記述方法によって表現できる.

論文において試験方法を記述している文章の大部分は,

- 装置(apparatus)
- 金属材料(material)
- 条件(condition)
- 実験過程(procedure)
- 実験結果(result)

で構成されている. 本研究では, 図1で示した生成語彙の4つの structure で記述される意味役割を, これらの内容が test というテーマに対して担っている役割に合わせる形で変更を加えた. その結果を図2に示す. 図2の test, apparatus, material, procedure, result は意味タイプを表している. 以下 test タイプとも呼ぶ場合がある.



Argument Structure: testを行うために必要なもの

Event Structure: testを行う時に生じる動作

Qualia Structure

Constitutive: その対象を構成する部分, testにより得られる結果

Formal: その対象の特性, 結果を導く式

Telic: testを行う目的, その対象を用いる目的

Agentive: testの結果を導く動作, その対象を生じる動作

Lexical Inheritance Structure: その対象の上位概念

図2 試験方法についての意味記述

### 3. tensile test の意味構造

合金研究に関する文献[8,9,10]での tensile test の説明から, 図2で示された記述方法に従って, 実際に使われている語を参考にして tensile test の意味構造を図3, 4のように記述した. 図中の四角で囲まれている語は意味のタイプを表す. 以下に図2, 3, 4の内容を併せて

説明する.

#### (1) Argument Structure(ARGSTR)

対象(ここでは試験方法)が存在する時に必要な意味タイプを指す. tensile test ならば, テストを行うために物質の標本(specimen)と実験装置(apparatus)が ARGSTR で関係づけられる.

