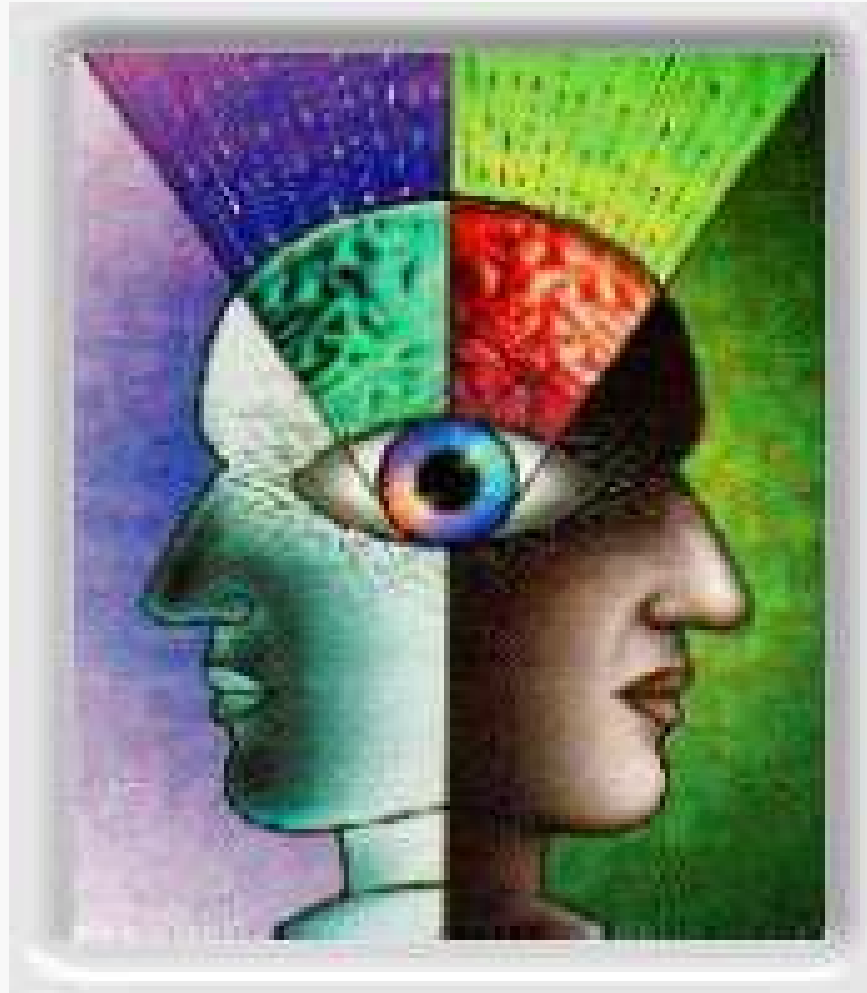


# Artificial Cognition™



Artificial Cognition, Inc.

