# Towards a Formal Framework of Cognitive Linguistics

Yingxu Wang [1] and Robert C. Berwick [2]

[1] International Institute of Cognitive Informatics and Cognitive Computing (IICICC)
Laboratory for Cognitive Informatics and Cognitive Computing
Dept. of Electrical and Computer Engineering, Schulich School of Engineering
University of Calgary
2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4
Email: yingxu@ucalgary.ca

[2] Dept. of Brain and Cognitive Science and Dept. Electrical Engineering and Computer Science
Massachusetts Institute of Technology
77 Massachusetts Ave., Cambridge, MA 02139, USA
Email: berwick@csail.mit.edu

**Abstract:** Cognitive linguistics is an emerging discipline that studies the cognitive properties of natural languages and the cognitive models of languages in computing linguistics, cognitive computing, and computational intelligence. This paper presents the theoretical framework of cognitive linguistics in order to systematically formalize the syntaxes and grammars of natural languages. An abstract language model of cognitive linguistics is created at the top level. Based on it, the cognitive structures of languages at the levels of lexis, phrase, clauses, sentence, paragraph, and article are formally modeled from the bottom up. Using contemporary denotational mathematics, the deductive grammar of English is formally modeled and rigorously analyzed. This basic research provides support for a wide range of applications in computational linguistics, cognitive informatics, online text processing, web search engines, machine language comprehension, autonomous machine learning, smart cell phones, semantic computing, and computing with words.

**Keywords:** *Cognitive informatics, cognitive computing, cognitive linguistics, natural language processing, lexical models, sentence models, paragraph models, article models, deductive grammar of English, deductive syntax, deductive semantics, concept algebra*

## 1. Introduction

It is recognized that the development of formal models of syntaxes, semantics, and grammars of natural languages is a fundamental challenge in cognitive linguistics. The mathematical models of natural languages are not only required in cognitive informatics, natural language processing, and computational linguistics in general, but also needed in word processing, web search engines, machine language comprehension, autonomous machine learning, smart cell phones, semantic computing, cognitive computing, and computing with words in particular. Without the mathematical models of grammars and linguistics, the empirical language knowledge of humans cannot be systematically and precisely conveyed to cognitive machines and systems, because the syntaxes and semantics of languages might not be exhaustedly enumerated by examples.

Studies on formal syntaxes and semantics are initiated in linguistics and natural language processing, which can be traced back to the works of Alfred Taski (1944) and Noam Chomski (1956). Languages are an oral and/or written symbolic system for thought, expression, and communications, reasoning, and knowledge representation. Linguistics is a discipline that studies the structures and properties of human or natural languages [Crystal, 1987; Pullman, 1997; Huddleston & Pullum, 2002; Wang, 2009a; Wardhaugh, 2006; O'Grady & Archibald, 2000]. Chomski formalized a general language structure known as the *universal grammar* [Chomski, 1956, 1957, 1959, 1962, 1965, 1982, 2007]. Then, the generalism and formal studies on syntactical structures have been proliferated. *Cognitive linguistics* is proposed by Gibbs et al. since 1996, which attempts to explain the cognitive processes of language and knowledge acquisition, storage, production, and comprehension [Gibbs, 1996; Taylor, 2002; Wang, 2003, 2007b, 2012a; Langlotz, 2006; Evans & Green, 2006]. Wang and his colleagues introduce the *deductive grammar* [Wang, 2009a] and *deductive semantics* [Wang, 2010b] in cognitive linguistics [Wang, 2007b, 2012a, 2012b] and cognitive informatics [Wang, 2002, 2003, 2007a, 2007b, 2007c, 2009b, 2009c, 2010a; Wang & Wang, 2006; Wang et al., 2006, 2009a, 2009b, 2011b]. A *concept algebra* based technology [Wang, 2008b, 2010c] is developed for rigorous semantic modelling and analysis [Wang, 2011a]. Then, a new paradigm of *denotational mathematics* [Wang, 2008a] known as *inference algebra* is created [Wang, 2011] for machine reasoning based on language knowledge bases [Wang, 2012b].

**Definition 1.** *Cognitive linguistics* is an emerging discipline that studies the cognitive properties of natural languages and the cognitive models of languages in cognitive computing and computational intelligence.

In order to rigorously explain the cognitive mechanisms of languages, syntaxes and semantics of a given language as well as their relations and interactions have to be systematically studied. In a language, syntaxes deal with relations and combinational rules of words in sentence formation; while semantics embody meanings of words and sentences. Syntactic and semantic analyses in linguistics rely on a set of explicitly described rules known as the grammar of the language. Therefore, contemporary linguistic analyses focus on the study of grammars, which is centered in language acquisition, understanding, and interpretation.

**Definition 2.** The *grammar* of a language is a set of rigorously specified rules that integrates phonetics, phonology, morphology, syntax, and semantics of the language.

Grammars govern the articulation, perception, and patterning of speech sounds, the formation of phrases and sentences, and the interpretation of utterance. However, the formalization of grammars by mathematical models is not only a fundamental need, but also a significant challenge in cognitive and computational linguistics.

This paper attempts to demonstrate that natural languages can be systematically formalized by using denotational mathematics and cognitive informatics methodologies. In the remainder of the paper, Section 2 formally describes the discourse of natural languages and its mathematical model at the top level. Section 3 analyzes and models the formal lexical and syntactical structures of English. Section 4 develops the general pattern of sentence structures and the deductive grammar of English in a rigorous approach. As a result, a rigorous theoretical framework of cognitive linguistics is established towards the applications in computational linguistics, cognitive informatics, online text processing, web search engines, machine language comprehension, autonomous machine learning, smart cell phones, semantic computing, and computing with words.

## 2. The General Mathematical Model of Natural Languages

A formal model of a general language such as English can be described as an abstract language using denotational mathematics [Wang, 2008a, 2011]. Based on the general abstract language model, the structural models of natural languages can be created at the lexis, phrase, clause, sentence, paragraph, and article levels from the bottom up.

