修　士　論　文

コーパスを使ったマルチエージェントシミュレーションによる言語意味の普遍性と相対性の研究

北陸先端科学技術大学院大学
知識科学研究科知識システム基礎学専攻

伊藤　正彦

2006 年 3 月

目　次

図　目　次

1. 始めに

人間はそれぞれが別々の経験をしている。同じ物事に対しても別のアプローチの情報を取得し、その内容は千差万別のはずである。しかし、どれほど別の経験をしていたとしても同じ言語を操る限り、基本的なコミュニケーションに困るということはない。逆に、まったく同じ言葉を聞いても、人によって違った意味にとることがある。その人のいる環境や立場によって言葉の定義が違うという現象もまた、珍しいものではない。

このように人は様々な情報を入手して生活し、一方で内容に関わらず同じ言葉の意味を取得し、一方で言葉の意味がその人の言語経験に依存している場合が存在する。このような言葉の意味が個人の経験差に関係なく一定の定義に収束する現象を言語の意味の普遍性と言い、個人の経験によって言葉の意味が変化する現象を言葉の意味の相対性と言う。

言葉の意味の普遍性と相対性の研究は「まぎらわしい」、「分りやすい」、「具体的」、「曖昧」、「抽象的」、「主観的」、「客観的」等の言葉の定義に対する解釈や人によって解釈が替わり易い単語の性質や、さらにそもそもなぜ人間は共通の言葉を持つことが出来るのかという根本的な話題にかかわるものであるが、どんな単語でどのように起きているかを計る方法は今のところないため、暗黙的にその存在は知られていたが具体的には社会学的な言葉の利用傾向といった研究で終わっていた。

言語というものは絶対的な上位者によってコントロールされているトップダウンのシステムではなく、数限りない言語の使用者がコミュニケーションに利用することにより少しずつその構成を変えていくボトルアップのシステムである。[Hashimoto 2004]そのため、言語の意味に関するこの二つの現象は各個人の言葉の意味を長時間見て始めて分析できるものである。

人間が言葉の意味をどのように定義しているかという研究は認知科学[Pinker 1995]、脳科学[Yamadori 1998]など様々な分野で行われており、様々なモデルが提案されているがまだ本格的な解明には時間がかかる。

言語の変化という大きな話題について様々な研究を行って成果を出している手法として構成論的アプローチが上げられる。とりあえず考えられる仕組みを実際に作成してみてその動きを見ながらその仕組みの妥当性を考えるというやり方である。この手法はコンピュータシミュレーションやロボットを用いて行われ、実際に観測しようのない現象や既に起こってしまって見ることの出来ない現象などがどのように起きたのかを考えるためによく使用される。その中でも特に個人もしくは生物一個体を一つの自律プログラムとして表現し、それを同時且つ大量に動作させることで何が発生するかを分析するマルチエージェントシミュレーションは、特に言語進化の分野で多く利用され、集団における共通語彙の生成についてもこの手法を使った研究が行われ、一定の成果を出している。

しかし、言葉の意味の普遍性と相対性という話題については、実際の言語を使われている環境が持つ規則性や多様性を表現する部分で現在のマルチエージェントシミュレーションでは扱いきれていない。

そこで、この研究では言葉の意味に対して限定的に実際に使われている現状をシミュレーションに組み込むことにより、単純なマルチエージェントシミュレーションでは不可能だった言葉の意味の普遍性と相対性についての分析を行おうと考える。具体的には各エージェントが実際に使用されている文書を使ったコーパスを使ったシミュレーション手法、Corpus Agent を提案し、言葉の意味がコミュニケーションを通して変化していく様子を再現する。その過程での各単語の意味がどのように収束、変化していくかを通して言葉の意味の普遍性と相対性の観測し、そして単語の特性との関連性を分析する。

以下のこの論文の構成を説明する。
第２章においてまず、この研究の対象である言葉の意味の普遍性と相対性とは何かを

説明し、言語進化の分野において一般的であるマルチエージェントアプローチを説明し、その事例として特にこの研究の直接の改良対象となった Naming Game を紹介する。さらに、意味の計量を行う分野としてコーパス言語学を紹介する。
第３章でこの問題についての今までの研究の問題点を明らかにし、研究の目的を述べる。
第４章にてコーパスにおける単語の共起分析を利用して現実の言葉の意味をマルチエージェントシミュレーションに組み込む手法、Corpus Agent を提案し、またその詳細を説明する。
第５章では実際にコーパスエージェントを実行し、どのような結果が出たかを紹介し、さらにそれを分析する。
第６章はこの研究の意味を改めて考え、結論にて研究を総括する。

2. 背景

まず背景として言葉の意味について現在ある学説を幾つか説明し、その後この研究の立場である認知意味論的立場を説明する。その後、認知意味論的立場の特徴であるプロトタイプ論の説明を行う。

次にこの研究の主題である言葉の意味の相対性と普遍性について説明する。この二つの概念はあまり知られていない概念であり、研究者ごとに違う言葉を使う場合があるため、この研究での定義を明確化する。

さらに、この分野の研究手法として一般的であるマルチエージェントシミュレーションについてその長所と問題点を解説し、その一つとして今回の研究の参考としたNaming Game の紹介する。

最後にマルチエージェントシミュレーションの問題点を解決する方法として現実の文書データから言葉の意味を分析するコーパス言語学について解説する。

2.1. 言葉の意味とは

言葉の意味はいったいどのようなものであるかを考える試みは古代哲学の時代から続く作業であり、多くの意見と定義が生まれてきた。言葉が物事に意味をつけること

は特殊な行為であると考えられてきたことは、旧約聖書にて神が最初に言葉を作ったと記されていることや、呪術的な風習の多くは物に特別な名前をつけることによって行われること、また、言葉で言ったことのように物事が進行すると考える日本の言霊信仰などからも伺える。
20 世紀以降、近代的な科学のアプローチや心理学、言語学の発達により、言葉の意味についての研究は大きな進歩をした。特にソシュールの記号言語学や、論理学から発した形式言語学の発達により言葉の意味についてより具体的な提案がされ、さらに認知科学的の分野では複数の言語への調査や脳科学の発展により概念に対する発見がなされた。
現在ある言語の意味についての考え方で主要なものは以下のとおりである。

1、認知意味論的考え方
言語の意味を外界の指示物を決定すると言うより認識された外界をカテゴリ化したものとしてとらえるもの。

2、形式意味論的考え方
言葉の意味をその言葉を正しく利用する条件とする。よって、言葉の意味はその条件を導き出すことによって一意に定義できると考える。

3、構造主義的な考え方
言葉の意味は他の言葉との関係によって表現される。その単語と他の単語の類似、対立などの関係によって理解されると考える。

現実の意味の捉え方はおそらくこれらが複合的に行われていると考えられるが、この研究では基本的に認知意味論的な態度を取る。その理由は個人的な言語経験の集積によって言語の意味が決まると考えるとき、カテゴリへの帰属を中心にその言葉の定義を考えることが出来る認知意味論的立場は柔軟性があるためである。

認知意味論における主要な概念としてプロトタイプ理論がある。
プロトタイプ理論とは、言語学・認知心理学上の概念で、人間が実際にもつカテゴリ

は、必要十分条件によって規定される古典的カテゴリではなく、典型事例とそれとの類似性によって特徴づけられるという考え方であり、このようなカテゴリをプロトタイプ的カテゴリと呼ぶ。たとえば「鳥」という語から想起されるのはカラスやスズメなどの空を飛ぶ小動物であり、ダチョウやペンギンなどは典型事例から外れている。典型性の差にもとづく現象は一般にプロトタイプ効果と呼ばれる。またこれに関連して、「鳥は飛ぶ」のように特別な文脈上の理由がないかぎりデフォルトとして仮定される状況は理想化認知モデルなどと呼ばれる。プロトタイプ理論は 1970 年代にロッシュらによって提唱された。

プロトタイプは以下のように形成されると考えられているが、基本的な言葉の意味については様々な制約の元で概念の形成を行っていく幼児期に行われるのでさらに複雑なものになっていると考えられる。

1、指示対象を指す言葉を与えられる
2、新しいカテゴリが作成される
3、直感的、もしくは誰かに教えられてそのカテゴリに該当するものが追加されていく
4、該当例から帰納的に典型的な特徴が抽出され、それにしたがってそのカテゴリに付いてのプロトタイプが形成されていく
5、ある程度のプロトタイプ形成がされた後新たな事例が発見された場合、現在形成されているプロトタイプに従って言葉の意味に近いかどうかが判断される。

2.2. 言葉の意味の普遍性と相対性

前節にて言葉の意味とはプロトタイプを中心としたカテゴリ化した概念であると述べた。それではどの言葉においてもその言葉にとっての必要条件と典型条件は同じように分布しているのだろうか。

例として前節でも登場した「鳥」と言う言葉の意味のカテゴリと、まったく別の言葉として「勉強」と言う言葉のカテゴリを考える。

鳥と言う言葉では、ある程度の年齢をつむと何が鳥であり、何が鳥でないかと言う生物的な条件がある程度一般的になり、必要条件が明確になってくる。
一方、「勉強」と言う言葉ではどのような行為が該当するか、典型はある程度明らかになっているが、必要条件の境目がどのようになっているかはあいまいである。

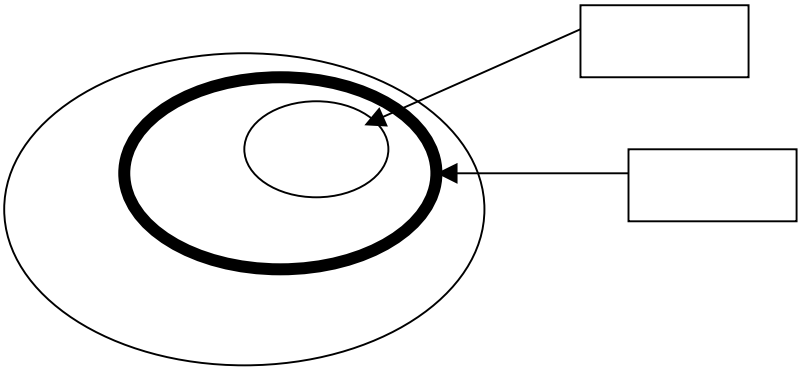


図 1:「鳥」の意味のモデル

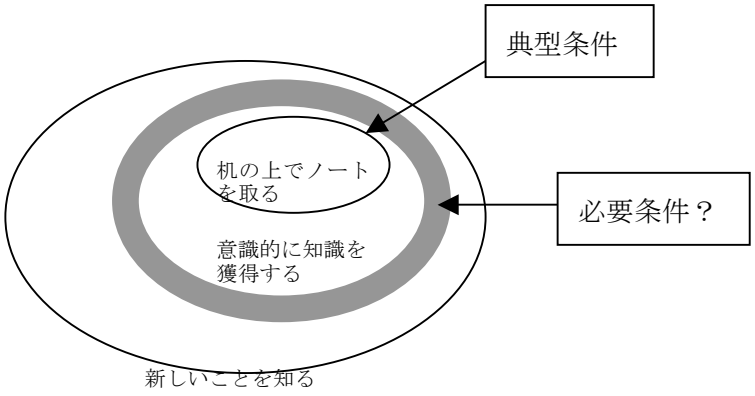


図 2:「勉強」の意味のモデル

また、典型例が人によって違う場合もある。例えば「ネットワーク」と言う言葉の意味を考えたとき、情報系の分野の人々は真っ先に LAN 等のコンピュータネットワークを思い浮かべるはずであり、その一方で社会科学の分野の人々は人と人のつながりを思い浮かべるはずである。

典型条件:理系

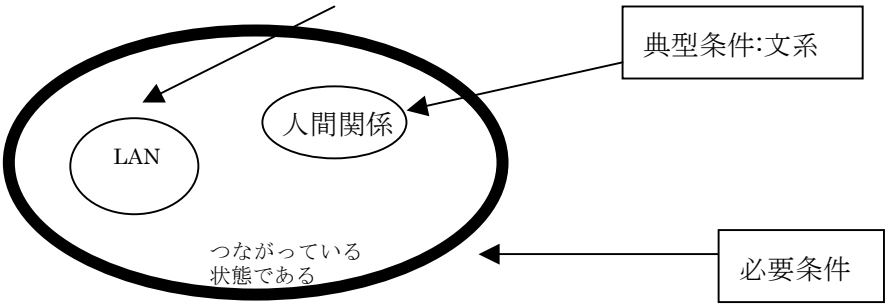


図 3:「ネットワーク」の意味のモデル

このように、言葉によってカテゴリにおいての必要条件や典型条件の枠組みは大分異なる。これは、言葉の性質によって言語経験から形成されるカテゴリの必要条件と典型条件の出来方が違うためである。