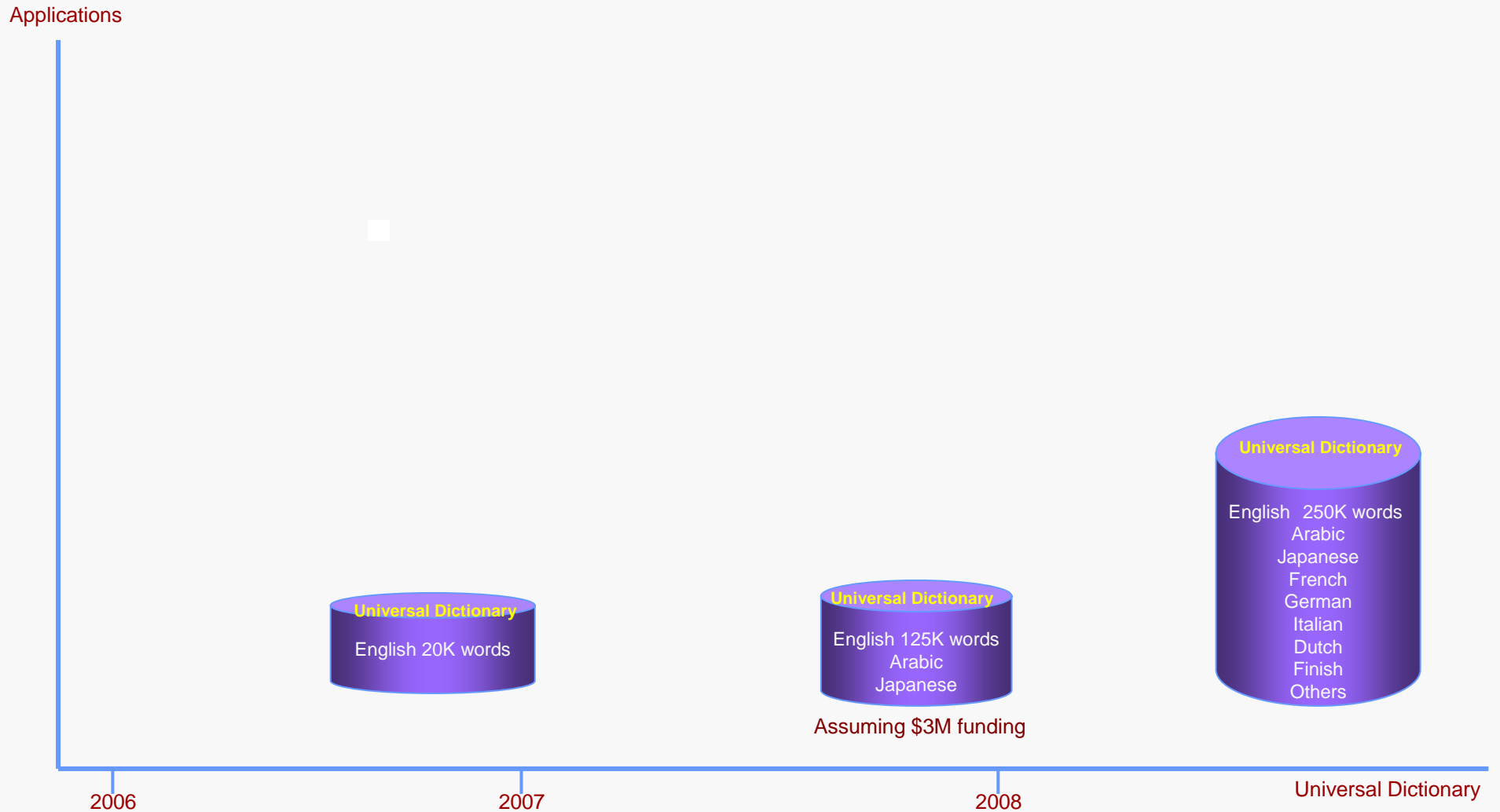
# Artificial Cognition, Inc. のご紹介

- **アーティフィシャル・コグニション (ACI: Artificial Cognition, Inc.) は、米国ニューハンプシャー州に本社がある会社です。 ACI社が開発したコンピューター用の辞書を用い、異なる二つの語が同じような意味をもつのか、あるいはそうでないのかをソフトウェアの数値演算によって決める技術 (特許申請) をもっています。**
- **ACIの目標は、コンピューターが数値データを扱うのと同じような容易さで、コンピューターが非数値的な概念的データを扱う方法を開発することです。**
- **中核をなす技術は、言葉 (語) を表現する方法です。語は、いくつかの領域に分けられた言葉の定義バイトで表現されます。それぞれの領域は、数値的演算で、それぞれの語が、類似しているのか、一部違っているのか、あるいは派生語なのかなどを決める情報を持っています。**

# Artificial Cognition, Inc. のご紹介

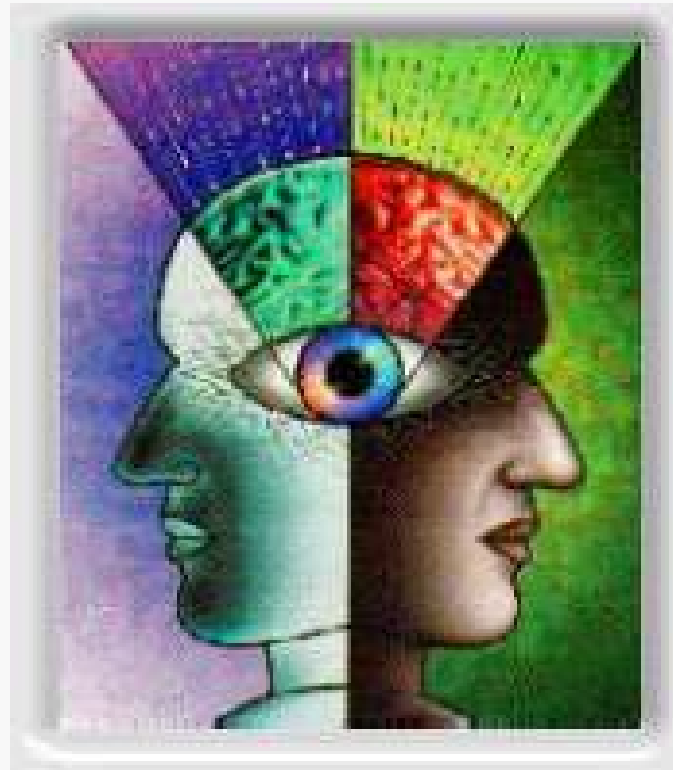
- この技術は、検索エンジンの前処理プロセッサ (Search engine front end processors)”, 文脈によるスペルチェッカー (Context sensitive spell checker), 会話の聴取・分析 (Communications monitor and analysis), 自然言語翻訳 (Natural language translators) などの多くの言語用途への適用が期待できます。
- 
- ACI社は、この技術を用いた商用の英語辞書の開発を進めており、さらに日本語の辞書も開発する予定です。
- ACI社は、辞書の構築、アプリケーションの開発のための日本からの投資、あるいは日本でのパートナーを探しております。

# Artificial Cognition™ Roadmap



# Artificial Cognition™

The technology in use



# Word/Text Processing, Parser

High school students are subjected to analogy questions as part of the SAT college entrance exams. Instead of logic, our software compares the differential of each pair of words numerically using ACI universal dictionary. After comparing all 5 sets, the differential that is closest to the original differential is the correct answer.