**Definition 3.** The *abstract language,* $\mathfrak{L}$, is a 5-tuple, i.e.:

$$\mathfrak{L} \triangleq (\mathfrak{E}, \mathfrak{X}, \mathfrak{W}, \mathfrak{R}, \mathfrak{S}) \quad (1)$$

where

- $\mathfrak{E}$ is a finite ordered set of alphabet of the language, $\mathfrak{E} \sqsubset \mathfrak{L}$;
- $\mathfrak{X}$ is a power set of *lexical relations* between the letters in the alphabet, $\mathfrak{X} \subset \text{Þ}\mathfrak{E} \sqsubset \mathfrak{L}$;
- $\mathfrak{W}$ is a finite nonempty set of *words* that are identified strings of the alphabet in the language, $\mathfrak{W} = \{\text{Þ}\mathfrak{E} \mid \mathfrak{X}\} \sqsubset \mathfrak{L}$. A subset of $\mathfrak{W}$ is known as the *primitive words* ($\mathfrak{W}^0$), $\mathfrak{W}^0 \subset \mathfrak{W}$, that directly represents real-world entities, proper names, meta-behaviors, and abstract concepts that cannot be further deduced onto more primitive concepts or behaviors;
- $\mathfrak{R}$ is a power set of *syntactic relations* between words, $\mathfrak{R} = \text{Þ}\mathfrak{W} \times \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}$;
- $\mathfrak{S}$ is a power set of *semantic relations* between words, $\mathfrak{S} = \text{Þ}\mathfrak{R} \rightarrow \text{Þ}\mathfrak{R}^0 = \text{Þ}\mathfrak{W} \times \text{Þ}\mathfrak{W} \rightarrow \text{Þ}\mathfrak{W}^0 \times \text{Þ}\mathfrak{W}^0$, $\mathfrak{R}^0 \subset \mathfrak{R} \sqsubset \mathfrak{L}$.

The formal grammar for a language can be categorized as the syntactical and semantic grammars. The former are grammars specifying the structural rules of the language; while the latter are grammars specifying the functional rules of the language. A special part of the semantic grammar of natural languages is the set of words, or the vocabulary, which can be classified as terminals and nonterminals.

**Definition 4.** The *vocabulary of a language*, $\mathscr{V}$, is a finite nonempty set of words that represent certain meanings or convey specific semantics, i.e.:

$$\mathscr{V} = \mathscr{T} \cup \overline{\mathscr{T}}, \ \mathfrak{W} \subseteq \mathscr{V} \sqsubset \mathfrak{L} \quad (2)$$

where $\mathscr{T}$ and $\overline{\mathscr{T}}$ denote the sets of terminals and nonterminals, respectively, in the languages.

**Definition 5.** The *terminals of a language*, $\mathscr{T}$, is a finite nonempty subset of $\mathscr{V}$ in which the vocabulary fall into the following special primitive categories, i.e.:

$$\mathscr{T} \triangleq \{\mathscr{V}_0 \mid \mathscr{V}_0 \subseteq (N_{e0} \cup N_{a0} \cup V_0 \cup \mathscr{M}_0) \subset \mathscr{V}\} \quad (3)$$

where $N_{e0}$, $N_{a0}$, $V_0$, and $\mathscr{M}_0$ denote words that represent real-world entities, meta-abstract concepts, meta-behavioral verbs, and common modifiers, respectively.

**Definition 6.** The *nonterminals of a language,* $\overline{\mathscr{T}}$, is a subset of $\mathscr{V}$ in which the vocabulary can be defined or derived by the terminals, i.e.:

$$\overline{\mathscr{T}} = \mathscr{V} \setminus \mathscr{T} \quad (4)$$

The abstract language $\mathfrak{L}$ as given in Definition 3 explains the universal discourse of languages as the contexts of linguistics. On the basis of the abstract language model $\mathfrak{L}$, the fundamental structural properties of natural languages at the lexis, phrase, clause, sentence, paragraph, and article levels can be rigorously modeled and analyzed in the following sections.

## 3. The Formal Lexical Structures of English

A set of three categories and 19 lexical and syntactic elements in language $\mathfrak{L}$ is elicited and summarized in Table 1. The lexes are a category of syntactic units that represent the six basic parts of speech. The modifiers are a category of syntactic units that that refines and restricts the semantics of words. The phrases are a category of syntactic units that compose multiple words into a complex unit.

**Definition 7.** The *set of lexical elements* $\mathcal{E}$ of a sentence $\mathcal{S}$ in language $\mathfrak{L}$ can be classified into the categories of *lexes* ($\mathcal{L}$), *modifiers* ($\mathcal{M}$), and *phrases* ($\mathcal{P}$), i.e.:

$$\begin{aligned}\mathcal{E} &\triangleq (\mathcal{L},\, \mathcal{M},\, \mathcal{P}) \\ &= \{\mathbb{N}, \mathbb{X}, \mathbb{V}, \mathbb{A}, \mathbb{D}, \mathbb{P}\} \\ &\quad || \{\tau, \kappa, \delta, \alpha, \neg, \gamma, \sigma\} \\ &\quad || \{NP, VP, AP, DP, PP, CP\}\end{aligned} \tag{5}$$

Each of the lexical elements $\mathcal{E}$ in English is formally modeled and explained in the following subsections.