結果として、言葉の意味は言葉の特質によって特に合意を得る作業がなくてもほぼ同じ内容を指す場合があれば、人(の言語経験)によって内容にずれがある場合もある。この現象を言語の意味の普遍性と相対性という。普遍性は「鳥」のケース、相対性は「勉強」や「ネットワーク」のケースである。

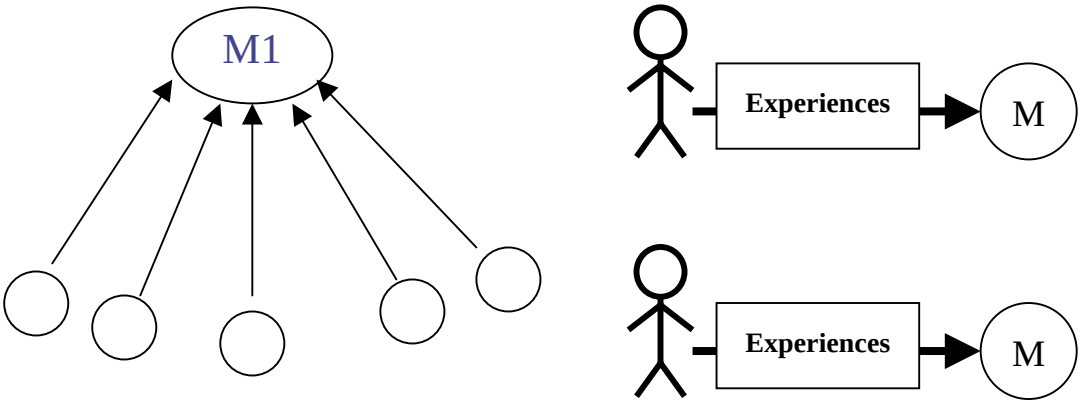


図 4:　言葉の意味の普遍性(右)　相対性(左)　のモデル

基本的に言語の意味は普遍性を得ようとする傾向が強い。何故なら言語の存在価値は

自分が伝えたい認識を相手になるべく正確に伝わるようにする事であり、個人ごとに定義の違う言葉は存在意義がないからである。しかし、言語の意味がそれぞれ個人の経験に依存して形成される中で、他人と言葉の意味にずれがあるものも発生する。

言葉の意味の普遍性と相対性は個人が如何に言葉の意味を形成するかと言う内部的な問題と、それを他人と共有できるかと言う外部的な問題が組み合わさった問題である。そのため、この現象を研究するためにはまずどのように発生しているのか、どの言葉で発生しているかを観測する必要がある。しかしあまりに大規模で同時多発的な現象なので実際に起こっているものを観測しようとするのはほとんど不可能であるといえる。

2.3　マルチエージェントシミュレーション

言葉の普遍性と相対性の研究はまずそれをどうにかして観測する方法を探すところから始まる。現在この研究が属する言語進化の分野では同じように観測したくても出来ない現象を観測するための手法として構成論的アプローチ、その中でも特にマルチエージェントシミュレーション（MAS）を多く利用される。ここでは MAS の紹介と、この研究と同じように集団においての共通語彙の形成について研究した MAS として Naming Game の紹介をする。また、同時にそれら既存の MAS をこの研究に使う際の問題点についても言及する。

2.3.1. MAS による構成論的アプローチとは

構成論的アプローチとは対象となる現象や事物に対して直接調査を行うのではなく、コンピュータシミュレーションやロボットなどを用いて対象のモデルを作成し、それを動かすことによって対象を理解しようとする方法論である。　[金子・津田 96, 橋本 02c]この方法は実証的観察が困難である現象や進化のように歴史性、一過性を持つ現象に特に意味を持つ。

MAS はコンピュータ上に作成された自立的に動くプログラム、エージェントを同時に複数動かしそれぞれを相互に干渉させることにより、現実に起こっている様々な現象をシミュレートさせる実験手法である。

科学に置いての基本的な現象の実証方法は対象に対して直接行う実験や観測であるが、社会的な現象や進化など、そもそも観測や実験そのものが不可能もしくは、困難である現象も多くある。また、観測することは出来てもそのメカニズムがどうなっているのかを調べることが不可能なもの、例えば、人間の認識と社会の関係などは研究の困難さから発展を止めているものも多くない。しかし、コンピュータ技術の発展により、高性能なコンピュータが安価で大量に使えるようになったことを背景に今まで観測が不可能だった対象をコンピュータ上に複製し、それを観測することによって今まで解明できなかった現象の解明を行う試みが行われ始めた。この方法は現象を細かく分析していく還元的な手法との対称として研究の対象の構成一度作ってみて考えていく所から構成論的アプローチと呼ばれている。

MAS は構成的アプローチの代表的なものであり、コンピュータ上に自立的に動くプログラム、エージェントを複数配置してそれを互いに干渉するようにして動作させ、それによって起きる状況を観察するものである。それぞれのエージェントは研究者が想定したロジックに従って行動し、それが大量に動き回る姿は、さながらモニター上を蟻がうごめいているように見える。マルチエージェントで行われている実験は生物の行動や社会行動といった一般的なもの以外にも、国際戦略、経済活動、文化の発生、そして先に紹介した Naming Game やこの研究が属する言語進化など、多くの分野の実験が行われている。

2.3.2. MAS の長所

MAS を導入することにより、今まではその一過性やあまりに長い期間で発生するために研究の仕様のなかった言語進化や、社会変化の分野においてどのような仕組みで

その変化が発生するのか、またどのような要素が変化を促すのかを研究することができるようになった。その理由は複雑な研究対象をモデル化して動かすことにより研究者が対象の注目している部分がどのような役割を持っているのかをより具体的に把握することができるからである。また、パラメータを変えたり、ロジックを変えたりして繰り返し実験することで一般的な方法では調べること自体が難しいものでも様々な条件で調査することができる。

2.3.3 Naming Game について

言葉の意味が言語経験によって形成されていく過程を分析した研究はあまり多くない。何故なら、基本的に人が同じ言葉の意味を共有していることや、逆に人によって言葉の意味に違いがあることは当たり前のことであり、それがあまりに当たり前であるからこそ、分析が難しく、また、人と人がとある言語について違う意味を持っていると言うことを計測する方法は今のところないからである。

数少ない研究の中から、集団がコミュニケーションの結果として同じ言語の意味を共有する過程を再現した研究として Naming Game [Steels 96]を紹介する。この研究は言葉の意味に普遍性が存在する仕組みと、この研究とは違う手法で言葉の意味の多様性が発生する仕組みを提案している。

2.3.3.1 Naming Game の概要

Naming Game はその名のとおり物の名前を付け合って集団の言葉の意味を形成しようとする MAS で、集団が共通の語彙を形成する過程を再現しようとしたものである。

Naming Game の基本的な構想はウィトゲンシュタインの「論考」[Wittgenstein 1916]より着想を得たもので言葉の意味とは何か対象を示すことであるとする定義から始まっている。

エージェントは言語知識として幾つかの単語と各語が指し示す対象の対<o, w>のリストを持つ。そしてその各対にはコミュニケーションにおいて使われた頻度とコミュニケーションの成功回数を表すスコアが付く。コミュニケーションの際、エージェントには speaker と hearer という異なる役割を割り当てられる。

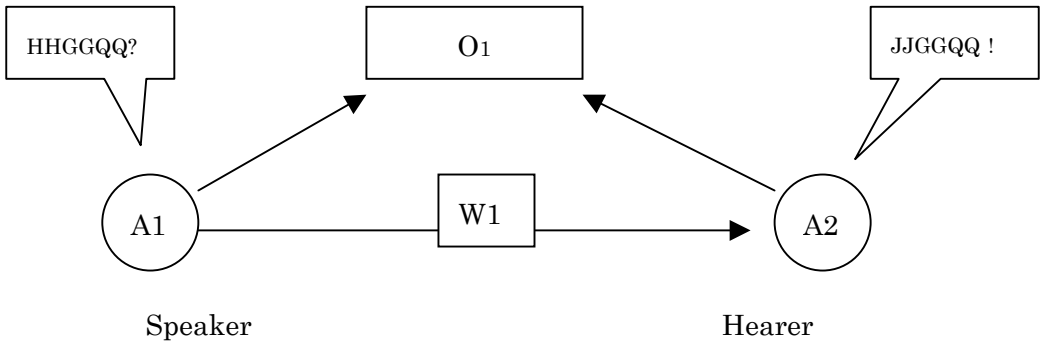


図 5:　Naming Game のモデル図

Speaker はある対象について単語列をもちいた表現を発話し、Hearer はその表現から対象をデコードしようとする。
両者の想定する対象が一致すればコミュニケーションは成功であり、その時に使われた単語、対のスコアは上昇する。失敗した場合はその理由に応じて、Speaker が新しい言葉を作ったり、Hearer が新しい組み合わせを採用したりする。
最終的に複数のエージェントがコミュニケーションと学習を繰り返した結果、エージェントたちは共通の語彙、すなわち単語と対照の組み合わせの集合を獲得する。
また、コミュニケーションの際に一定の確率で伝達の誤情報が発生し、それによってコミュニケーションの内容は多様性を発生させるようになっている。

2.3.3.2 Naming Game の問題点

新しい手法として注目されている MAS による仮想言語による共通語彙の形成の手法はコミュニケーションによって共通した語彙を形成させることには成功し、普遍性に

ついてはある程度の存在を証明したがいくつかの点で実際の言語習得と明らかに違う部分を持っていることが指摘できる。

1.相対性についてはコミュニケーションの失敗によって発生した誤情報や、非言語的なコミュニケーションが単語の利用法に多様性を導き出していると言う考え方は言語の習得方法をあまりに単純に見ている。言語の相対性や多様性は別にコミュニケーションエラーに依存しなくても言語自体が持っている用法の広さによって起きるはずであり、それは使用方の間違いではなく、言語の特性そのものであると考えるべきである。

2、Naming Game ではその名のとおりエージェントが自分がまだ名前を付けていないものに対して名前を付けあい、それを他のエージェントが確認し、自分の語彙に追加する事によって語彙を増やしていく方法となっているが、この過程で必要なのは自分と相手が同じオブジェクトを見ていることであり、同じ対象について話していることを確認することができることである。しかし、実際には同じ対象を確認しながらを会話することができるのは幼児のときの言語習得の課程ぐらいであり、それ以外は相手が何を話しているか推測しながら言語の意味を判断していくこととなる。つまり、言語を覚える際は対象を確認しあうことではなく、言語の使われ方だけを見て言語を覚えることの方が多いはずだ。また、主要な単語の使い方については対象を確認しながら覚えることが出来たとしても、ある程度年齢が経ってから覚える概念的な単語や抽象的な単語についてはその対象を見ながらその内容を確認しあうと言うことは原理的に不可能である。そのため、幼児期の言語獲得期のシミュレーションは別として、言語の意味形成のシミュレーションは二人が確認できる対象に対して名前をつける手法では現実との齟齬が発生するはずである。

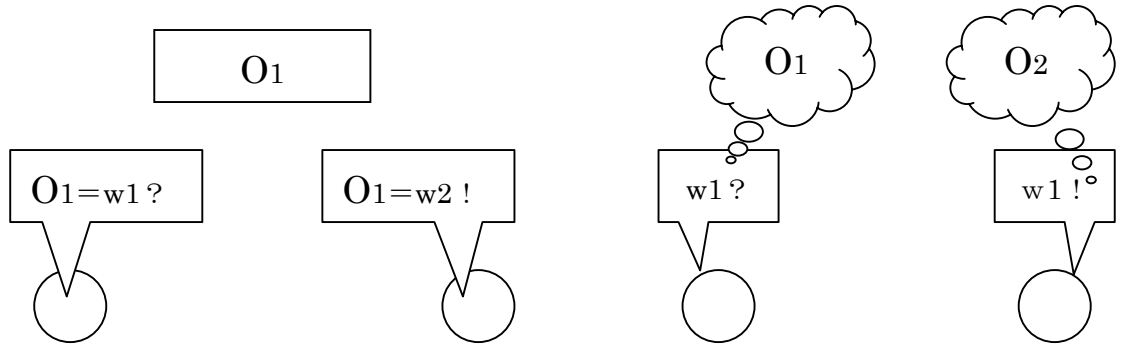


図 6：Naming Game のモデルと実際の会話モデル

3、Naming Game において言語は単純に対象を指し示す記号として取り扱われているので、対象の特質や文法的な違いなどには考えを置いていない。全ての単語がまったく同じ立場で習得される前提において作成されている。しかし、現実においては単語の特質によって覚えやすさや重要性はまったく違っており、幼児期の言語獲得においても言葉は明らかな優先順位をもって習得される。[今井 1997]
言語の意味の普遍性と相対性はこの言語そのものの特質に大きく依存したものである可能性が高い。しかし、すべての言語を同じように扱うことしか出来ない Naming Game のやり方は現実の言語の意味形成とは大きな違いがある。