• **DALMATIAN : DOG ::**

• (A) oriole : bird

(B) horse : pony

(C) shark : great white

(D) ant : insect

• (E) stock : savings

• **CUB : BEAR ::**

(A) piano : orchestra

(B) puppy : dog

• (C) cat : kitten

(D) eagle : predator

• (E) fork : utensil

• **DOCTOR : HOSPITAL ::**

• (A) sports fan : stadium

(B) cow : farm

(C) professor : college

(D) criminal : jail

• (E) food : grocery store

• **6.) AUTOMOBILE : MOTORBOAT ::**

(A) gas generator : water wheel

(B) balloon : raft

(C) freight train : ocean liner

• (D) rickshaw : rowboat

(E) windmill : sailboat

# Word/Text Processing, Parser

The Correct Answer (according to *our* math) is highlighted in blue. How did you do?

Conventional wisdom says that this type of work cannot be done by computers at all. The second thing is to ask, "How quickly could you do it?" ACI's technology allows the computer to find the answer to an analogy problem such as these in less than a millionth of a second."

• DALMATIAN : DOG ::

- (A) oriole : bird
- (B) horse : pony
- (C) shark : great white
- (D) ant : insect
- (E) stock : savings

• CUB : BEAR ::

- (A) piano : orchestra
- (B) puppy : dog
- (C) cat : kitten
- (D) eagle : predator
- (E) fork : utensil

• DOCTOR : HOSPITAL ::

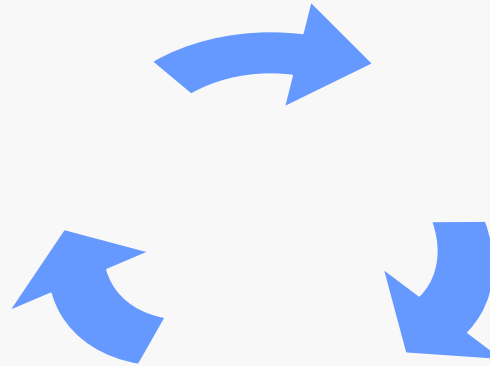
- (A) sports fan : stadium
- (B) cow : farm
- (C) professor : college
- (D) criminal : jail
- (E) food : grocery store

• 6.) AUTOMOBILE : MOTORBOAT ::

- (A) gas generator : water wheel
- (B) balloon : raft
- (C) freight train : ocean liner
- (D) rickshaw : rowboat
- (E) windmill : sailboat

# Natural Language Translations

The art of translation, what occurs when an unknown word is encountered



1. A Japanese caller speaks a sentence containing the word “Tomesode”
2. Her cellphone translates this into an intermediate computer language
3. The English speaking users’ cellphone translates this from the intermediate language to English

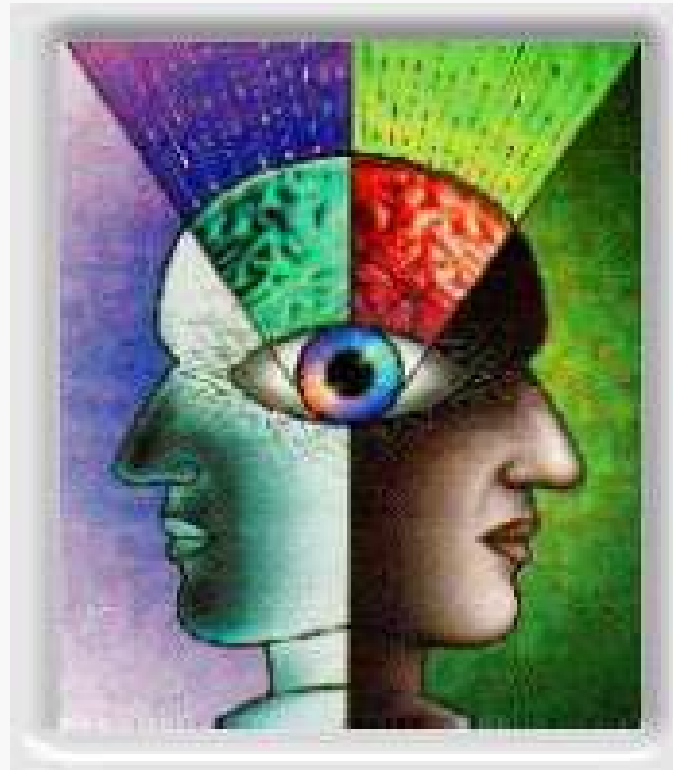
A search of the database finds no match in the English language for the equivalent of the word Tomesode.