### 3.1 Parts of Speech

**Definition 8.** The *lexes*, $\mathcal{L}$, in language $\mathfrak{L}$ encompass six categories of lexical roles known as the parts of speech such as noun, pronoun, verb, adjective, adverb, and preposition where each category can be formally modeled as a specific set of words in the lexical categories $\mathcal{N}$, $\mathcal{X}$, $\mathcal{V}$, $\mathcal{A}$, $\mathcal{D}$, and $\mathcal{P}$, respectively, in $\mathfrak{L}$, i.e.:

$$\begin{aligned}&\mathbb{N} \triangleq \{n \in \mathcal{N} \,|\, \mathcal{N} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\ &\mathbb{X} \triangleq \{x \in \mathcal{X} \,|\, \mathcal{X} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\ &\mathbb{V} \triangleq \{v \in \mathcal{V} \,|\, \mathcal{V} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\ &\mathbb{A} \triangleq \{a \in \mathcal{A} \,|\, \mathcal{A} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\ &\mathbb{D} \triangleq \{d \in \mathcal{D} \,|\, \mathcal{D} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\ &\mathbb{P} \triangleq \{p \in \mathcal{P} \,|\, \mathcal{P} \subset \mathfrak{W} \sqsubset \mathfrak{L}\}\end{aligned} \tag{6}$$

Table 1. The Syntactic Categories of Linguistics

| No. | Category | Subcategory | Symbol | Description |
|---|---|---|---|---|
| 1 | **Lexis** | | $\mathcal{L}$ | ***Basic parts of speech*** |
| 1.1 | **(Part of speech )** | Noun | $\mathbb{N}$ | Entities, abstract concepts, and behaviors |
| 1.2 | | Pronoun | $\mathbb{X}$ | Representation of nouns |
| 1.3 | | Verb | $\mathbb{V}$ | Actions, states, and possessions |
| 1.4 | | Adjective | $\mathbb{A}$ | Properties of a noun |
| 1.5 | | Adverb | $\mathbb{D}$ | Properties of a verb |
| 1.6 | | Preposition | $\mathbb{P}$ | Relations between two words or phrases |
| 2 | **Modifier** | | $\mathcal{M}$ | ***Semantic refinement and restrictive*** |
| 2.1 | | Determiner | $\tau$ | Articles and special pronouns or adjectives that introduce or restrict a noun |
| 2.2 | | Qualifier | $\kappa$ | Special adverbs that modify or restrict the extent or manner of a verb |
| 2.3 | | Degree | $\delta$ | Special adverbs or adjectives that modify or restrict the extent of an adjective or an adverb |
| 2.4 | | Auxiliary | $\alpha$ | Verbal prefixes that specifies the tense, voice, mood, person, and/or number of a verb |
| 2.5 | | Negative | $\neg$ | Special words that express the opposite meaning of a phrase or sentence |
| 2.6 | | Conjunction | $\gamma$ | Connections between equivalent words, phrases, and clauses such as *coordinating*, *subordinating*, and c*orrelatives* |
| 2.7 | | Interjection | $\sigma$ | A word of emotional expression |
| 3 | **Phrase** | | $\mathcal{P}$ | ***A syntactic unit with multiple words*** |
| 3.1 | | Noun phrase | NP | A composition of a noun with a determiner or adjective (phrase) |
| 3.2 | | Verb phrase | VP | A composition of a verb with a noun (phrase) or adverb (phrase) |
| 3.3 | | Adjective phrase | AP | A composition of an adjective with a degree word or noun (phrase) |
| 3.4 | | Adverb phrase | DP | A composition of an adverb with a determiner or verb (phrase) |
| 3.5 | | Prepositional phrase | PP | A composition of a preposition with a determiner or noun (phrase) |
| 3.6 | | Complement phrase | CP | A composition of an AP, NP, or DP to a subject ($\mathcal{J}$) or object ($\mathcal{O}$), and a predicate (VP). |

An important property of lexes is known as inflections, which represents the morphological flexibility of natural languages that allows the change of forms of words in accordance with grammatical roles, relations, and agreements between parts of speech in a sentence. Each part of speech including related inflection rules will be rigorously specified in the following subsections with formal models.

**Definition 9.** A *noun*, $\mathbb{N}$, in language $\mathfrak{L}$ is a part of speech that identifies and represents a real-world entity, a proper name, a meta-behavior, or an abstract concept in the following pattern:

$$\begin{array}{ll}
\mathbb{N} \triangleq N_0 \mid N_1 \mid N_n & \text{// Mass, singular, and plural nouns} \\
\quad \mid N_n^{'} & \text{// Irregular plural nouns} \\
\quad \mid N_p & \text{// Proper nouns} \\
\quad \mid \mathbb{X} & \text{// Pronouns} \\
\quad \mid V_{ing} & \text{// Gerunds} \\
\quad \mid \mathbb{N}^{+} & \text{// Complex nouns} \\
\quad \mid (\mathbb{N},)^{+} & \\
\quad \mid \mathbb{N} \ \gamma \ \mathbb{N} &
\end{array} \qquad (7)$$

where $\gamma$ represents a conjunction that will be defined in the category of modifiers, and $\mathbb{N}^{+}$ represents a compound noun that repeat $\mathbb{N}$ for one or more times such as *NN* and *N*, *N*,…, and *N*, respectively.

In addition to the lexical and syntactical patterns, a set of relational grammar rules between nouns and verbs, pronouns, and determiners can be specified known as $R_2$ (number), $R_3$ (tense), $R_4$ (mood), $R_5$ (number), $R_6$ (gender), $R_7$ (role), and $R_9$ (determiner), respectively. Because of the limit of space, the mathematical models of the relational grammar rules will be reported separately [Wang & Berwick, 2012].

**Definition 10.** A *pronoun*, $\mathbb{X}$, in language $\mathfrak{L}$ is a part of speech that is used to represent or indicate a personal, impersonal, and proper noun in nine categories in the following pattern:

$$\begin{array}{ll}
\mathbb{X} \triangleq X_{01} \mid X_{0n} & \text{// Impersonal, single/plural} \\
\quad \mid X_{11} \mid X_{1n} & \text{// First person, single/plural} \\
\quad \mid X_{21} \mid X_{2n} & \text{// Second person, single/plural} \\
\quad \mid X_{31m} \mid X_{3nm} & \text{// Third person (male), single/plural} \\
\quad \mid X_{31f} \mid X_{3nf} & \text{// Third person (female), single/plural} \\
\quad \mid X_{p} & \text{// General person} \\
\quad \mid X_{i1} \mid X_{in} & \text{// Indefinitive, single/plural} \\
\quad \mid X_{th1} \mid X_{thn} & \text{// Demonstrative, single/plural} \\
\quad \mid X_{wh} & \text{// Interrogative} \\
= & \text{// Categories: Subjective (J) | Objective (O) |Possessive (A) | Possessive-noun (N)} \\
\quad \mid (X_{01\text{-}J} \mid X_{01\text{-}O} \mid X_{01\text{-}A} \mid X_{01\text{-}N}) = (it \mid it \mid its \mid its) & \\
\quad \mid (X_{0n\text{-}J} \mid X_{0n\text{-}O} \mid X_{0n\text{-}A} \mid X_{0n\text{-}N}) = (they \mid them \mid their \mid theirs) & \\
\quad \mid (X_{11\text{-}J} \mid X_{11\text{-}O} \mid X_{11\text{-}A} \mid X_{11\text{-}N}) = (I \mid me \mid my \mid mine) & \\
\quad \mid (X_{1n\text{-}J} \mid X_{1n\text{-}O} \mid X_{1n\text{-}A} \mid X_{1n\text{-}N}) = (we \mid us \mid our \mid ours) & \\
\quad \mid (X_{21\text{-}J} \mid X_{21\text{-}O} \mid X_{21\text{-}A} \mid X_{21\text{-}N}) = (you \mid you \mid your \mid yours) & \\
\quad \mid (X_{2n\text{-}J} \mid X_{2n\text{-}O} \mid X_{2n\text{-}A} \mid X_{2n\text{-}N}) = (you \mid you \mid your \mid yours) & \\
\quad \mid (X_{31m\text{-}J} \mid X_{31m\text{-}O} \mid X_{31m\text{-}A} \mid X_{31m\text{-}N}) = (he \mid him \mid his \mid his) & \\
\quad \mid (X_{31f\text{-}J} \mid X_{31f\text{-}O} \mid X_{31f\text{-}A} \mid X_{31f\text{-}N}) = (she \mid her \mid her \mid hers) & \\
\quad \mid (X_{3n\text{-}J} \mid X_{3n\text{-}O} \mid X_{3n\text{-}A} \mid X_{3n\text{-}N}) = (they \mid them \mid their \mid theirs) & \\
\quad \mid (X_{p\text{-}J} \mid X_{p\text{-}O} \mid X_{p\text{-}A} \mid X_{p\text{-}N}) = (who \mid whom \mid whose \mid whose) & \\
\quad \mid (X_{i1} \mid X_{in}) = ((anything,\ anyone,\ anybody) \mid & \\
\qquad\qquad (something,\ someone,\ somebody)) & \\
\quad \mid (X_{th1} \mid X_{thn}) = ((this,\ that) \mid (these,\ those)) & \\
\quad \mid X_{wh} = (which,\ what,\ ...) &
\end{array} \qquad (8)$$

where the first six categories are personal or impersonal pronouns; while the last three categories are relative pronouns for introducing indefinite, demonstrative, and interrogative noun phrase (*NP*) and/or subordinate clauses.

It is noteworthy that most pronouns except the relative pronouns have inflections in the forms of *subjective* ($X_{\text{-}J}$), *objective* ($X_{\text{-}J}$), *possessive* ($X_{\text{-}J}$), and *possessive-noun* ($X_{\text{-}J}$) pronouns as shown in the second part of Eq. 8. A set of relational grammar rules between pronouns and nouns or verbs can be described by $R_1$ (person), $R_3$ (tense), $R_4$ (mood), $R_6$ (gender), $R_7$ (role), and $R_8$ (antecedent), respectively [Wang & Berwick, 2012].

**Definition 11.** A *verb*, $\mathbb{V}$, in language $\mathfrak{L}$ is a part of speech that denotes a state, an existence, a behavior, an action, and an occurrence in the following pattern:

$$\begin{array}{ll}
\mathbb{V} \triangleq V & \text{// Stems} \\
\quad \mid V_s & \text{// 3rd-person-singular} \\
\quad \mid V_s^{'} & \text{// 3rd-person-singular (irregular)} \\
\quad \text{// } \textbf{\textit{Relation to objects}} & \\
\quad \mid V_t & \text{// Transitive} \\
\quad \mid V_i & \text{// Intransitive} \\
\quad \mid V_{tt} & \text{// Ditransitive} \\
\quad \text{// } \textbf{\textit{Tenses}} & \\
\quad \mid V_p = (V \mid V_s \mid V_s^{'}) & \text{// Present} \\
\quad \mid V_f = \alpha \ V & \text{// Future} \\
\quad \mid V_{pt} & \text{// Past} \\
\quad \mid V_{pt}^{'} & \text{// Past (irregular)} \\
\quad \mid V_{pp} & \text{// Past participle} \\
\quad \mid V_{pp}^{'} & \text{// Past participle (irregular)} \\
\quad \text{// } \textbf{\textit{Nonfinite}}\ (verbals) & \\
\quad \mid V_{to} & \text{// Infinitives} \\
\quad \mid V_{ing} & \text{// Gerund} \\
\quad \mid V_{ing} & \text{// Present participial} \\
\quad \mid V_d & \text{// Past participlial} \\
\quad \text{// } \textbf{\textit{Special verbs}} & \\
\quad \mid \text{BE} = be \mid am \mid is \mid are \mid was \mid were \mid to\ \ be \mid being \mid been & \\
\quad \mid \text{HAVE} = have \mid has \mid had & \\
\quad \mid \text{DO} = do \mid does \mid did \mid done & \\
\quad \text{// } Complex\ \ verbs & \\
\quad \mid (\mathbb{V},)^{+} & \\
\quad \mid \mathbb{V} \ \gamma \ \mathbb{V} &
\end{array} \qquad (9)$$

The inflections of verbs are specified in Eq. 9 according to person, object, tense, and their nonfinite forms. The irregular forms of verbs, particular those of frequently used ones such as *BE*, *HAVE*, and *DO*, are highlighted in Eq. 9.

The grammar rules of verbs encompass those of person, number, tense, mood, and voice, as well as modifiers and special structures. A set of relational grammar rules between verbs and nouns, pronouns, modifiers, and special sentence structures such as $R_1$ (person), $R_2$ (number), $R_3$ (tense), $R_4$ (mood), $R_5$ (voice), $R_{10}$ (qualifier), $R_{12}$ (negation), $R_{13}$ (inverted order), and $R_{14}$ (There BE structure) is specified, respectively, [Wang & Berwick, 2012].