4、人間は言語の意味を学習、修正していく。子供の頃は語彙が少ないため一つの言葉が多くの言葉の意味を含んでいるが、大人になると言葉の意味を細分化させ一つの言葉が担当する範囲は狭くなる。また、場面によって違う言葉の用法に適応し、話をする相手によって使う言葉の用法を変える。言葉を一度覚えた後もその意味は生きている間絶えず調整していくのである。しかし、Naming Game においては一度覚えた単語の意味は二度と修正する機会はなく、同じ対象をさす単語に置き換えられていくのを待つだけである。言語習得のシミュレーションには覚えた言語をさらに状況に最適化していく仕組みが必要である。

2.4. コーパスによる意味分析

現実の言葉を扱った分野としてはまず、コーパス言語学が挙げられる。コンピュータの普及以前から文献を統計的に扱う試みは行われていたが、コンピュータによって多くの文献を簡易、迅速に扱えるようになり、そのため今まではデータとして取り扱わなかった一般文書や会話に着いての情報も研究対象となり、さらに、特定の分野の研究だけなく、コーパスから言語そのものを解明しようとする研究も行われるようになった。

2.4.1. コーパス言語学とは

実際に使用された大量の言語資料を集めたものをコーパスと呼び、そのコーパスを利用して、より実際的な言語の仕組みを探る学問がコーパス言語学である。
チョムスキーの生成文法が言語能力（Competence）を調査対象とするのに対し、コーパス言語学は言語運用（Performance）に焦点を当てる。
具体的には、言葉の原理を演繹的に考えていこうとする一般の論理言語学（一般のMAS において利用される言語学は要素を抽象化するためこちらを利用される）とは対照的に、大量のテキストデータをもとに言語を帰納的、統計的に考えていこうとするものである。

実際に使用された文章は自動的に一定の規則（文法、統語、音韻）の上で書かれており、さらに、文章は単語の意味を利用して書かれている。文法的には正しいが何の意味もない文章（ex: The red speaks to the bottom of sky）はといういものは存在せず、文章に使われる単語はその文中で使われる単語と一定の関係性を持って存在している。その関係をいちばん簡単な形で表現しているのが特定のコーパス内においてどのような単語の共起関係をしているかをあらわしたコロケーションである。

コロケーションを言語研究に Firth（1951）[Firth 1951]は利用し始めたのが最初だが、その後 Harris（1971）[Harris 1971]が「分布」の概念を持ち込み、単語の意味をその後の利用状況によって表現する手法が始まった。彼の説は以下のとおりである。

分布とは、同じコーパスにおいての統語環境であり、ハリスは分布と意味を以下のように関連付けた。

1、分布が同じなら意味は同じである。

2、分布が部分的に同じならその程度に応じて意味も同じになる。

3、分布がまったく異なるなら意味はまったく異なる。

この、分布の考え方を導入することにより、言葉の意味を定量的に表現することが可能になった。

race	0.02
big	0.032
stray	0.001
run	0.02
bite	0.06

表 1： “dog” についてコーパスから共起を計算したもの

コロケーションによる単語の意味分析はコンピュータの発達とともに盛んになり、コロケーション分析では現在出版されている新聞から単語の一般的な意味を導き出す試み[Church、Hanks 1991]が行われている。

2.4.2. MAS に対するコーパス言語学の長所

多くの MAS の問題点は実際のデータを根拠に置かず作成者の構想のみを具体化してしまうため、そこから出力された結果の有用性が低くなることである。特に言語進化の分野では誰も現実での同様の状況を観測したわけではないのでそれを間違いということは出来ないが、そのような研究の増加によって学問分野の有用性まで疑われることになる。

そのため、MAS では現在の言語進化の研究においてはこの問題に対しては言語をできるだけ数学的に表現しモデルを抽象化することで、数学のもつ論理によって研究の有用性、説得力を維持しようとしている。しかし、言語の利用には多くの要素が含まれており論理的に現在使用されている言語を構築するにはまだ課題は多い。その状態で

抽象化、簡略化することはまだ解明されてはいないが言語の意味を構成する重要な要素をスポイルしてしまう可能性がある。
人間が言語の意味を脳内でどう定義しているかを分析することはまだ脳科学分野でも結論は出ておらず、さらに、どのような経験がそれをどのように変化させていくかの研究は始まったばかりである。よって、現実の言語をデータとして使いたいが、そのために人間の言語を学習、操作するメカニズムを再現することは大変な困難である。

コーパス言語学では言語の意味を取り扱う問題点を逆に言語の意味そのものを取り扱わない手法で行うことで解決する。コーパス言語学では、一般的な言語規則以外においては言語の細かい吟味はしない。するのはその単語と他の単語がどのような関係性があるかを統計学的手法によって計算された結果である。つまり、その単語がどのような意味があるかを知らなくても、どう使われたかを分析することでその内容を把握できるという立場である。

コーパス言語学での意味分析を利用することにより、逆にどのような単語の意味についても端的な結果を表現する事が可能になる。これにより Naming Game では同じ対象を確認できる環境でなければ二人が同じ意味を共有しているかを確認することが確認できなかった問題を克服している。また、この研究の趣旨である新たな言語体験によって言葉の意味が変化していくという部分についても、使用された用法を重視するコーパス言語学の趣旨に合致する。
何より言うべきことは、今まで架空の言語でしか表現できなかった意味の変化を実際に利用されているものを利用することで研究の説得力が増すことが、コーパス言語学を利用する最大の利点である。

3. 問題点/研究目的

3.1. 問題点

関連研究で述べたとおり、言語の意味の普遍性と相対性の研究は困難が多く、分かっている部分より分かっていない部分のほうが多い。

1.認知言語学で行われている研究は言語の使用されている領域全体から比べればほんの一部分に過ぎず、応用を行える部分はかなり少ない。また、サンプルの収集の困難さからどうしても違う言語との比較によって相対性、普遍性の分析を行うことしか出来ず、同じ言語の中でもこれらが発生していることを分析することが出来ない。また、この問題は人間の言語の習得に大きく関わっているものであるはずだが、同じく実験の困難さから学習時間とこれらの特質との関連性の分析はほとんど行われていない。

2、Naming Game により、コミュニケーションによって共通語彙の発生が発生することは証明されたが言語習得の再現について幾つか問題があり、それによって相対性と普遍性の確認にはまだいたっていない。

結局のところ、どのように普遍性と相対性が人間の言語生活において発生しているかを確認する研究はまだされておらず、長い時間がかかることや、サンプルの抽出が極めて難しいことを考えると一般的な人を使った実験による分析はやはり不可能であることが推察される。今まで行われた架空の言語環境によるシミュレーションを使ったものも、架空であるがゆえに現実の言語環境の重要な部分を反映出来ていない。もし、これに改良を加えてその部分の反映を行ったとしても、今度はどうしても恣意的なパラメータを追加せざるを得ず、シミュレーションそのものの信頼性の低下から逃れることが出来ない。

言語がダイナミックなものであり、私たちは生活において言語が変化していく中で行われているのは明白であるが、それが具体的にどのように起きているのかは結局のところまだ計測/確認されていない。現時点のこの分野の研究はつまり、あるのは確実だが触ることが出来ないものの前で足踏みしている状態なのである。
また、確認できたとしても（今後の Naming Game の発展系において前述の欠点の幾つかが解消されたものが開発されるかもしれない）今度はどんな言葉が普遍性を持ち、どんな言葉が相対性を持つかと言う問題が注目される。一般的な例を見れば言葉によって相対性、普遍性の出現の差があることは明白であるが、その差を研究することは、そもそも二つが計測されていない現在では行われていない。

3.2. 研究の目的

問題点から、この研究の目的を述べる。

1、まず世の中に存在している言葉の意味の普遍性と相対性をできるだけ具体的な形で測定し、その存在を明らかにする。
今までに Naming Game 等で研究された集団での語彙形成の研究をさらに進め、現実に使われている言葉によって言葉の意味の普遍性と相対性が世界にどのように存在しているかを表現する。具体的にはマルチエージェントシミュレーションにコーパス言語学を組み合わせて言葉の意味を実際に計りながらエージェントが学習していくシミュレーションを構築する。

2、言葉ごとの意味の普遍性、相対性を測定し、その出現と言葉の特性との関連性を解明する。
1、のシミュレーションを構築することにより実在の言葉の普遍性、相対性がどのようになっているか観測する。その結果とその単語の意味や用法を比較し言葉と普遍性・相対性はどのような関係があるのかを分析する。

4. 研究手法

この章ではこの研究での手法を説明する。
まず、問題点/目的から手法に要求されるものを導き出す。
その要求に対する回答としてマルチエージェントシミュレーションによる構成論的手法とコーパス言語学による意味分析を併用したシミュレーション技法 Corpus Agent を説明する。
そして、それらをどのように使うかと言う方法を説明した後、システムとしての実際の作成方法を説明する。

4.1. 手法に要求されるもの

目的や、今までの関連研究に欠けていた部分を組み合わせると、この研究に要求されている部分は以下のようになる。

1、明確な形で相対性、普遍性を定義し、それの発現を確認できる手法であること
言葉の意味の普遍性、相対性の出現についての研究の困難さが困難なのは一つには具体的にどのようにそれの出現がなされたのかを確認する方法がないことということだ。言葉の意味が場所によって違うことがあることも、どんな場合でも同じ言語なら

ば基本的には言葉が通じることも、感覚では分かっているが、それが具体的にどのように起こっているかを計測することはまだされていない。一番基本的な要求は万人が納得できる確認方法で二つを出現させることができることである。

2、言葉の意味を定量的に計測できる手法であること。

1、を確認するためにはまず言葉の意味そのものが明確に計測できる形になっている必要がある。何故なら、意味の普遍性と相対性は言葉の意味の変化に基づくものであるからだ。意味の普遍性とは人と人が事前のコミュニケーションを行わなくても言葉の意味をほぼ同じように理解していることであり、相対性とはそれまでのコミュニケーション活動によって言葉の意味が変化している現象である。つまり、人と人との言葉の意味が遠いか近いかを計測できなければ理解しようがないものなのである。そのため、この研究のためには個人が使っている言葉の意味と他者の言葉の意味がどれだけ近いかを定量的に計測できる手法をとる必要がある。

3、他者とのコミュニケーションによる意味の学習のプロセスを持っていること。

言葉の意味の相対性、普遍性は言語の意味学習に際してのコミュニケーションへの依存度の違いと言うこともできる。そのため、その研究にはコミュニケーションによって言語の意味を変化させていく過程が必要であり、それによってどのように意味が変化したかを計る必要がある。

4、現実の言葉を利用した手法であり、どの単語について分析したのかを明確にできる手法であること

Naming Game における問題の一つは現実の言葉を取り扱わず架空の言語を作成していったため現実の言語には当たり前に存在であろう属性を取りこぼして考えている可能性がある。

研究の目的の一つは、言葉の意味の普遍性、相対性が言葉のどのような特質に依存しているかを分析することである。そのため単語毎に相対性、普遍性が出現しているかを確認できる手法を取る必要がある。

5、現実的な期間、環境において実行できる手法であること

何よりも、この分野の研究がなかなか進まなかった理由の一つは実験などを行うのに

必要な期間と準備が実際の人間を使った場合膨大になってしまうからである。一番必要な条件として実現可能な実験方法であることが望まれる。

4.2. 要求達成のための基本的なアイデア

背景で紹介したMASとコーパスによる意味分析は、まったく別の研究分野ではあったが、言葉の意味の変化を分析すると言う部分において二つは相互補完する事が出来る。つまり、