- The database software analyzes the definition and finds that it means “formal kimono for married woman”
  - The receiving cell phone finishes the translation and speaks



# Artificial Cognition™

Comparison between ACI vocabulary and Wordnet



# ACI VocabularyとWordnetの比較

## 1. 歴史と目的

Wordnet	ACI Vocabulary
<p>Wordnetは、プリンストン大学心理学学科にて、心理学と言語資料の研究を目的に開発された。この目的のために作成されたデータをコンピュータを用いて参照するものだが、コンピュータの元来の処理能力について特別に配慮されたものではない。従来、同様の単語リストが僅かしか存在しなかったため、Wordnetが人工知能等に利用されてきた経緯がある。</p>	<p>ACI Vocabularyは、コンピュータが単語の持つ概念を理解できるように開発された。コンピュータの機械語がデータの比較や演算を行えるよう、具体的に設計されている。恐らく、コンピュータがACI技術を用いて単語の概念を理解するのに必要な能力は、コンピュータが浮動小数点数を理解するために要する能力とほぼ同等といえる。(注参照)</p>

# ACI VocabularyとWordnetの比較

注:

コンピュータは何かを理解することができる、と考えるべきである。コンピュータは数字のみしか理解できないものである、とよく言われる。コンピュータの数値演算ユニット(ALU)は通常、ビットか2の補数の整数のために設計されたオペレーションを有している。これらのフォーマットに固有のオペレーションがあるということは、コンピュータがこうしたフォーマットを比較・演算して意味のある結果を出すことができることを意味する。コンピュータに数字「2」のビット・パターンを与えたとしても、コンピュータは数字の「2」が象徴する二重性という概念を理解しているわけではない。2を3と比較して2の方がより小さい数であることを示す値を返すことはできるが、「小さい」が意味するところの概念は持っていない。また、2と3を足すことはできるが、足すということの意味は理解していないし、演算の対象となる数値や計算結果の実際の意味も理解していない。これが、コンピュータにとって数字を「理解」するということである。つまり、適切な記憶フォーマットが与えられているという前提のもと数字を比較したり演算したりする能力は備えている。浮動総小数点数の処理は、ALUの整数よりも若干手間を要するが、浮動小数点演算装置を用いれば浮動総小数点数の演算は自然に行われる。

ACIの技術を用いれば、コンピュータは「カラス」と「鳥」の意味を比較し、前者が後者の下位語であることを求めることができる。「カラス」と「ワシ」を比較して、両者が同じグループのメンバであることを求めることができ、それを「鳥類」というグループとして認識することができる。さらに、「ライオン」と「ワシ」の意味の違いを求め、その違いが「自動車」と「小型旅客飛行機」の違いとほぼ同じであることさえ求めることができる。こうした演算はALUに若干の負荷をかけるが、浮動小数点数用のものと類似した処理装置を開発することが可能であり、そうすればこうした演算は1又は2クロックサイクルにて行える。我々は、これが、コンピュータが数字を「理解する」と言われてきたと同じように単語の概念の「理解」を構成するものであると主張している。

# ACI VocabularyとWordnetの比較

## 2. 機能

Wordnet	ACI Vocabulary
<p>Wordnetは、単語を「synset」と呼ばれる同義語のグループに分けて分類を行う。上位語、下位語やある語の部分を表す語等の単語の関係に関する情報が、これらsynset毎に構築されている。個別の単語はASCII形式で定義されており、コンピュータが理解できるように作成されているのではない。利点としては、単語がグループ化されていて、各グループの単語の関係が明確なことである。</p>	<p>ACIは、単語をコンピュータが理解可能な形で定義する。上位語・下位語・ある語の部分を表す語の関係は同様に存在するが、これらの情報は各々の単語の定義そのものに簡潔に包含されている。他の分類も維持されており、通常ひとつの定義についてより多数の分類が存在する。また、Wordnetでは不可能な様々な方法で、単語を操作することができる。</p>

# ACI VocabularyとWordnetの比較

## 3. オペレーション

Wordnet	ACI Vocabulary
<p>Wordnetでは、単語の理解はデータベース上に広がる相関関係情報によって維持されている。こうした関係を調べるためには、コンピュータの記憶装置を参照することを要する - これは通常はディスクへ何度も照会することを意味する。各単語のグループの関係情報への到達手段に関する動作を除いては、Wordnetは単語の意味をオペレーション可能なフォーマットで保存していない。</p>	<p>ACIでは、単語に関する理解は単語を記憶するフォーマットそのものに組み込まれている。単語の意味は、コンピュータのレジスタに適合し、普通の機械命令を使って操作できるように設計されている。ACI技術は、意味による違いの判別や関連する共通点の比較など、単語の意味に基づく様々な処理が可能である。</p>