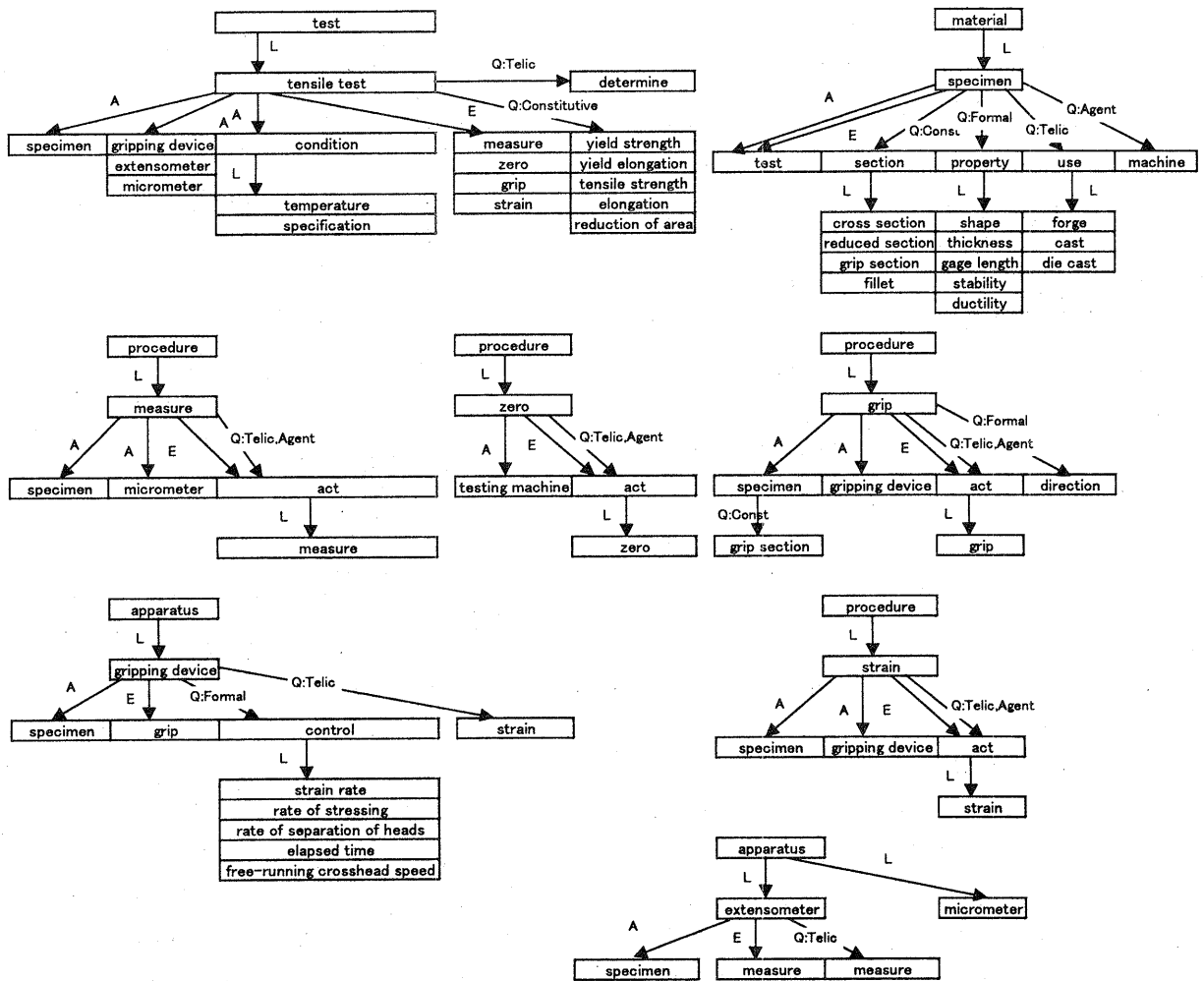


図3 tensile test についての意味記述 (test, specimen, procedure, apparatus)

## (2) Event Structure(EVENTSTR)

対象に伴う動作を表す意味タイプを指す。tensile test では、measure (引張る前の標本のサイズを測定する)、zero (伸び計の目盛りを正確にゼロに合わせる)、grip (標本の両端を引張り機械で固定する)、strain (標本を引張る) といった試験過程で生じる動作が EVENTSTR で関係づけられる。

## (3) Qualia Structure

### (3)-1 Constitutive

対象を構成する部分や、対象の表すことがらにより得られる結果の項目を指す。tensile test では、yield-strength (降伏応力)、yield-elongation (降伏伸び)、elongation (破断伸び)、reduction of area (絞り) とする。specimen (標本) であれば、それを構成する部分とするので、cross section, reduced section, grip section, fillet が Constitutive に関係づけられる。

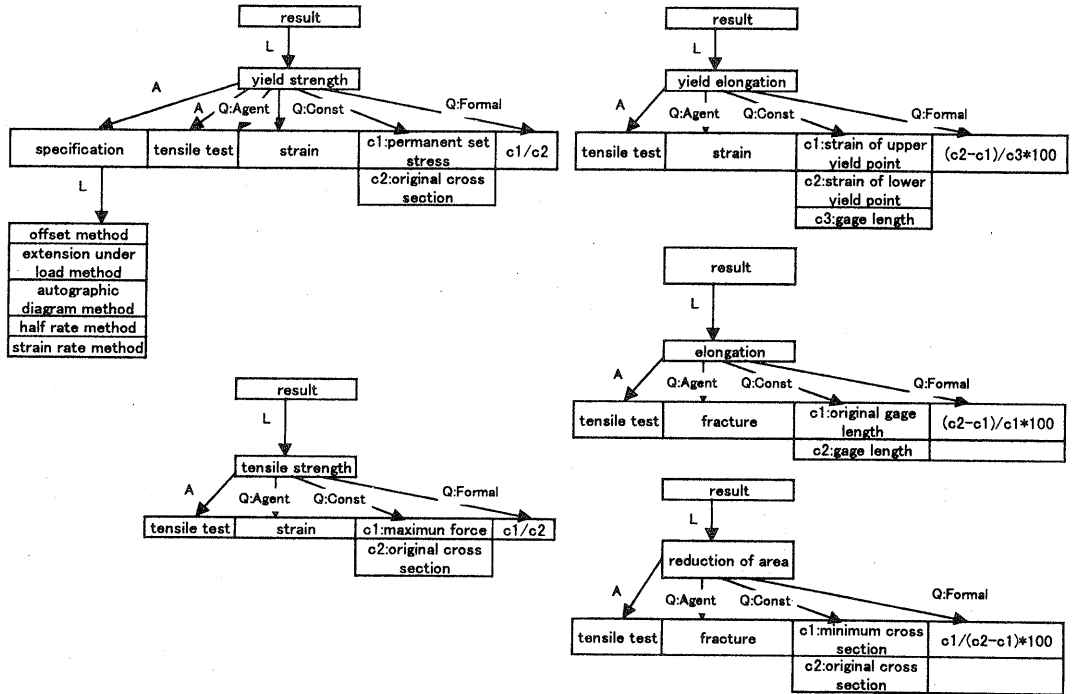


図4 tensile test についての意味記述 (result)

### (3)-2 FORMAL

FORMAL は対象の特性を指す。specimen であれば shape, gage length, thickness, stability, ductility とする。result タイプの語であれば、結果として得られる値を求める式に必要な項目とする。

### (3)-3 TELIC

そのことがらをなす目的やその対象を用いる目的を指す。tensile test であれば(3)-1 で述べた yield strength などを決定することなので、determine とする。specimen タイプでは、鍛造するためなどに標本が用いられることから forge 以下の内容とする。

### (3)-4 AGENTIVE

対象となることがらを導く動作や、対象を生じる動作を指す。specimen は加

工して作られるので AGENTIVE は machine とする。

### (4) Lexical Inheritance Structure

語の意味タイプについて上位・下位関係を設定し、下位のタイプは上位のタイプの内容を上書きしない限り継承する。

### 4. 実際の文章から作成した意味記述例

3章で述べた内容にしたがって tensile test について書かれている実際の文章について意味記述を行った。二例を以下に示す。{}の中は用いる意味タイプを表している。{}の付いていない記述は実際の文章で用いられている語を ARGSTR 等に分類して表している。

(1) Tensile testing was performed at 22°C using a crosshead speed of 0.127 mm (0.005 in.) per minute.