**Definition 12.** An *adjective*, $\mathbb{A}$, in language $\mathfrak{L}$ is a part of speech that is used to modify or restrict a noun or pronoun in the following pattern:

$$\begin{aligned}
\mathbb{A} \triangleq\ & A && \text{// Positive} \\
& \mid A_{er} && \text{// Comparative} \\
& \mid A_{er}^{'} && \text{// Comparative (irregular)} \\
& \mid A_{st} && \text{// Superlative} \\
& \mid A_{st}^{'} && \text{// Superlative (irregular)} \\
& \mid (N_0\ 's \mid N_1\ 's) && \text{// Possessive nouns (singlular)} \\
& \mid N_n\ ' && \text{// Possessive nouns (plural)} \\
& \mid \mathbb{X}_P && \text{// Possessive pronouns} \\
& \mid V_{to}P && \text{// Infinitive phrases} \\
& \mid V_{ed}P && \text{// Participial phrases} \\
& \mid PP && \text{// Propositional phrase} \\
& \mid (\mathbb{A},)^{+} && \text{// Complex adjectives} \\
& \mid \mathbb{A}\ \gamma\ \mathbb{A}
\end{aligned} \tag{10}$$

where $A, A_{er}$, and $A_{st}$ represent the positive, comparative, and superlative levels of adjectives, respectively.

**Definition 13.** An *adverb*, $\mathbb{D}$, in language $\mathfrak{L}$ is a part of speech that is used to modify or restrict a verb, an adjective, or another adverb in the following pattern:

$$\begin{aligned}
\mathbb{D} \triangleq\ & D && \text{// Positive} \\
& \mid (more \mid much \mid less)\ D && \text{// Comparative} \\
& \mid D_c^{'} && \text{// Comparative (irregular)} \\
& \mid (most \mid least)\ D && \text{// Superlative} \\
& \mid D_s^{'} && \text{// Superlative (irregular)} \\
& \mid V_{to}P && \text{// Infinitive phrases} \\
& \mid PP && \text{// Propositional phrase} \\
& \mid \mathbb{D}\ \mathbb{D} && \text{// Complex adjectives} \\
& \mid (\mathbb{D},)^{+} \\
& \mid \mathbb{D}\ \gamma\ \mathbb{D}
\end{aligned} \tag{11}$$

where $D, D_{er}$, and $D_{st}$ represent the positive, comparative, and superlative levels of adverbs, respectively.

**Definition 14.** A *preposition*, $\mathbb{P}$, in language $\mathfrak{L}$ is a part of speech that is used to link and relate a word phrase, or clause to anther counterpart in the following pattern:

$$\begin{aligned}
\mathbb{P} \triangleq\ & P && \text{// Primitive} \\
& \mid \mathbb{P}\ \gamma\ \mathbb{P} && \text{// Complex adjectives}
\end{aligned} \tag{12}$$

**3.2 Modifiers**

The parts of speech described in the preceding subsection describe the roles of words in a language. When analyzing relations of words in a sentence, some words behave as modifiers that affect the syntax or semantics of the target words. The modifiers are a subset of adjectives, adverbs, pronouns, nouns, and infinite verbs.

**Definition 15.** The *modifiers* in $\mathfrak{L}$ are words or phrases that describe, limit, qualify or connect another word or phrase in the categories of *determiners* (τ), *qualifiers* (κ), *degrees* (δ), *auxiliaries* (α), *negatives* (¬), *conjunctions* (γ), and *interjections* (σ) where each category can be formally modeled as a specific set in $\mathfrak{L}$, i.e.:

$$\begin{aligned}
&\tau \triangleq \{\tau_i \in \tau \mid \tau \sim \mathbb{N} \wedge \tau \subseteq (\mathcal{T} \cup \mathcal{X} \cup \mathcal{A}) \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\
&\kappa \triangleq \{\kappa_i \in \kappa \mid \kappa \sim \mathbb{V} \wedge \kappa \subseteq \mathcal{D} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\
&\delta \triangleq \{\delta_i \in \delta \mid \delta \sim (\mathbb{A} \cup \mathbb{D}) \wedge \delta \subseteq \mathcal{D} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\
&\alpha \triangleq \{\alpha_i \in \alpha \mid \alpha \sim \mathbb{V} \wedge \alpha \subseteq \mathcal{U} \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\
&\neg \triangleq \{\neg_i \in \neg \mid \neg \sim (\mathcal{P} \cup \mathcal{S}) \wedge \neg \subseteq (\mathcal{D} \cup \mathcal{A} \cup \mathcal{X}) \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\
&\gamma \triangleq \{\gamma_i \in \gamma \mid \gamma \subset \mathfrak{W} \sqsubset \mathfrak{L}\} \\
&\sigma \triangleq \{\sigma_i \in \sigma \mid \sigma \subset \mathfrak{W} \sqsubset \mathfrak{L}\}
\end{aligned} \tag{13}$$

where $\sim$ represents a relation between a certain type of modifies and the suitable type(s) of elements that it may operate.

Each category of modifiers will be rigorously specified in the following subsections with formal models.

**Definition 16.** The *determiners*, $\tau$, in language $\mathfrak{L}$ are a set of *articles*, *demonstrative adjectives*, *restrictive adjectives*, and *possessive pronouns* that introduces or restricts a noun in the following pattern:

$$\begin{aligned}
\tau \triangleq\ & ARTICLE \mid Demonstrative_A \mid Restrictive_A \mid Possessive_X \\
= (\ & (ARTICLE = \{a, an, the\}) \\
& \mid (Demonstrative_A = \{\text{this, that, these, those}\}) \\
& \mid (Restrictive_A = \{\text{every, each, any, all, some, ...}\}) \\
& \mid (Possessive_X = \{my, his, her, our, your, their, its\}) \\
)\ &
\end{aligned} \tag{14}$$

where $\tau \subseteq (\mathcal{T} \cup \mathcal{X} \cup \mathcal{A})$.