# ACI VocabularyとWordnetの比較

## 4 . 文脈情報

Wordnet	ACI Vocabulary
Wordnetのデータベースには、文脈情報は備わっていないと思われる。	ACIのデータベースには、各語の定義に豊富な文脈情報が備わっている。

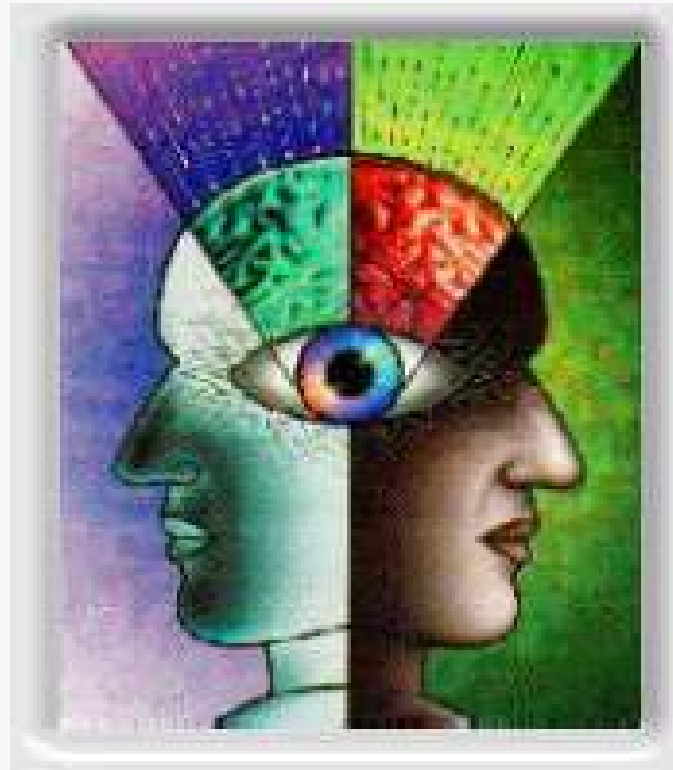
# ACI VocabularyとWordnetの比較

## 5. サイズと費用

Wordnet	ACI Vocabulary
<p>Wordnetは20年間をかけて開発された。最新の統計によれば、Wordnetには語の意味の関係が207,016あり、それらが117,597のsynsetの中の155,327の文字列と関係している。プリンストン大学の学生・教員が協力して開発したもので、無償労働も含まれたはずである。300万ドルの補助金を使用した。</p>	<p>ACI Vocabularyの作成には、およそ1.5人年を要した。ACIには、約50,000の単語対ルール関係があり、それらが約11,000の文字列と約10,000のルールに結び付いている。今までのところ外部からの財政的支援は受けていない。今年中に当面の財政的支援なしにデータベースを150%拡大するというゴールを掲げているが、それは妥当なものと考えている。</p>