A, マルチエージェントシミュレーションは物事の変化を再現し、分析することができるが前提としている条件は全て架空のデータを使っており、恣意性を否定できない。

B, コーパスによる意味分析は現実に利用されている言葉の意味を定量的に表現できるがそれはあくまである程度のデータの統計的な結果であり、それがどのような変化の結果であるかはわからない。

そこで、この研究では二つを組み合わせて、新しい形のシミュレーション手法を提案する。すなわち、現実の文書データを、エージェントに言語経験として持たせ、それを改変することによって意味を学習していくシミュレーション手法、Corpus Agentである。

Corpus Agentのアイデアは以下のとおりである。
1、全てのエージェントは自身の言語経験としてコーパスを持っている。

2、エージェントは自分の知っている言葉の意味をコーパス内でのその単語の文章内での共起として表現する。これは頻度として定量的な表現方法で表される。

3、各エージェントは自分の言葉の意味を相手に見せることでコミュニケーションを取る。もしその内容があまりに違っていれば、二人はコミュニケーションが取れない。

4、エージェントは他エージェントとのコミュニケーションの結果に従い学習をすることができる。学習とは、新たな情報を入手して次のコミュニケーション時に意味が近くなるようコーパスを修正しておくことである。

無論、人間の言語活動は統計的な手法を取っていると言う研究はないし、言語経験が全ての情報を等価値に並ぶ一枚の文書データとしてまとめられているわけではない。しかし、入力として新しい情報が入ること、そして自分の持っている言語経験から言葉の意味を導き出すという過程については現実と同じであり、自身の考える言語の意味を他者と比較するという点のみを考えれば、その内容は十分に妥当なものであると考えられる。

4.3. Corpus Agent の説明

ここでは Corpus Agent の説明を行う。まず、元となる人間のコミュニケーションモデルを提示し、それを再現したシミュレーションモデルを解説する。

4.4.1. モデルとなるコミュニケーションスタイル

この研究では、人間の一般的なコミュニケーションを以下の形式で表現する。
シャノンが作ったコミュニケーションモデル[Shannon 1949]に学習フェーズが付加されたもので、妥当性の高いものであると考える。

1、話し合う相手を決める

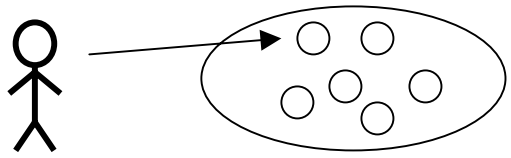


図 7.a: 人間のコミュニケーションモデル 1

2、話す相手に話題とする単語を伝える。

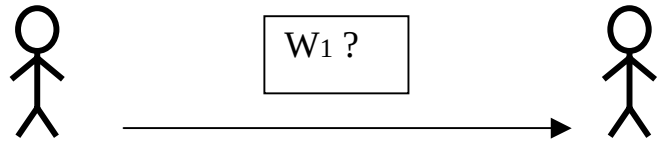


図 7.b: 人間のコミュニケーションモデル 2

3、双方で自分の持つ言葉の意味を提出し、比較しあう。これは実際にはそれぞれが自分の言語の意味に基づいて会話を行うことを示している。

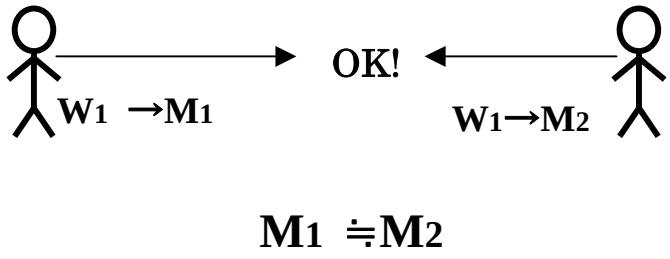


図 7.c: 人間のコミュニケーションモデル 3

4、もし相手と自分の言葉の意味が十分に近ければ会話は意味が通ったものとして成立する（会話が続くなら 2 へ、そこで会話が終わるなら 1 へ戻る）が、遠ければ言葉の意味が違いすぎるため意味が通らなかったことになる。

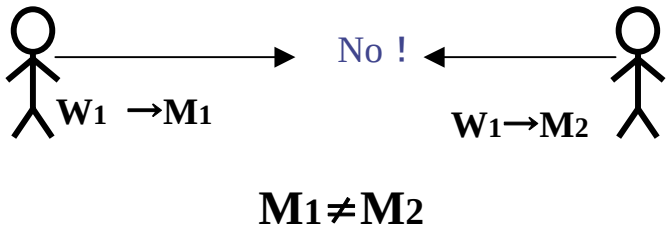


図 7.d: 人間のコミュニケーションモデル 4

5、学習をする。

言葉の意味が通じなかった場合、自分の言葉の意味がなるべく一般に通じるようにその言葉についての情報を入手する。

Naming Game ではこのフェーズで逆に自分が言いたいことを伝えられる言葉を捜して行動するようになっているが、どちらが人間の行動として妥当であるかというより、人間は二つの行動を同時に行っており、その注目する部分が研究によって違うと考えるべきであろう。

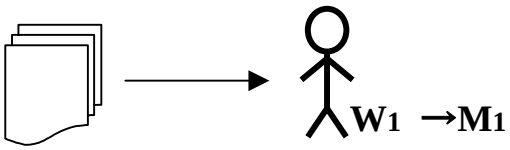


図 7.e: 人間のコミュニケーションモデル 5

4.3.2. 対応するシミュレーションのステップ

前述のコミュニケーションモデルに対応するステップとして、Corpus Agent は以下のステップで稼動する。
なお、アルゴリズムの効率性の関係で幾つかの処理は人間の行動より前倒しになっている。また、この研究では各単語に注目するため、一回の実験で対象にする単語は常にひとつの単語である。

1、エージェントはコミュニケーションを行う相手を選ぶ

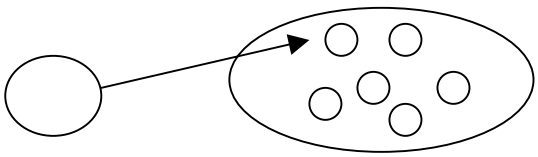


図 8.a: エージェントのコミュニケーションモデル 1

2、エージェントは自分のコーパスの一部（時系列的に昔に取得した内容）を削除し、新たに情報を取得したコーパスを作成する。

この時点で新しいコーパスはまだ一時的なものであり、昔のコーパスは内容を保持される。

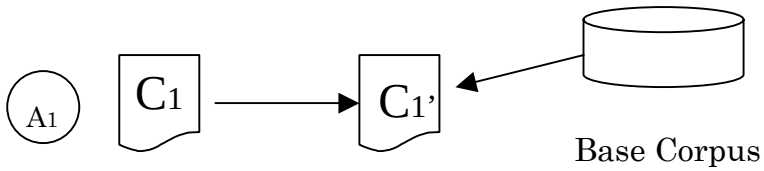


図 8.b: エージェントのコミュニケーションモデル 2

3、コーパス（新旧双方とも）から言葉の意味となるフレーズ上での共起を計算する。

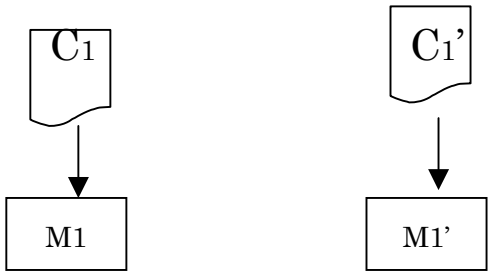


図 8.c: エージェントのコミュニケーションモデル 3

4、相手となるエージェントの単語の意味と自分の新旧両コーパスでの単語の意味を比較し、どちらが近いか決める。

基準はそれぞれの意味のユークリッド距離となる。

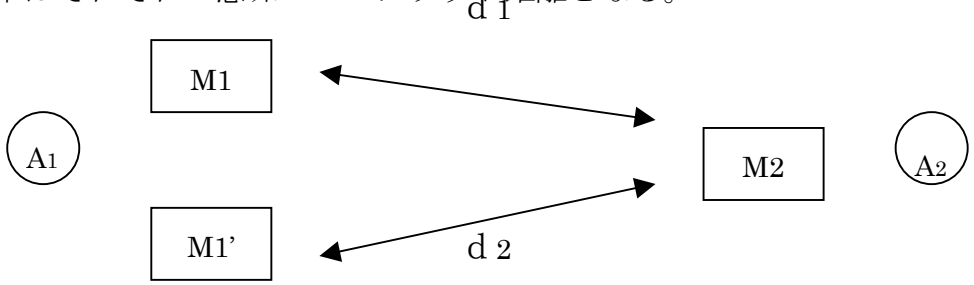


図 8.d: エージェントのコミュニケーションモデル 4

5、もし新コーパスのほうが近ければ新たなコーパスとして登録する。旧コーパスの方が近ければコーパスの内容は維持される。

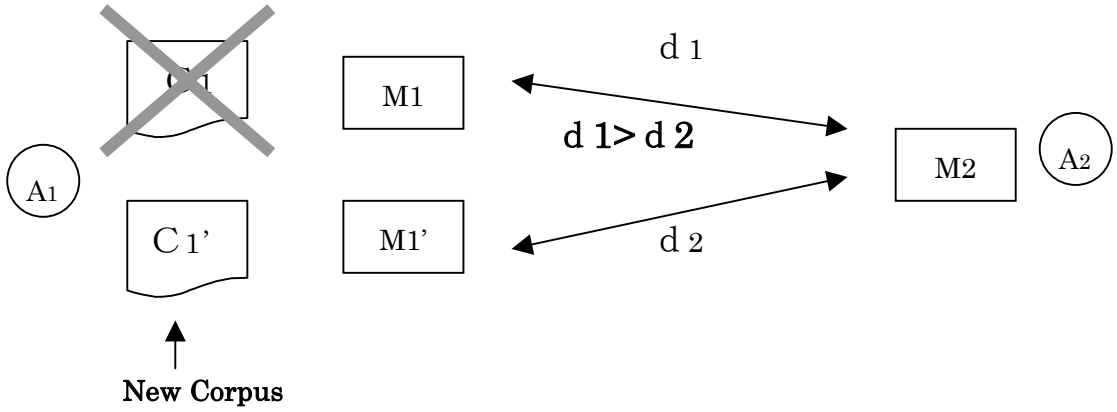


図 8.e: エージェントのコミュニケーションモデル 5

4.4. システム構築の実際

4.3 でこの研究のため作成するシステムの概要を説明した。ここでは実際にシステムをどのように作成されたかを説明し、システムの詳細を紹介する。
紹介は以下のとおりである
1、シミュレーションの世界状況を構成するためのベースとなるコーパスの作成
2、さらに各コーパスから単語に対する共起を計算するアルゴリズム
3、複数のエージェントが相互に干渉しあうマルチエージェントシステムの構築

4.4.1. コーパス作成

この研究では、各エージェントが自分の情報として持つ情報のほか、学習の際周囲の世界から情報を習得するために得る情報源として、大量の文書データが必要となる。この研究ではできるだけ広範囲で且つ偏りのない内容の分布をもつコーパスを大量に作成するためのシステムを作成し、結果として 11G バイト以上のデータ量を持つコーパスを作成した。

4.4.1.1. データ取り出し

この研究の基本となる文書データは極めて膨大、且つ広範囲であるが、逆に現在人々が普段読んでいるテキストが結果として欲しいため文章の質そのものはそれほど考慮しない。そのため、コーパスを作成するデータソースはインターネットから取得することに決めた。

現在インターネット上には様々な情報が存在しており、そのほとんどは HTML 形式で記述されたテキストファイルであり、そこには様々な文章が記載されており、一般に公開されている。この研究では半自動的にこれらのファイルを取得するシステムを作成し、大量の HTML ファイルを取得することに成功した。

このシステムは自分が取得したファイルから URL を取り出し、さらにその URL からファイルを取り出すことを繰り返すことで取得した HTML からリンク先が完全になくなるまで無限に HTML を取得しつづけるものである。

また、このシステムは取得したファイルを 10M バイトごとにまとめ、コーパス全体を大量のファイルの集合体として扱うようにした。また、作業効率を上げるために 30 スレッドのマルチスレッドによる設計で同時に複数箇所の URL への問い合わせを行って間断なくファイルを取得できるようにしたため、実験に利用する文書コーパスのサイズとしては十分なデータ量を短期間で入手することが出来た。