The number rules of determiners are specified by $R_9$ in [Wang & Berwick, 2012].

**Definition 17.** A *qualifiers*, κ, in language $\mathfrak{L}$ are a set of special adverbs that modifies the extent or manner of a verb as follows:

$$\kappa \triangleq \{almost,\ always,\ often,\ perhaps,\ never,\ ...\},\ \kappa \subset \mathbb{D} \tag{15}$$

The modification rules of qualifiers are specified by $R_{10}$ in [Wang & Berwick, 2012].

**Definition 18.** The *degrees*, *δ*, in language $\mathfrak{L}$ are a set of special adverbs or adjectives that modifies or restricts the extent of an adjective or an adverb as follows:

$$\delta \triangleq \{too,\ so,\ very,\ more,\ quite,\ ...\},\ \delta \subset (\mathbb{A} \cup \mathbb{D}) \tag{16}$$

The modification rules of degrees are specified by $R_{11}$ in [Wang & Berwick, 2012].

**Definition 19.** The *auxiliaries*, α, in language $\mathfrak{L}$ is a set of verbal prefixes that indicates the tense, voice, mood, person, and/or number of a verb in the following pattern:

$$\begin{aligned}
\alpha &\triangleq ASPECT \mid Positive_MODAL \mid Passive_MODAL \mid Dummy_MODAL \\
&= (\ ASPECT = ( \\
&\qquad\qquad \mid \text{Habitual:} \quad V_p = (V \mid V_s \mid V_s') \\
&\qquad\qquad \mid \text{Progressive:} \ BE\ V_{ing} \\
&\qquad\qquad \mid \text{Complete:} \quad HAVE\,(V_{pp} \mid V_{pp}') \\
&\qquad\qquad \mid \text{Future:} \quad will\ V \\
&\qquad\qquad \mid \text{Past:} \quad (V_d \mid V_d') \\
&\qquad\quad ) \\
&\quad \mid (Positive_MODAL = \{will,\ would,\ shall,\ should,\ may,\ might,\ can, \\
&\qquad\qquad could,\ must,\ \ldots\}) \\
&\quad \mid (Passive_MODAL = \{BE\,(V_{pp} \mid V_{pp}'),\ HAVE\ been\,(V_{pp} \mid V_{pp}'),\ ...\}) \\
&\quad \mid (Dummy_MODAL = DO\ (V \mid VP)) \\
&\quad )
\end{aligned} \tag{17}$$

The tense rules of auxiliaries are specified in $R_3$ in [Wang & Berwick, 2012].

**Definition 20.** The *negations*, ¬, in language $\mathfrak{L}$ is a category of words that expresses the opposite meaning of a phrase or sentence, i.e.:

$$\begin{aligned}\neg \triangleq \{&not,\ no,\ never,\ nothing,\ nobody,\ none, \\ &neither...nor,\ ...\},\ \neg \subset (\mathbb{D} \cup \mathbb{A} \cup \mathbb{X})\end{aligned} \tag{18}$$

The negation rules are specified by $R_{12}$ in [Wang & Berwick, 2012].

**Definition 21.** The *conjunctions*, $\gamma$, in language $\mathfrak{L}$ is a special part of speech that connects and relates words, phrases, clauses by using the *coordinatives, subordinates,* and *correlatives* in the following pattern:

$$\begin{aligned}
\gamma &\triangleq (\gamma_{\text{Coordinatives}} \mid \gamma_{\text{Subordinates}} \mid \gamma_{\text{Correlatives}}) \\
&= (\ (\gamma_{\text{Coordinatives}} = \{and,\ or,\ nor,\ but,\ for,\ so,\ \ldots\}) \\
&\quad \mid (\gamma_{\text{Subordinates}} = \{when,\ while,\ after,\ before,\ as, \\
&\qquad because,\ who,\ if,\ since,\ unless,\ until,\ ...\}) \\
&\quad \mid (\gamma_{\text{Correlatives}} = \{both\ ...\ and,\ either\ ...\ or, \\
&\qquad not\ only\ ...\ but\ also,\ whether\ ...\ or,\ ...\}) \\
&\quad )
\end{aligned} \tag{19}$$

It is noteworthy in Eq. 19 that the coordinative and correlative conjunctions connect two or more equivalent elements in parallel where the elements can be at the levels of word, phrase, clause, and sentence. However, the subordinate conjunctions relate a subsentence to the main sentence.

**Definition 22.** The *interjections,* $\sigma$, in language $\mathfrak{L}$ is a special part of speech that expresses an exclamation, qualification, or degree, which extends the context of a sentence in the following pattern:

$$\begin{aligned}
\sigma &\triangleq (\text{Exclamation} \mid \text{qualifiers } (\kappa) \mid \text{degrees } (\delta)) \\
&= \ (Oh \mid Wow \mid Whew,\ ...)!\ \mathcal{S} \\
&\mid \ \kappa,\ \mathcal{S} \\
&\mid \ \delta,\ \mathcal{S}
\end{aligned} \tag{20}$$

The interjections usually appear in the front of sentence separated by the exclamation or comma to denote a strong or weak emotional expression, respectively.

### 3.3 Phrases

Phrases are the second-level structures beyond lexes in a sentence. A phrase is not a complete sentence where its principal verb is not in the finite form. Otherwise, it is a clause. Phrases can be classified into six categories as follows.

**Definition 23.** A *phrase* in $\mathfrak{L}$ is a combination of multiple words that forms an element of sentence in the categories of *noun phrase* (*NP*), *verb phrase* (*VP*), *adjective phrase* (*AP*), *adverb phrase* (*DP*), *proposition phrase* (*PP*), and *complement phrase* (*PP*) where each category of the phrases can be formally modeled as a specific set in $\mathfrak{L}$, i.e.:

$$\begin{aligned}
NP &\triangleq \{p_i \in NP \mid NP \subset \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}\} \\
VP &\triangleq \{p_i \in VP \mid VP \subset \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}\} \\
AP &\triangleq \{p_i \in AP \mid AP \subset \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}\} \\
DP &\triangleq \{p_i \in DP \mid DP \subset \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}\} \\
PP &\triangleq \{p_i \in PP \mid PP \subset \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}\} \\
CP &\triangleq \{p_i \in CP \mid CP \subset \text{Þ}\mathfrak{W} \sqsubset \mathfrak{L}\}
\end{aligned} \tag{21}$$

where $p_i$ represents a phrase, and Þ$\mathfrak{W}$ denotes a power set or combination of words in language $\mathfrak{L}$.