# Artificial Cognition™

## Demonstration



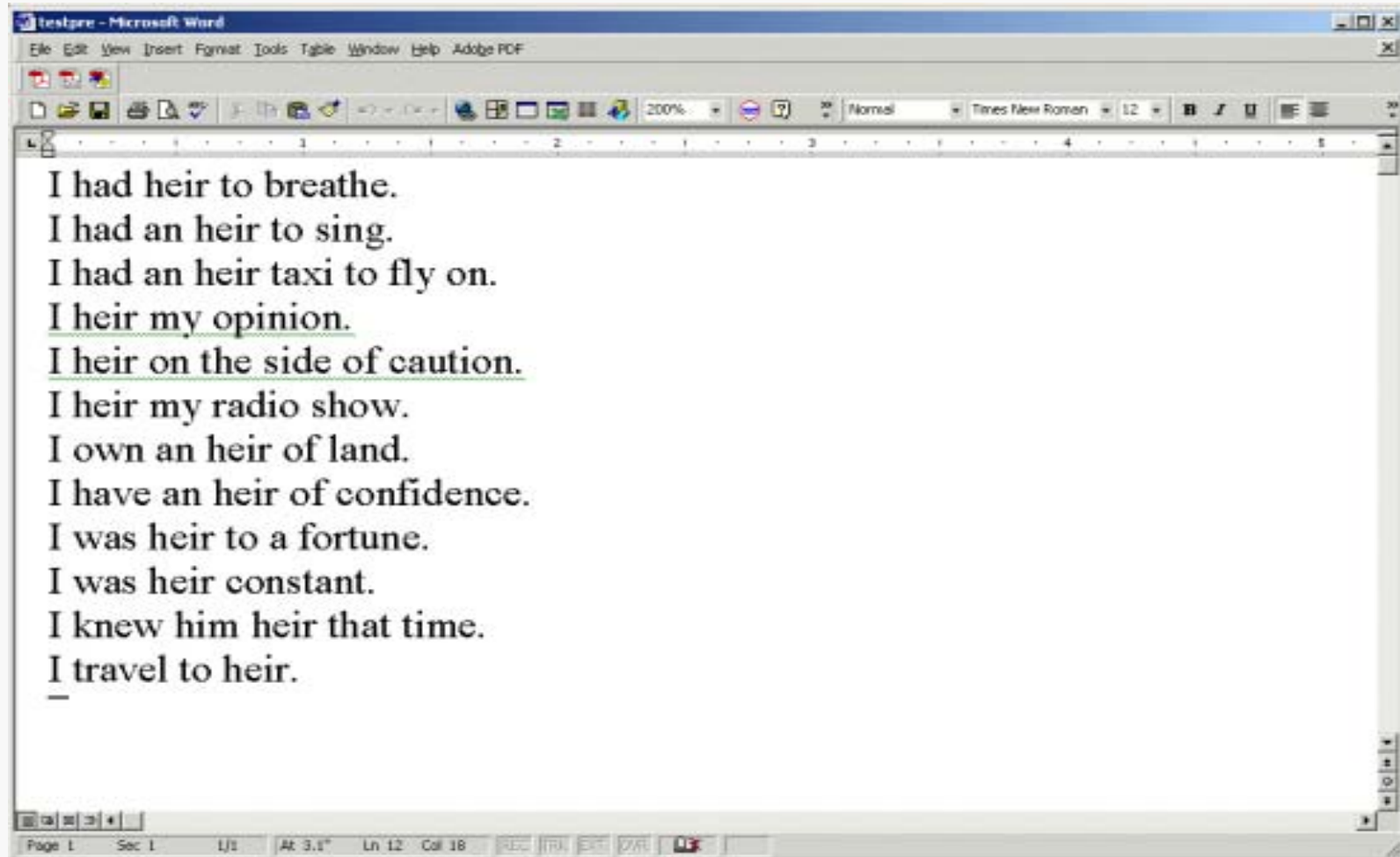


# Artificial Cognition™ デモ

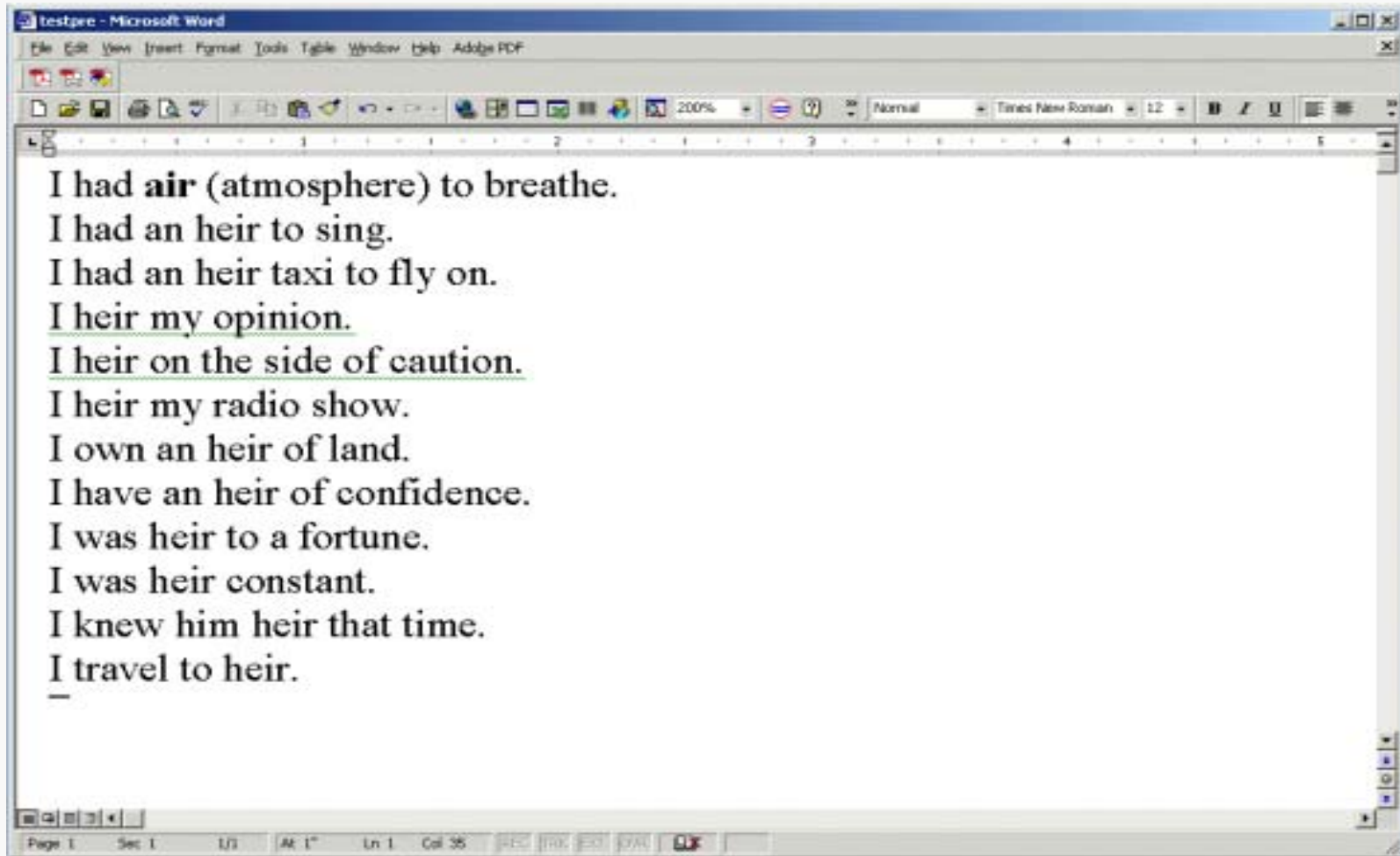
## Artificial Cognition™ は、曖昧性を除去します

- 今日のコンピューターにその定義が記憶されている語彙は、それぞれの本来の意味との相対的近さや関係を示す手がかりを持っていません。