4.4.1.2. 下処理

前節のシステムによって 55 ギガバイトと言う相当量の文書データを手に入れたが、現在利用されている HTML ファイルは多くの割合が通常の文章以外のデータによって構成されるため、中から文章のみを抜き出すため下処理を行った。
下処理は Perl によって行い、以下の処理を施した所 55G バイトのデータは 11G まで減少した。
以下が行った下処理である

- タグ内の情報を全て削除
- JavaScript の情報を全て削除
- ２重のスペースを一つにする
- タブを全て削除
- ピリオドとカンマの後ろを全て改行
- レイアウトのために行われている改行を削除し行を繋げる
- 前置詞と冠詞、代名詞を削除
- ［　］　に囲まれた情報を削除

4.4.2. 共起頻度計算システム

コーパスから単語の共起を計り、意味を算出するアルゴリズムを開発した。この研究ではこの計算が相当回数行われるため、単純な共起のみで言葉の意味を表現することにした。
以下がその処理手順である。

1,コーパスからターゲットの単語が含まれる文を抜き出す。
2,文を単語で区切り配列を作成する。
3,単語のリストを作り配列の単語を登録していく。もし単語が既に登録されていたらその単語の個数のパラメータを一つ追加する。
4,コーパスの文を全て走査したらターゲットの単語が含まれていた文の数で登録されていた単語の個数を割り共起頻度を計算する。
5,上位 20 個の共起頻度をその言葉の意味としてメモリに登録する。

4.4.3. マルチエージェントシステム作成

一番時間がかかり、かつ仕様が複雑になったのはマルチエージェントシステムである。言語は Java を使用し、マルチスレッドによって各エージェントを表現した。

各エージェントは自分の言語経験としてコーパスを持っており、その内容を書き換えていくのがこの研究の大きな特徴であるが、それをそのままシステムとして作成した場合、以下の課題が出現した。

１、各エージェントのコーパスの内容をどのように維持保管するか

もし各エージェントのコーパスを実際に作成し行動ごとにその内容を読み、修正ごとにファイルに書き出しを行わせた場合、特に修正の部分で大きな処理負担になることが予想された。また、処理を早く行うためにコーパスの内容をメモリに蓄積した場合、大量のメモリ空間を消費してしまう。コーパスの情報をどのように蓄積するか

2、二つのコーパスの近さをどのように計るか。

各エージェントはコミュニケーションの際、二つのコーパスの意味を比較するが、何を持ってその評価を行うか

3、新コーパスは何を基準に作成するか

学習の過程として新規のコーパスを現在のコーパスから作成することになるが、もっとも妥当なコーパスの更新方法は何であるか

これらに対し、本システムは以下の構成で作成した

1、ファイルリストによるコーパス管理

この研究で世界全体を表現するのに使用されるベースコーパスは大きな一枚のテキストファイルではなく、総計 5480 ファイルの大量のファイルによって構成されている。これは管理をしやすくするのと同時に情報をセグメント化してどの単位で情報を扱われるかを表現するためである。

そこで、各エージェントは自分のコーパスを独立したファイルを作成するのではなく、自分がベースコーパスの中からどのファイルを利用しているのかを記したファイルリストをメモリ上に所有し、コーパス読み込みの際はそのリストを元に各ファイルを読み込み全体として一つのコーパスを作成するという形式を取った。この形式のおかげで、エージェントはファイルを一々書き出し自分のファイルを管理する手間を省略することが出来た。

2、ユークリッド距離による相関比較

各エージェントが算出した自分の新旧コーパスからの共起(上位 20 個のみを採用)は相手コーパスの共起とのユークリッド距離を算出した値によって比較される。ユークリッド距離による比較を採用した理由は双方がどのような共起を持っていたとしてもどちらが近いかの判断を明確に行える点、どの言葉との共起が強かったかと言う度合いの違いも距離に反映される点、最後に計算が速く処理が簡潔に終わる点がある。

(ユークリッド距離の計算式)

二つの意味の値が $M_a = \{a_1, a_2, a_3, a_4, a_5.....\}$　$M_b = \{b_1, b_2, b_3, b_4, b_5.....\}$　である時

$$Dis\tan ce = \sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + (a_3 - b_3)^2 + (a_4 - b_4)^2 + (a_5 - b_5)^2 .....}$$

ただし、片方にしかない単語が存在する場合、相手の値は 0 とする。

3、キュー方式によるファイルの交換

コーパスを構成するファイルリストはキュー形式になっており、ファイルの取得順序が履歴として残るようになっている。一時的に作成されたコーパスでは新規に追加されたファイルはキューの後ろに追加され、前の同数のファイルが削除される。この形式を採用することにより昔の情報が順次記憶から消えていく過程も再現された。

5. 実験

前章の手法を元に作成したシステムにより、実験を行った。この実験で明らかにすることは実験目的である
a､言葉の意味の普遍性と相対性を観測すること
b､二つの特質と言葉の特性を分析すること
である。

この章ではまず実験の目的である言葉の意味の普遍性と相対性の発現を実験においてどのように定義する。
次にどのような条件で実験を行ったかを記し､その結果を述べる。最後に結果の分析を行い､実験によって目的が達成されたかを評価する。

5.1. どのように普遍性、相対性を計測するのか

前章の説明により､システムにおいて､エージェントがコーパスを書き換えていくことによって自己の言葉の意味を自分の環境に最適化していく過程を解説した。言葉の意味の相対性と普遍性は個人がコミュニケーションを行うことによって起こるが、現象としては集団を観測して分かるものである。そこで、この研究では各エージェントのコミュニケーション時における意味の距離の変化を時系列で一覧して観測することにした。

以下がこの実験で利用するグラフである。X 軸はコミュニケーションを行った延べ回数、Y 軸はコミュニケーション時におけるエージェント間の意味の距離をあらわしている。

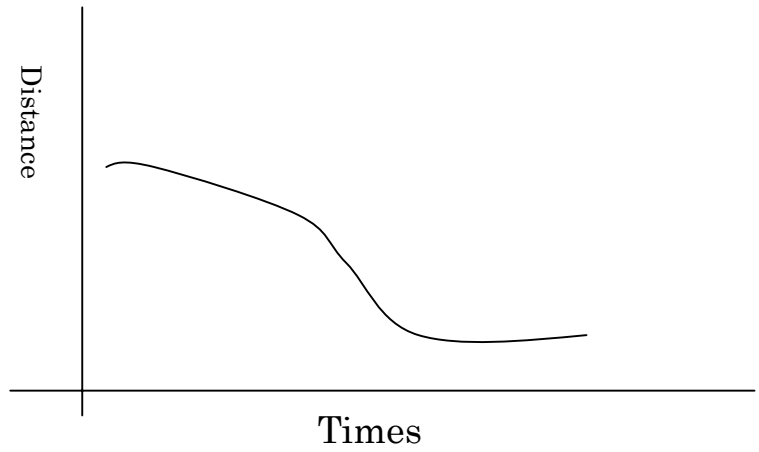


図 10: 実験から出力されるグラフ例

さて、このグラフの中で普遍性と相対性はどのように表現されているのであろうか。普遍性の特徴は以下のとおりである。

・同じ言語を使っている場合どのコミュニティでも言葉の意味はほとんど同じ

・話している間に内容が変化しない

・言葉の意味の変化はほとんどない。

以上を踏まえると、普遍性の高い言葉の典型は以下のように変化するはずである。

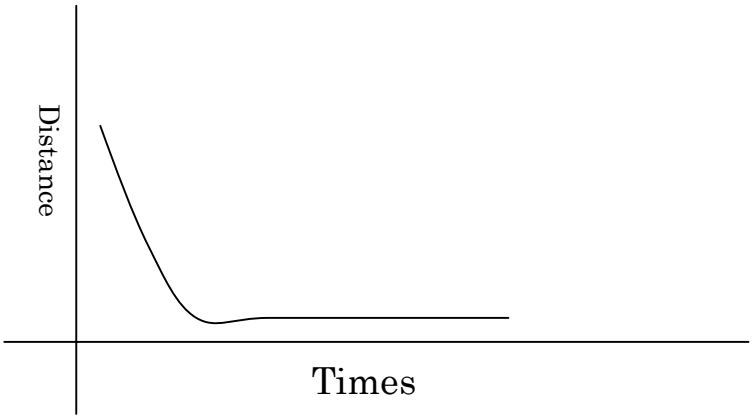


図 11: 普遍性が高い時のグラフ例

つまり、どのエージェントとも言葉の意味が簡単に縮まり、コミュニケーションを行っても距離の変化がほとんどない。言葉の意味がコミュニケーションへの依存が小さい状態である。

相対性は普遍性の逆に以下のような特徴をもつ
・コミュニティごとの意味が違うため、違うコミュニティから来た同士では言葉の意味が大分違う
・話す相手によって言葉の使い方が違うことが多い。
・話している間に言葉の定義を調整して話すこともある。
以上を踏まえると、相対性の高い言葉の典型は以下のように変化するはずである。

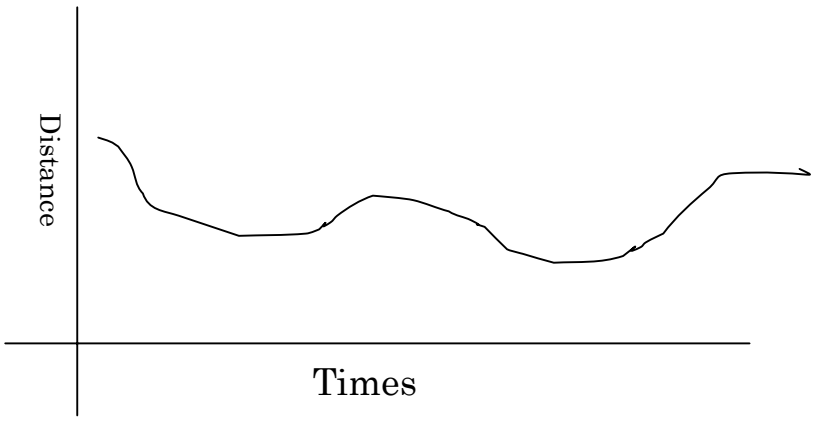


図 12: 相対性が高い時のグラフ例

相対性はコミュニケーションによって言葉の意味が変化しつづける現象である。そのため、エージェント間の言葉の距離は変化しつづけ収束するということがない。

言葉の相対性と普遍性は相反する傾向であり、このグラフでも収束するかどうかという尺度で二つの区別を行う。よって、もし言語にその存在がなければ、もしくはこの実験で二つの存在が発見できなければグラフは片方の傾向に集中するはずである。

5.2. 実験の条件

実験は以下の条件で行った。

エージェント数:15

エージェントごとのコーパスのサイズ : 220M 程度（実際には作業効率化のため 20M 程度になっている）

コーパスを書き換える際のデータ変換割合 10%

行動回数：　エージェントごとに 1000 回

実験対象とした単語：

dog, bird, green, movie, head, Mouse, god, body, water, food,
country, company, group, beautiful, justice, freedom, bad,
black, white, Information, cool

※基本的に普通名詞と動詞、さらに形容詞の中から利用頻度の高いものを無作為に選択して行った。ただし、システムの問題からその単語の文字列を含む無関係な単語も含んでしまうため(例：Cat に対する Catch)その恐れのある単語についてははじめから実験を行わなかった。

5.3. 実験結果

前節の実験条件に従い実験を行い、計 28 件の結果を出力し、これらの結果は 3 種類のグループに分類された。
この章ではまず、実験結果についてまずこの三つのグループの分類を説明し、さらに目的である言葉の意味の普遍性と相対性の発現の確認と言葉の特質と言葉の普遍性と相対性の関係の分析を行う。
なお、実験で算出される数値は共起頻度から算出したユークリッド距離、コミュニケーション時におけるコーパスの交換が起こる確率、そしてコーパス交換が起きた時の起きた時と起こる前の差である。

5.3.1. グループ 1

このグループは実験開始直後から強い収束傾向をもち以下の共通点を持っている。
・早い段階（1000 回以内）にエージェントの言葉の意味が近い距離は 0.05 以内に集中
・コミュニケーションにおいてコーパスの変換があった場合変化の差は平均 0.0072 以内の非常に小さな変化
・変換発生の確率は 0.06 で、かなり少ない

このグループは、以上の点から言葉の普遍性が極めて高く、初期値での誤差をコミュニケーションによって修正しさえすれば同じ言葉の意味を簡単に共有できる単語の集合である。