{tensile test}

ARGSTR:temperature:22°C

{gripping device}

QUALIASTR:Formal:crosshead

speed:0.127mm(0.005 in) per minute

(2) Tensile tests were conducted at both constant strain rate and constant crosshead speed at various combinations of temperature.

{tensile test}

ARGSTR:temperature:various combinations

{gripping device}

QUALIASTR:Formal:strainrate:constant

Formal:crosshead speed:constant

## 5. 考察

2, 3章での記述方法に沿って4章で実際の文章に現れる語を記述したが, 本研究が目的とする, 論文内容の抽出や, 複数の論文間での差異を詳しい内容にまで立ち入って抽出することに利用するためには以下の項目が有効である.

- (1), (2)の文章とも, crosshead speed についての説明があるが, crosshead speed というのは, 図3の gripping device の Formal の箇所にある. このことから, 文章中には明に実験装置を用いたとは言及されていなく

ても, 何らかの gripping device が用いられていることがわかる.

- (1)と(2)において, 採用している意味タイプが一致しているものについては比較できるが, (2)における strain rate の記述は(1)にはない. このことから(1)と(2)における strain rate の有無という点が差異の一つであると考えられる.
- 図3の tensile test の意味記述において, specimen の ARGSTR が test であると同時に, EVENTSTR も test である. このように, 名詞と動詞の test を意味役割という点から統合的に表現することで意味記述が簡潔になっている.
- 図3の tensile test の意味記述中では specimen が ARGSTR に関係づけられ, その specimen も個別に意味記述を持っている. このように意味記述を1つだけではなく, その中にある意味役割としての語のタイプにも個別に意味記述を持たせることで, 内容についての粒度の細かい記述ができる. 例えば, tensile test の記述では, 装置として gripping device が必要であることと, 発生する動作の1つが strain であることが分かる. 一方 gripping device の記述から, その装置を使う目的が strain であるということが分かる. よって両者の内容を統合して, tensile test での引っ張る動作は gripping device によって行われることがわかる.

## 6. 今後の課題

### (1) 意味記述の自動更新

意味記述の作成について、用意すべき意味タイプがある程度収束するまでは人間の手作業で行わなければならない。しかしそれ以降は、新しく文章が入力された場合にその文章に現れる語を既存の意味記述の枠に統合させるような手法の開発が必要である。加えて未知の語の意味タイプを判定するしくみも必要である（開発中）。

### (2) 意味解析システムの開発

用意した意味記述を操作して、考察で述べたような、文章間の差異や内容の統合ができるシステムを開発する予定である。そのためには、生成語彙における3種類の生成的演算も採り入れて考える必要がある（開発中）。

## 参考文献

- [1] 郡司隆男, 田窪行則, 石川彰 (訳) : 現代の文法理論, 産業図書(1988).
- [2] 郡司隆男 (訳) : 制約にもとづく統語論と意味論—H P S G入門, 産業図書, (1994).
- [3] Pustejovsky, J. : The Generative Lexicon. MIT Press(1995).
- [4] Jackendoff, R. : Semantic Structures. MIT Press(1990).
- [5] 松尾利行, 西田豊明, 星本健一 : 金属材料論文からの技術情報空間の構築と探訪の知識支援, 人工知能学会誌, Vol.12, No.1, pp.68-77(1997).
- [6] 難波英嗣, 奥村学 : 論文間の参照情報を考慮した学術論文要約システムの開発, 情処研報 Vol.98, No.82, pp.79-85(1998).
- [7] 郡司隆男, 阿部泰明, 白井賢一郎, 坂原茂, 松本裕治 : 意味 (岩波講座言語の科学4), 岩波書店(1998).
- [8] American Society for Testing Materials, Annual Book of ASTM Standards, Vol. 03.01, Metals-Mechanical Testing; Elevated and Low Temperature Tests; Metallography, ASTM(1995).
- [9] 金属術語辞典, アグネ(1965).
- [10] 日本規格協会, JIS G202 鉄鋼用語 (試験) (1987).