- 語彙の定義が量的に与えられ、それがコンピューター上で一算術・論理演算で一操作できるシステムを想像してください。
- このシステムでは、概念をデジタル形式で表現することを開発したので、コンピューターは概念を理解することが出来ます。単語airが異なる文脈で使われ、さらにミススペルのある11の例文を例にとって、文脈の意味に基づいてスペルチェックを行います。

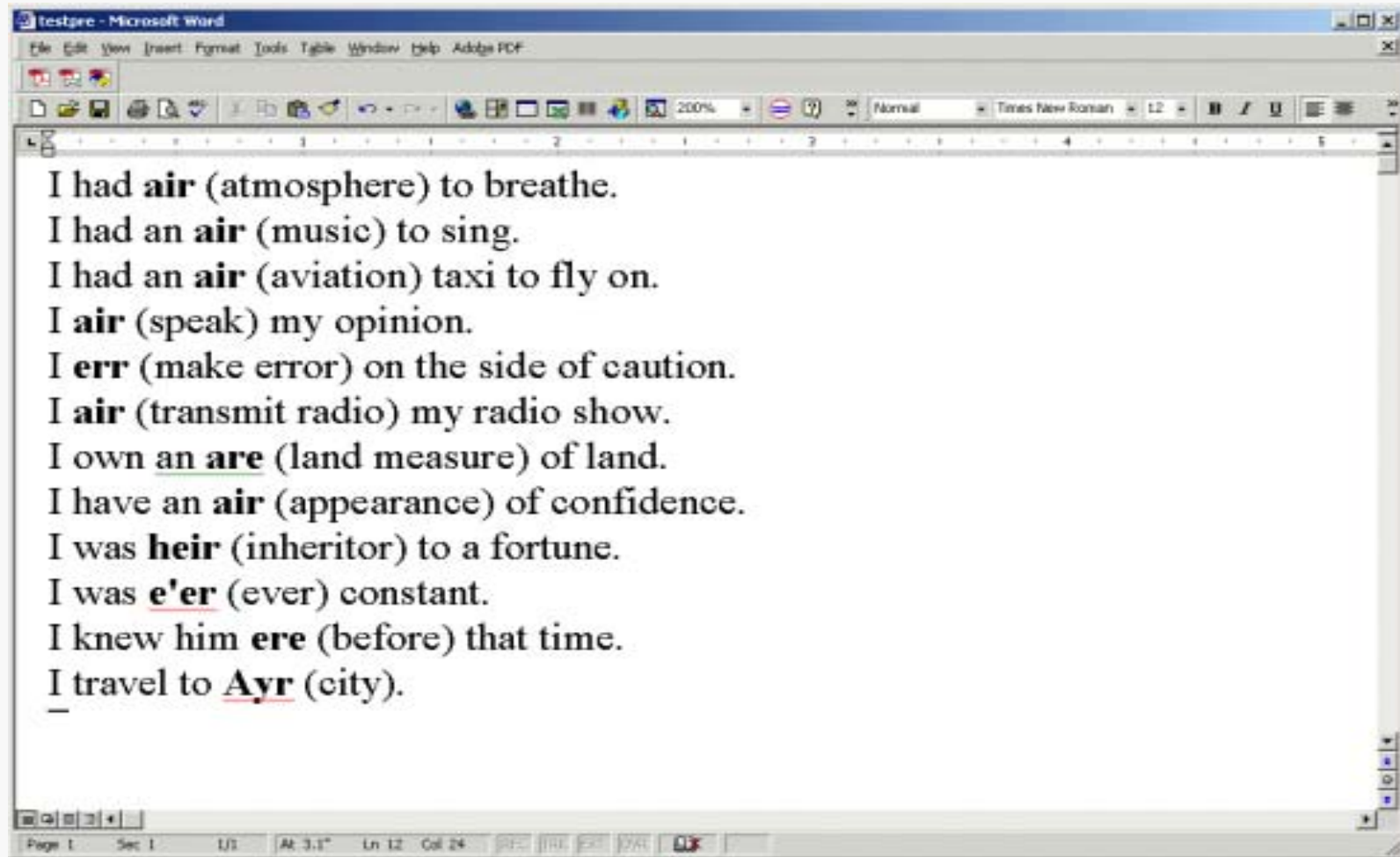
# WRONG!