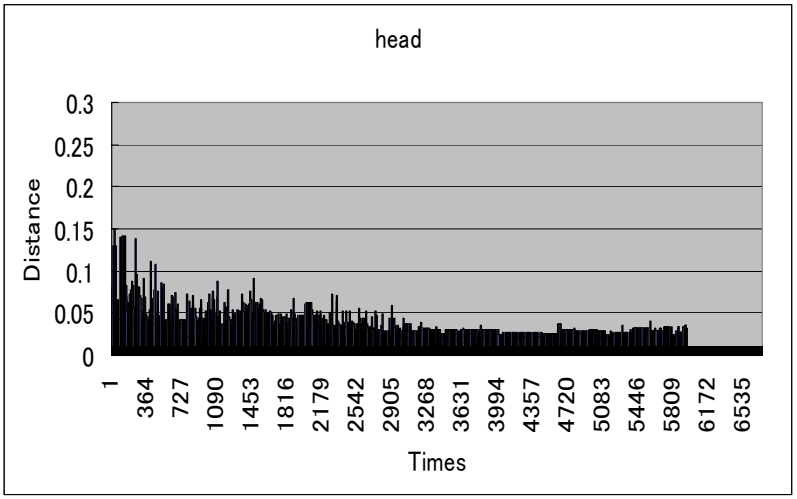


図 13: “Head”をターゲットにシミュレーションした結果：距離

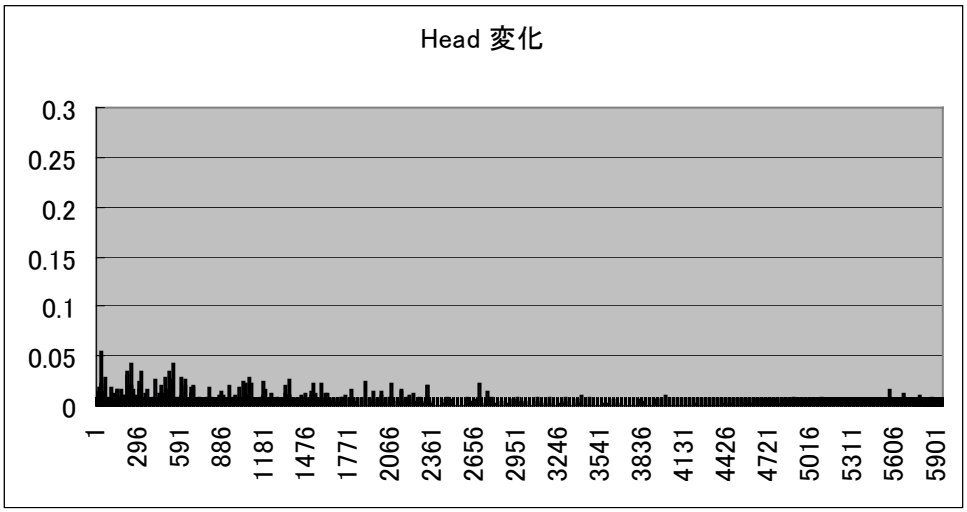


図 14: “Head”をターゲットにシミュレーションした結果：距離の変化差

このグループに該当する単語は white,　country, head, green, movie, fight　で、傾向として明確なオブジェクトとしての対象が存在する言葉、もしくは対象の確認が簡単な言葉が該当している。

5.3.2. グループ 2

このグループは実験開始直後において意味の距離はかなり遠いが、コミュニケーションを繰り返すことにより次第に意味を収束させている。

・初期からしばらくはそれぞれのエージェントの距離は遠いが、コミュニケーションによってエージェント間の意味の距離は短くなり、最終的にはエージェントの言葉の意味が近い距離 0.05 以内に集中する。

・コミュニケーションにおいてコーパスの変換があった場合変化の差は平均 0.015 から 0.02 までの中規模の変化。

・変換発生の確率は　0.1 前後

このグループは、元々はグループ 1 のように特定の意味を持っていたわけではないがコミュニケーションによって話している集団がその言葉を特定の意味で使用する合意に至ったと考えられる。コミュニケーションによって言葉の意味が大きく変わったという面では相対性があることが確認できるが、ある程度コミュニケーションを行った後は言葉の意味の差がかなり低くなっているため、それ以降は普遍性を確認できるという事も出来る。

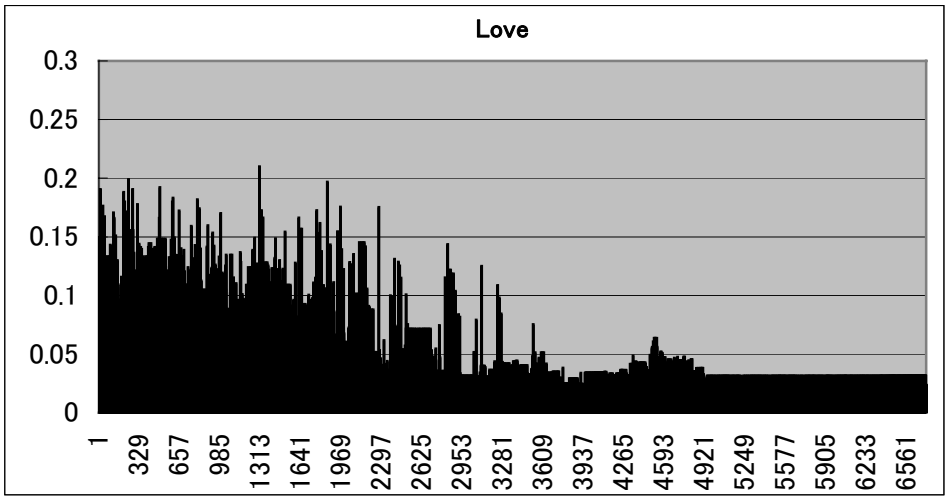


図 15: “Love”をターゲットにシミュレーションした結果　:距離

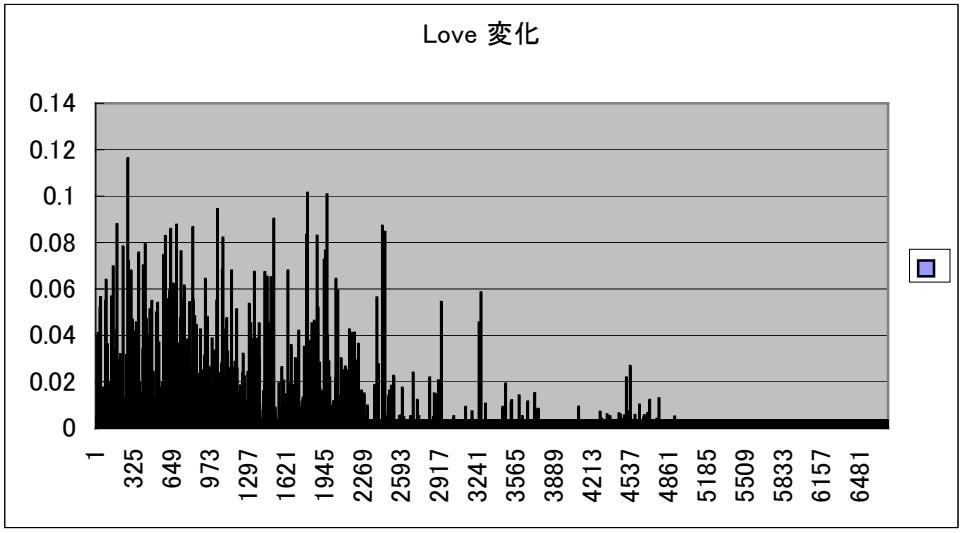


図 16: “Love”をターゲットにシミュレーションした結果　:距離の変化差

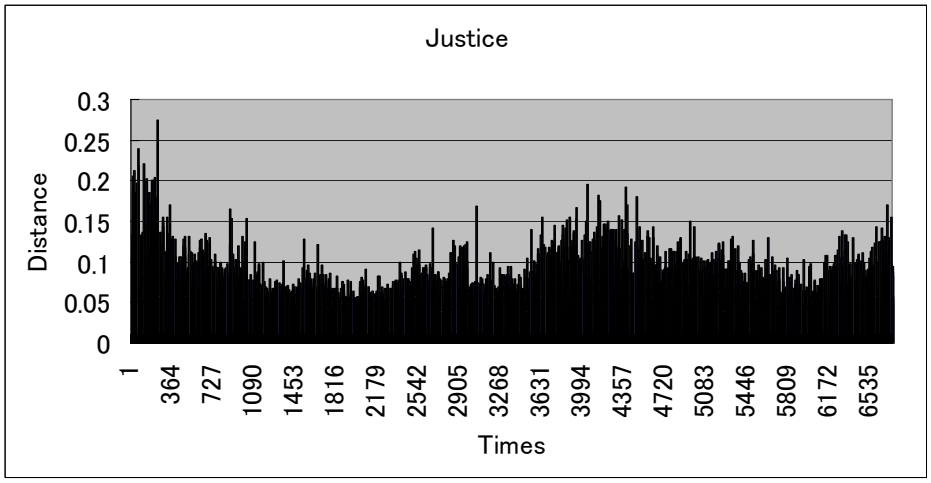
このグループに該当する単語は　people, love, food, information などで、傾向として抽象的、概念的な言葉で、場面場面によって様々な使われ方をするが具体的な定義があまり明確でない言葉が多い。

5.3.3. グループ 3

このグループは実験開始から終了まで、そのときそのときの傾向はあるが最終的に言葉の意味は収束しない。

・それぞれのエージェントの距離は遠く、コミュニケーションによってエージェント間の意味の距離は一時的には短くなるが、一方でエージェント間の距離を短くすることが別のエージェントとの距離を広げるということを繰り返しており、最後まで収束する気配がない。
・コミュニケーションにおいてコーパスの変換があった場合変化の差は平均　0.02　以上、かなり大きな変化である。
・変換発生の確率は 0.2 以上。頻繁に発生している。

このグループは単語の中でも相対性が強く、一回ずつのコミュニケーションによって大きく言葉の意味が変化している。



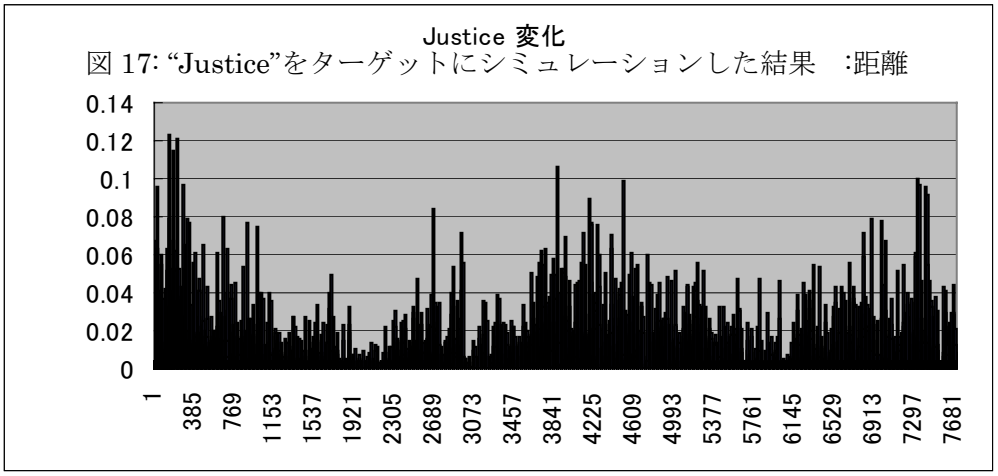


図 17: “Justice”をターゲットにシミュレーションした結果　:距離

図 18: “Justice”をターゲットにシミュレーションした結果　:距離の変化

このグループに該当する単語は justice, beautiful, freedom, cool,　god で、傾向として価値観や感情、評価に関する単語が多い。

6. 最後に

ここでは実験の結果に対する分析、そして研究の問題点と今後の改良点について記す。

6.1. 分析

分析すべき内容としてはまず、二つの目的が達成されたかどうかを述べ、そして実験結果から普遍性と相対性について何が言えるかを考える。

6.1.1. 言葉の意味の相対性と普遍性は発現したのか

もし、特定の単語に対する他の単語の出現率が均一であり、どの単語とも特に強い関連性がなかった場合、各エージェントは初期状態で偶然持った共通点を元に暫時的に意味を収束させるグループ 2 の形態になるはずである。