Each category of phrases will be rigorously specified in the following subsections with formal models.

**Definition 24.** A *noun phrase*, *NP*, in language $\mathfrak{L}$ is a noun composed with a determiner or adjective in front or following by alternative prepositional phrase and/or compliment phrase:

$$\begin{aligned} NP \triangleq & [\tau \mid \tau\ \mathbb{A} \mid \tau\ AP]\ \mathbb{N}\ [PP \mid CP \mid PP\ CP] \\ & \mid NP\ \gamma\ NP \\ & \mid (NP,)^{+} \\ & \mid (NP)^{+} \end{aligned} \tag{22}$$

where $\mathbb{N}$ is the principal noun including the subjective pronouns $X_{-J}$, objective pronouns $X_{-O}$, and gerunds $V_{ing}$ as defined in Definition 9.

**Definition 25.** A *verb phrase*, *VP*, in language $\mathfrak{L}$ is a verb modified by an auxiliary, qualifier, adverb, adverb phrase, and/or negation in the following pattern:

$$\begin{aligned} VP \triangleq & [\alpha \mid \neg \mid \alpha\,\neg \mid \kappa \mid DP \mid \alpha\ DP \mid \alpha\,\neg\ DP]\ \mathbb{V}\ [DP] \\ & \mid\ V_{to}P = to\ V && \text{// Infinitive phrase} \\ & \mid\ V_{ing}P = V_{ing} && \text{// Gerund phrase} \\ & \mid\ V_{ing}P = BE\ \ V_{ing} && \text{// Present participial phrase} \\ & \mid\ V_{d}P = HAVE\ \ V_{d} && \text{// Past participial phrase} \\ & \mid\ VP\ \gamma\ VP && \text{// Complex VP} \\ & \mid\ (VP,)^{+} \end{aligned} \tag{23}$$

where $\mathbb{V}$ is the principal verb as defined in Definition 11.

**Definition 26.** An *adjective phrase*, *AP*, in language $\mathfrak{L}$ is an adjective modified by an adverb, an adverb phrase, or a degree word in the following pattern:

$$\begin{aligned} AP \triangleq & [\delta \mid D \mid DP]\ \mathbb{A} \\ & \mid (NP \mid XP)\ AP \\ & \mid AP\ \gamma\ AP \\ & \mid (AP,)^{+} \end{aligned} \tag{24}$$

where $\mathbb{A}$ is the principal adjective as defined in Definition 12.

**Definition 27.** An *adverb phrase*, *DP*, in language $\mathfrak{L}$ is an adverb modified by a degree word δ in the following pattern:

$$\begin{aligned} DP \triangleq & [\delta]\ \mathbb{D}\ [CP\ ] \\ & \mid\ DP\ (VP \mid DP \mid AP) \\ & \mid\ VP\ DP \\ & \mid\ (\mathbb{D} \mid DP)\ ,\ \mathcal{S} \\ & \mid\ (\mathbb{D} \mid DP)\ !\ \mathcal{S} \\ & \mid\ DP\ \gamma\ DP \end{aligned} \tag{25}$$

where $\mathbb{D}$ is the principal adverb as defined in Definition 13.

**Definition 28.** A *prepositional phrase*, *PP*, in language $\mathfrak{L}$ is a phrase of preposition with object that modifies a noun or noun phrase in the following pattern:

$$PP \triangleq [DP\ ]\ \mathbb{P}\ (\mathbb{N} \mid NP) \tag{26}$$

where $\mathbb{P}$ is the principal preposition, and NP including the subjective pronouns $X_{-O}$, as defined in Definition 14.

**Definition 29.** A *complement phrase*, *CP*, in language $\mathfrak{L}$ is an adjective or noun phrase that completes the sense of a verb, a subject, and an object in the following pattern:

$$CP \triangleq \mathbb{A}\,[\mathbb{N}] \mid AP\,[NP] \tag{27}$$

CPs are very widely used phrases in sentences, which can be a supplemental part of N/NP, V/VP, A/AP, or P/PP [Hodges and Whitten, 1982; O'Grady and Archibald, 2000; Taylor, 2002].

## 4. The Formal Sentence Structures of English

On the basis of the syntactic structures, mathematical models, and grammar rules described in Section 3, a formal model of sentence in English can be rigorously developed in this section.

### 4.1 The Formal Model of Sentences

A *sentence* is a basic unit of a complete expression for a thread of thought or a behavior. The general structure of a sentence in a natural language encompasses a subject and a predicate where the latter can be further divided as a verb phrase (behavior) and its object(s).

**Definition 30.** An *abstract sentence*, $\mathcal{S}$, in language $\mathfrak{L}$ can be formally described as a triple:

$$\mathcal{S} \triangleq [\mathcal{J}]\ \mathbb{V}\ [\mathcal{O}] \tag{28}$$

where

- $\mathcal{J}$ is the *subject* of the sentence, $\mathcal{J} \in (\mathbb{N} \cup \mathbb{X} \cup NP \cup XP\ ) \subset \mathcal{N} \subset \text{þ}\mathfrak{W} \sqsubset \mathfrak{L}$;
- $\mathbb{V}$ is the *predicate* of the sentence with an inflective form of *verb* or *VP*, $\mathbb{V} \in \mathcal{V} \subset \text{þ}\mathfrak{W} \sqsubset \mathfrak{L}$;
- $\mathcal{O}$ is the object of the predicate and the sentence, $\mathcal{O} \in (\mathbb{N} \cup \mathbb{X} \cup NP \cup XP) \subset \mathcal{N} \subset \text{þ}\mathfrak{W} \sqsubset \mathfrak{L}$, where the brackets [ ] represent an optional term.