# Line 1 Corrected



# Line 12 Corrected, Try That Microsoft!



# Comparison between ACI vocabulary and Wordnet:

## 1. History and Purpose

¶ Wordnet was developed by the Psychology department of Princeton University for the study of psychology and linguistic data. It uses computers to track the data it prepared for those purposes, but without special regard to the native capacities of computers. Historically it has been used for artificial intelligence, etc. because there are few similar such lists of words.

¶ ACI's vocabulary was developed so computers could have a native understanding of concepts. It is designed specifically to allow the machine language of computers to compare or operate on data. Arguably, the ability of the computer to understand concepts using ACI's technology is approximately the same as the ability of the computer to understand floating point numbers. (See note.)

## 2. Function

¶ Wordnet categorizes words by placing them in "symsets," sets of synonyms. Hypernym, hyponym, and meronym relationships are built for the sets. Definitions of words are in ASCII form and are not understandable by the computer. The advantage is that the words are grouped and relationships of groupings are explicit.

¶ ACI defines words in a form that is understandable by the computer. The hypernym, hyponym, and meronym relationships are present, but they are concisely contained in the definitions themselves. Other groupings are maintained, and there is typically a greater number of them per definition. Additionally, the words can be operated on in a variety of ways not available with Wordnet.

## 3. Operation

¶ Wordnet maintains its understanding of words in relationships spread through a database. Examining these relationships requires tracking through storage memory - usually meaning a series of references to disk. Aside from actions relating to maps to the sets relationships, Wordnet does not store meaning in operable format.

¶ ACI maintains its understanding of words in the format of the word storage itself. Word meanings are designed to fit into computer registers and be operated on using ordinary machine instructions. ACI's technology can do a variety of functions based on meaning, such as finding meaning-based differences and comparing related analogies.

## 4. Contextual Information

¶ Wordnet's database appears not to have any contextual information built in.

¶ ACI's database as rich contextual information built in to the definitions of each word.

## Comparison between ACI vocabulary and Wordnet:

### 5. Size and cost

¶ Wordnet has been developed over a period of twenty years. Its latest statistics say it has 207,016 word-sense relationships, related to 155,327 strings in 117,597 synsets. It was developed by a combination of labor of students and faculty at Princeton, some of which was doubtless unpaid, but has used \$3 million in grant money.

¶ ACI's vocabulary has required approximately one and a half man-years to produce. It has approximately 50,000 word-rule relationships, related to approximately 11,000 strings and about 10,000 rules. It has not received outside funding as yet. It is considered a reasonable goal to increase the database by 150% by year's end without immediate funding.

### NOTE:

We should consider the idea that a computer can understand something. It is often said that computers natively understand only numbers. The arithmetic-logic unit (ALU) of a computer typically has operations that are designed for bits and twos-compliment integers. The fact that there are native operations for these formats means that the computer can compare and operate on them producing meaningful results. The computer, given a bit pattern for the number 2, does not have a concept of duality. It can compare 2 with 3 and return a value indicating that 2 is the lesser number, but it does not have a concept of what "less" means. It can add 2 and 3, but it does not understand what it means to add, let alone the actual meanings of operands or results. This is what it is for a computer to "understand" number: it has a native ability to compare and operate on them, given the appropriate storage format. Floating point numbers require a little more work than integers in an ALU, but floating point units operate on them natively.

Using ACI's technology, the computer can compare the meanings of "crow" and "bird" and find that the former is a hyponym of the later. It can compare "crow" and "eagle" and find that both are members of the same set, which it can identify as the set of all birds. It can even find the difference between the meanings of "lion" and "eagle," and see that this difference is nearly identical to the difference between "car" and "small passenger plane." These operations require a little work in the ALU, but a processing unit similar to that for floating point numbers could be built, and would perform these operations in one or two clock cycles. We argue that this constitutes "understanding" concepts in the same way that computers have been said to "understand" numbers.