しかし、実験により、各エージェントの動きは単語毎に違っており、意味の形成もかなり違うと言うことが明らかになり、言葉の意味の相対性と普遍性の存在が明らかになった。

グループ 1 において収束がすぐに開始し、単語の意味の距離が短い状態で維持されたのは、元から単語の使われ方には強い一般的な傾向があり、それが短いコミュニケー

ションの間に全てのエージェントで共有されたと考えるべきであろう。その証拠として、コーパス交換時の距離の変化が活発の期間が短く、さらにその変化の幅も小さい。つまり、コーパスを構成するファイルがランダム指定される初期状態を設定した時点で既に一定のかなり近い距離を持つことができているのだ。

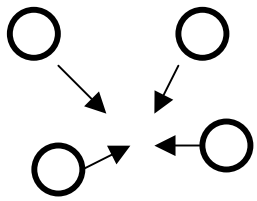


図 19: グループ 1 での意味の変化

反対にグループ 3 においては、なかなか言葉の定義が収束しない。各エージェントの動きとしては他のエージェントの意味に近づくよう行動しているのに何故収束できないのか、各エージェントの動きを細かく分析した。

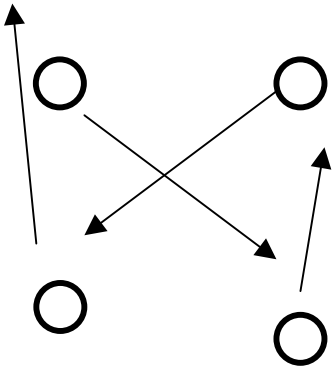


図 20: グループ 3 での意味の変化

各エージェントの特定の他エージェントとの距離の推移を見ると、まずコーパス交換時の距離の差が大きい。つまり一回の学習で言葉の定義が大きく変わっている。必ずその後また長い距離に離れてから接近を開始している。一度の交換で接近する距離が大きいためその交換以前に接近した他のエージェントとの距離が初期化されてしま

うことがわかった。

以上のように、コミュニケーションにほとんど依存せず言葉の意味の合意が出来てしまうグループ 1 の状態、また、コミュニケーションによって意味が変化しすぎ共通の意味をもてないグループ 3 の状態。両者は共に言葉の意味の普遍性と相対性の状況を如実に表している。その中間であるグループ 2 では相対性はあるが低いため、コミュニケーションを行うことで合意を得られる域であったと言うことができる。

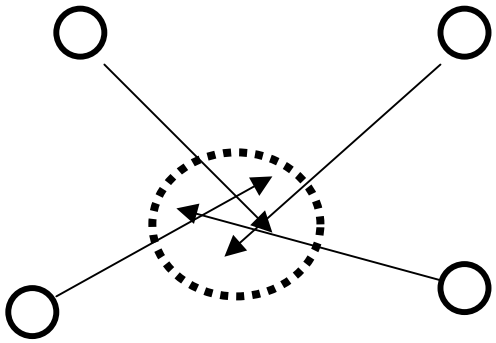


図 21: グループ 2 での意味の変化

6.1.2. 言語との関係性は見つかったのか

計 27 語について実験を行った結果言葉を 3 つのグループに分けることが出来た。この結果とそれらの言葉の一般で使われている意味や用法を比較し、定量的な計測がどれほど一般的な人間の言語感覚に合致しているかを分析してみた。

まず、グループ 1 の所属した言語は head, dog, green,　などであった。これらの単語は可視的な物体や､物理的な状態として対象が存在する、明確でどのようなときに使うかが具体的にイメージできる単語である。
もしこれらの単語が相対的であった場合、コミュニケーションにおいて致命的な食い違いが発生する。また、これらの単語はあまりに基本的な部分で利用されるため話の文脈などからそれがどのようなものであるかを知ることはとても困難である。また、幼児の言語獲得においても先に習得される単語であり、人間の基礎概念と密接に結びつく単語であるため普遍性が高いのではないかと推察する。

グループ 2 に属する言葉は people, love, food, information 等で、グループ 1 と比べると内容がカテゴリーを表している名詞や、抽象的であり、概念的である言葉が多い。これらの言葉はグループ 1 のように話し手と聞き手が対象を確認しながらコミュニケーションを取る Naming Game のような状況ではなく、会話によってその意味を推測する単語である。ここに該当した単語は状況や話している集団によって使われ方が違い、コミュニケーションエラーをよく発生させるものであるが、それぞれの分野では暗黙的に使い方が決まっているケースが多い。これはコミュニケーションがその言葉の意味についての集団合意にいたらせた相対性の結果であると考えられる。

グループ 3 に属する言葉は justice ,beautiful ,freedom , cool ,god 等である。
このグループに属する言葉は感情や評価、価値観に関する単語が多い。これらの単語はその言葉がどう言う意味なのかは凡例を示すことは出来ても定義を示すことは出来ない。また、その人がいた環境や経験によって容易に変化し、とある個人にとっての凡例が他の人の受け入れられるものであるかは分からない。
これらの状況は相対性の典型的な発現例であると考える。

以上のように、各グループの特徴はそれに属する言葉の実際の意味を的確にカテゴリー化した結果となっている。

具体的な言葉、対象が明確な言葉　→　普遍性が高い
抽象的、概念的な言葉　→　相対性があるがコミュニケーションで合意を作る
感情、評価、価値観に関する言葉　→　相対性が高く、人によって意味が違う

これにより、言葉の意味と普遍性、相対性には一定の関係性の傾向があることが分かった。

6.1.3. 何が普遍性と相対性を生むのか

この研究であった二つの目的である、言葉の意味の相対性と普遍性の存在の確認、またそれらと言葉の特質がどのような関係性があるかは解明された。さらにここではコーパス交換時の距離の差に注目して分析を行う。

まずグループ 1 ではまず距離の差が問題にならない程小さい。これはほぼ単一の言葉の使い方が支配的に意味を決めているためである。
グループ 3 では簡単に言葉の意味の差が縮まるため逆に他のコーパスから距離が離れてしまい、なかなか収束できない。しかし、グループ 2 と 3 ではシミュレーション初期の距離はほとんど変わらない。いったいどのような違いが二つを分けているのであろうか。追加される情報は常に同じ数のファイル数なのでいくらかの誤差はあれ、そこに含まれる単語の量はある一定の範囲内である。
この研究では距離の計算を二つのコーパスにおいて、共起する上位 20 個の単語との共起確率をユークリッド距離によって算出している。そのため、一定量の新たな情報によって距離が大きく変化するためには新たに追加される単語の集合は様々な単語に均等に分布しているより少ない個所に集中して追加されたもの、つまり単語分布が偏ったもののほうが望ましい。

S=a+b+c+d……（a,b,c,d>0）の時
$S^2 > a^2 + b^2 + c^2 + d^2$.....

このシステムではより近い意味単語の頻度分布を持つコーパスがコミュニケーションを通して選ばれる。単に偏った単語を持つ情報が入っただけではコーパスの単語の意味は遠くなるために採用されない。偏った情報が採用されるためには次の状況のどちらかが必要である。

a,コミュニケーションの対象となるエージェントのコーパスの単語分布が偏っており、その分布状況と新規に入った単語の分布が似ている時

b,主体となるエージェントのコーパスの単語分布が偏っており、新規に入った単語がその偏りを修正する方向に偏っている時。

どちらの場合にしても、コーパス自体の情報の偏りが必要となる。

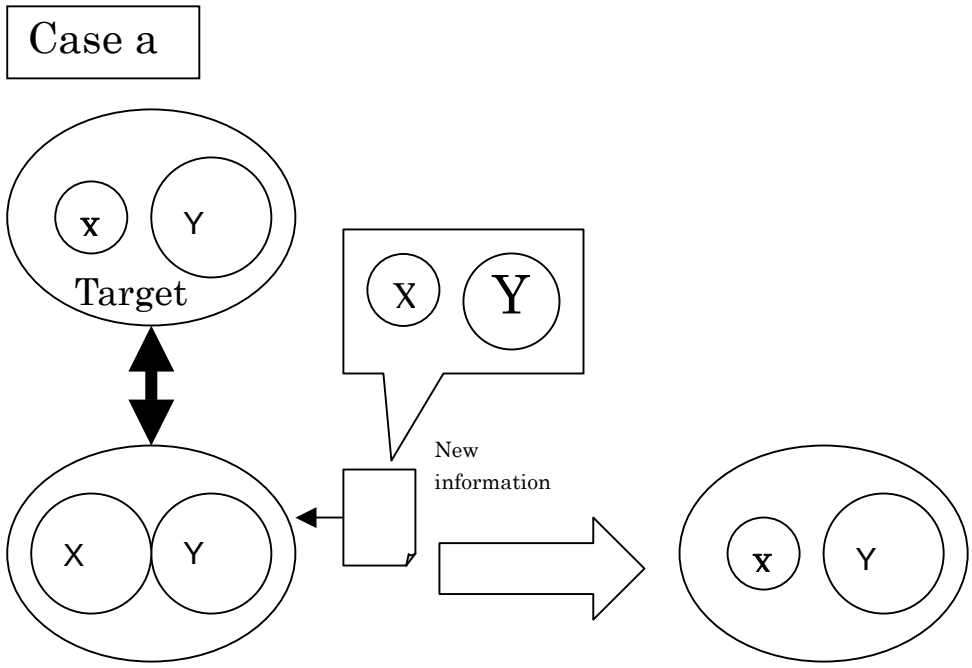


図 22: a のケースでのコーパスの変化

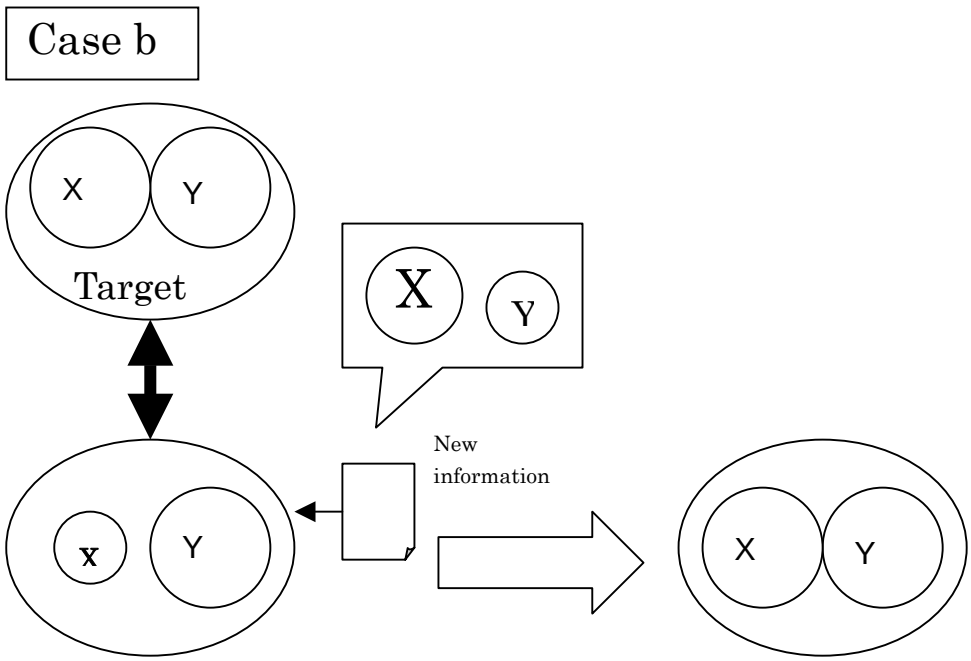


図 23: b のケースでのコーパスの変化

さらに、もしこのコーパスの偏りが単独のエージェントのものであるなら、コミュニケーションは b, のケースが連続して発生し偏りが解消されるはずである。

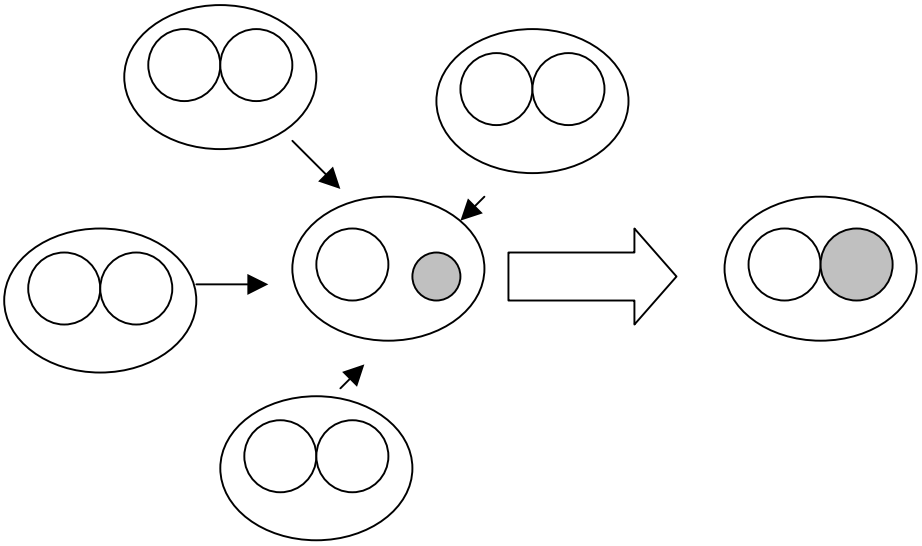


図 24: b が連続発生し偏りが解消

そのため、大きな差を発生させるコーパス交換を永続して発生させるためには何らかの形で a,の状態を維持させ続けなければいけない。そのためには、各エージェントのコーパス、そしてコーパスに新規追加される情報が複数の単語の使用法のどれかに偏っている状態が想定される。つまり、以下のようなステップが繰り返される。

1､エージェント *a1* は偏り *b1* の属しており、偏り *b2* を持つエージェント *a2* とコミュニケーションを取る。

2、*a1* に追加された情報 *inf1* の偏りが *b2* であった場合､*a1* は *inf1* によって偏りを *b2* に変える。

3､しかし別のところで別の偏りが *b1* であるエージェントとコミュニケーションを取り､且つその時に追加された情報の偏りが *b1* であれば新たに *b1* を偏りとするエージェントが発生する。

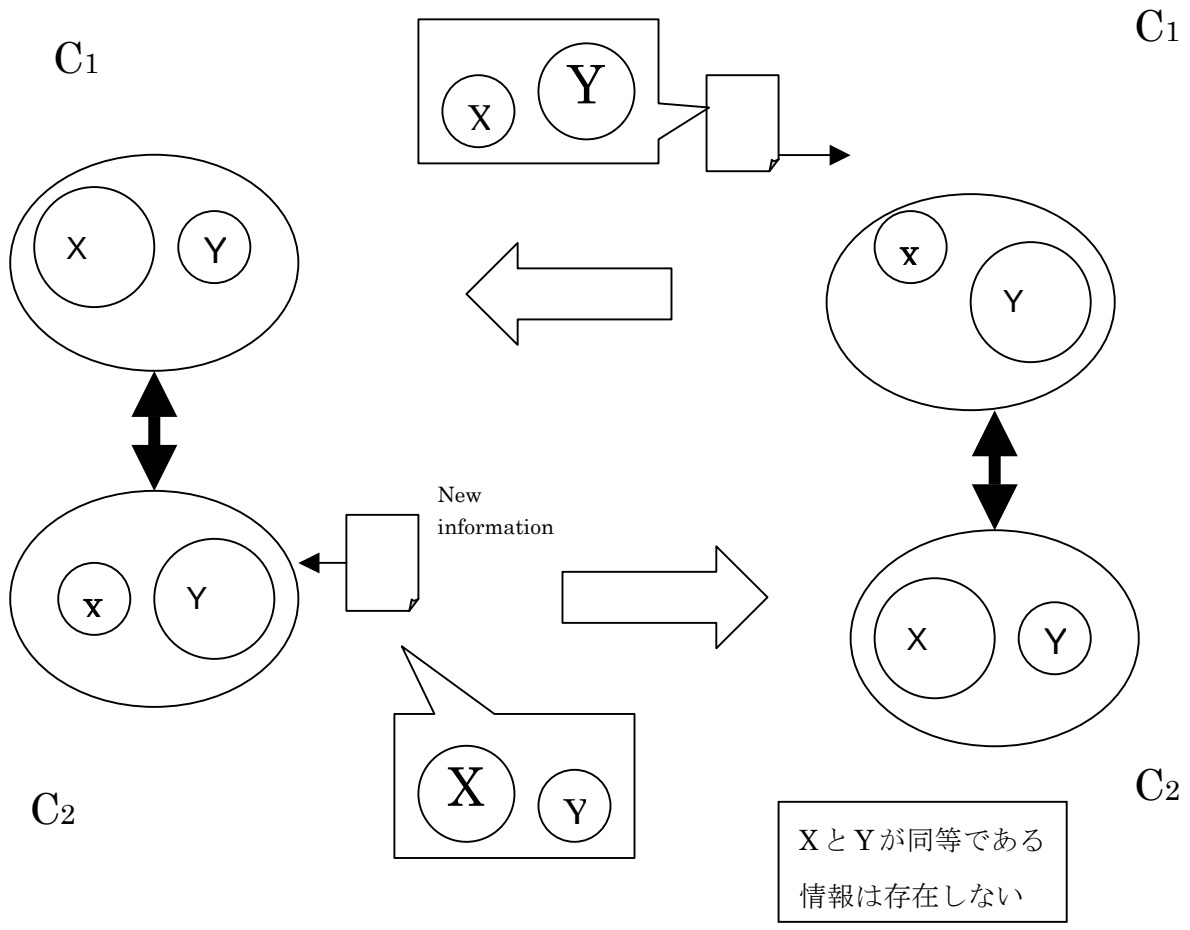


図 25: 相対性が維持されるコーパスのモデル

つまり、高い相対性はその言葉の意味が相反的な少数の利用方法において使用されていると言うことであり、言葉の意味が無秩序に定義されているわけではない。その事は、グループ 3 に該当する単語は意味があいまいな単語というより比較的相反しやすい人の価値観に関するものであったり（God、Justice, beautiful)、ほぼ同じくらい利用頻度のある意味がある単語(Mouse,動物とパソコン器具)であったりする所からも推察できる。

そして、グループ 3 ほど大きな距離の差がなかったグループ 2 は結果として言葉の偏りがそれほどなく、あったとしても b,のケースで偏りが解消されていったものと考え

られる。グループ 2 に該当する単語があいまいで様々な場面で利用されるが、専門的に使っている分野では明確な定義が出来ている言葉が多いのは、おそらく頻繁なコミュニケーションによって意味が収束し、定義が明確になったからだと考えられる。

6.2. 結論

実験と分析により、以下のことが達成されたことを確認した。

1, コーパスを使ったマルチエージェントシミュレーションを提案し、現実の言葉を使って言葉の意味の相対性と普遍性の存在を確認した。

2, 1,の実験結果を元に言葉の特質と相対性,普遍性を関連付け、言葉の特質と相対性・普遍性がどのような関係にあるかを分析した。その結果、人間が一般的に感じているあいまいな言葉の解釈に関する問題や価値観の衝突などは言葉の意味がシミュレーションによって解明され、現在使っている言葉の普遍性や相対性をある程度分類する目安を作った。

3, 限られた結果からの推論ではあるが、どのように普遍性と相対性が発生するかを検証し、言葉の利用法の偏りが主な原因であるとわかった。

6.3. 今後の課題

今回の研究において幾つかやりたかったが出来なかったこと、またはこの研究をさらに発展させるであろう指針などを記す。この研究に興味を持った人などがいたら参考にしてもらえるとありがたい。

6.3.1. 各エージェントの意味の分析

今回の研究では時間や分析方法への知識の問題で取得できたデータは各エージェントの他のエージェントの関係だけで、エージェント内部の意味分析を行うことが出来なかった。
エージェント間の距離は、エージェントにとっては大変分かりやすい評価指標になるが、これはエージェントにとっての距離であって、客観的な指標ではない。
やはり、全体的な分析方法によって各エージェントの意味がどのように分布しているかを測定する必要があったのではと考える。

言葉の意味はそれぞれ 20 項目の名前と値からなるデータで、内容は各エージェントごとに少しずつ違い、さらに変化もあるため、最終的には 1000 項目以上の多変量解析を行うことになるはずである。

6.3.2. コーパスを使う限界

今回の研究では合計 11G のコーパスを作成し、そこからエージェント用のコーパスを切り出して行った。一つのコーパスとしては 11G と言うサイズは分析にはかなり大きなものになり、各エージェントに割り当てられた 220M というサイズも十分なものであるが、何度も情報取得する対象としては同じ情報にあたってしまう可能性がかなり高いためさらに大きな情報源が必要であると感じた。
もしくは、今後の研究では Web などインターネットを自分で検索する機能を作り、自分で検索して意味を取得していくような形態にする必要があるかもしれない。

6.3.3. 意味分析における形態素解析

今回の研究では意味の測定は単純に一つの言葉に対する同じ単語の同フレーズ内で

の共起でのみ行った。しかし、実際の言語を考えると動詞に対する主語としての名詞か目的語としての名詞かで意味は大きく違い、また前置詞のかかり方でも変わってくるはずである。しかし、今回の研究ではその部分を技術的な問題から考慮に入れず、前置詞も下処理の段階で削除してしまった。

さらに正確さを期するためには意味測定には形態素解析によってその単語がどのような関係でその単語と共起したかを考慮した共起計算が必要になると思われる。

謝辞

今回の研究において、まず、Ho　Tu Bao 教授並びに知識創造方法論研究室の方々には大変にお世話になった。特に、直前まで方法論ばかり走ってしまい、一体何をやりたいのかわからない状態を耐えながら研究に対して辛抱強く助言と指導をしていただいた Ho 教授には言葉で言い尽くせない感謝をする。また、河崎さおり助手からは忌憚ない批評と一般の視点から見た客観的かつ冷静な意見をもらい、ともすると感覚的な部分で満足しがちな自分に喝を与えていただいた。
意味の比較についてのアイデアを提供してくれた Zang さん、研究環境整備の協力をしてくれた Dung さんにも感謝する。マルチエージェントが本業ではない研究室でくしくもマルチエージェント研究をともに行うことになった岸田君とは、全く違う分野の研究ながら多くの意見交換を行った。

また、他研究室では研究初期のまだ方向性が固まらない時期に、藤波努助教授、杉山公造教授、池田満教授に貴重な意見をいただき、大変に参考とした。また、思いついたアイデアを最初にぶつけて感想を聞く相手として川島君には大変世話になった。

最後に、今までのことをさておいて好きな研究をさせてくれた両親に尽くせぬ感謝したい。

[参考文献]

[Hashimoto 2004]言語進化とはどのような問題か? ～ 構成論的な立場から
Takashi Hashimoto (2004)　Proceedins of 18th Annual Conference of the Japanese Society for Artificial Intelligence (JSAI)

[Pinker 1995]Steven Pinker, Language Acquisition, Chapter to appear in L. R. Gleitman, M. Liberman, and D. N. Osherson (Eds.), An Invitation to Cognitive Science, 2nd Ed. Volume 1: Language, Cambridge, MA: MIT Press, 1995.

[1998 Yamadori] 山鳥重　人はなぜことばを使えるのか一脳と心の不思議、講談社,1998

[橋本 02c] 橋本敬: 構成論的手法, In: 杉山公造, 他(編), ナレッジサイエンス? 知を再編する 64 のキーワード, 紀伊国屋書店(2002) pp.132?135.

[金子・津田 96] 金子邦彦・津田一朗: 複雑系のカオス的シナリオ, 朝倉書店(1996).

[Axelrod 2005] Robert Axelrod　Advancing the Art of Simulation in the Social Sciences

[Steels 96] Steels, L: Self-Organizing Vocabularies, In:Langton, C and Shimohara, T (Eds.), Artificial Life　V, MIT Press (1996).

[Wittgenstein 1913]WITTGENSTEIN, Ludwig (1953) Tractatus logico-philosophicus suivi de Investigations philosophiques, [Pierre Klossowski], tel, Gallimard, 1961/1986.

[今井 1997] 今井むつみ（1997）『ことばの学習のパラドックス』共立出版

[清野　1997]　清野智昭：テクスト・データベースを用いたドイツ語形容詞の意味分析(熊本大学文学会「文学部論叢」第 55 号　文学篇) 1997

[Firth 1951]Firth, J.R.: 'Modes of meaning,'Essays and Studies, 1951.

[Harris 1970]Harris, Z.S.:"Distributional structure". In: Harris, Z.S.:Papers in structural and tranformational linguistics. Formal linguistic series. Vol.1, S.775-794.Dordrecht, 1970.

[SHANNON 1949] SHANNON, Claude E. & WEAVER, Warren (1949)『コミュニケーションの数学的理論』(The Mathematical Theory of Communication)[長谷川淳・井上光洋]明治図書出版, 1969.