Sentences may be classified according to their purposes known as *statements, commands, questions*, and *exclamations*. Sentences may also be classified according to their syntactic structures known as *simple* and *complex* ones. A complex sentence is usually constructed by conjunctions of subsentences known as clauses.

**Definition 31.** A *clause* is the third-level structures beyond a phrase that is a subsentence related to another one in a complex sentence usually connected by a coordinative or correlative conjunctions as in *independent* clauses and introduced by subordinate or relative pronouns as in *dependent* clauses in the following pattern:

$$
\begin{aligned}
\mathcal{S} \triangleq\ & (\text{Independent\_clause } (IC) \,|\, \text{Dependent\_clause } (DC)): \\
& = \text{IC} = (\ \ (C, C) \\
& \qquad\qquad |\ (C; C) \\
& \qquad\qquad |\ (C\ \gamma_{Coordinative}\ C) \\
& \qquad\qquad |\ (C\ \gamma_{Correlative}\ C) \\
& \qquad\qquad ) \\
& \ \ |\ \text{DC} = (\ \ (C_m\ \gamma_{Subordinate}\ C_s) \\
& \qquad\qquad\quad |\ (C_m\ X_r\ C_s) \\
& \qquad\qquad\quad )
\end{aligned}
\tag{29}
$$

where $X_r$ represents relative pronouns such as the general personal ($X_p$), *indefinite* ($X_i$), *demonstrative* ($X_{th}$), and *interrogative* ($X_{wh}$) pronouns, respectively, as defined in Eq. 8.

The connections between two clauses in a complex one can be established by the *conjunctions* (γ) as described in Definition 21. The relations of clauses in a complex sentence can be *equivalent* and *subordinate*. In the former case, the clauses are known as *independent clauses* (IC); while in the latter case, the clauses are known as the *dependent clauses* (DCs) where the *main* clause is in the form of a complete sentence, and the *subordinate* clause is a sentence introduced by a conjunction or relative pronoun.

There are noun, adjective, and adverb clauses from the syntactical point view. In some cases the introductory conjunction for a subordinate clause may be omitted when the syntax and semantics are reserved and be clearly understood. This form of clause is called *elliptical* clauses.

## 4.2 The Deductive Grammar of English

On the basis of the mathematical models of both the individual syntactic structures as developed in Section 3, a formal model of the English grammar known as the deductive grammar can be rigorously established as shown in Fig. 1.

**Definition 32.** The *deductive grammar* is an abstract grammar that formally denotes the syntactic rules of a language based on which as a generic formula, valid language sentences can be deductively derived.

The Deductive Grammar of English (DGE) can be formally described in EBNF as shown in Fig. 1. DEG specifies the formal structures of sentence, clauses, phrases, modifiers, and terminals (lexes), as well as their syntactical rules, in a top-down hierarchy. Based on the formal DGE and supported by a lexical database for all terminals words, a parser can be derived to autonomously process texts expressed in English according to the formal grammar.

The syntactic rules of individual elements such as those of lexes, modifiers, and phrases as well as their inflections have been formally defined in Section 3. In addition to the syntactic rules, there are 15 relational semantic rules between different elements of sentences in English such as those of $\mathcal{J}$-$\mathbb{V}$ agreement, $\mathbb{X}$-$\mathcal{J}$ agreement, negation, and inverted structures [Wang & Berwick, 2012].

**The Deductive Grammar of English**

```
// Sentence structure
𝒮 ::= [𝒥] 𝒫                                   // Sentence
    | 𝒮 γ 𝒮                                   // Complex sentence with conjunctions
    | σ 𝒮                                     // Complex sentence with interjections
J ::= NP                                      // Subject
𝒫 ::= VP [𝒪⁺]                                 // Predicate
𝒪 ::= NP                                      // Object

// Phrase structures
NP ::= [τ | τ AP] ℕ [PP | CP | PP CP]                          // Noun phrase
     | NP γ NP | (NP,)⁺ | (NP)⁺
VP ::= [α | ¬ | α ¬ | κ | DP | α DP | α ¬ DP] 𝕍 [DP]          // Verb phrase
     | VP γ VP | (VP,)⁺
AP ::= [δ | DP] 𝔸                                              // Adjective phrase
     | AP γ AP | (AP,)⁺
DP ::= [δ] 𝔻 [CP]                                              // Adverb phrase
     | DP γ DP
PP ::= [DP] ℙ NP                                               // Prepositional phrase
CP ::= AP [NP]                                                 // Complement phrase
// Relational syntactic rules
ℛ ::= (R_1 | R_2 | … | R_15)                                   // See [40]
// Lexes
ℕ ::= 𝔗_N | ℕ                                                 // Nouns
𝕏 ::= 𝔗_X | 𝕏                                                 // Pronouns
𝕍 ::= 𝔗_V | 𝕍                                                 // Verbs
𝔸 ::= 𝔗_A | 𝔸                                                 // Adjectives
𝔻 ::= 𝔗_D | 𝔻                                                 // Adverbs
ℙ ::= 𝔗_P | ℙ                                                 // Propositions

// Modifiers
τ ::= <determiners>
κ ::= <qualifiers>
δ ::= <degrees>
α ::= <auxiliaries>
¬ ::= <negative>
γ  ::= <conjunctions>
σ ::= <interjections>
// Terminals (𝔗)
𝔗_N ::= <terminal_nouns>                                       // 𝔗_N ⊆ 𝔚
𝔗_X ::= <pronouns>                                             // Definition 11
𝔗_V ::= BE | HAVE | DO | <terminal_verbs>                      // 𝔗_V ⊆ 𝔚
𝔗_A ::= <adjectives>                                           // Definition 12
𝔗_D ::= <adverbs>                                              // Definition 13
𝔗_P ::= <propositions>                                         // Definition 14
𝔗_γ ::= <conjunctions>                                         // Definition 21
𝔗_σ ::= <interjections>                                        // Definition 22
BE ::= <be> |<am>| <is> | <are> | <was> | <were> | <to be> | <being> | <been>
HAVE ::= <have> | <has> | <had>
DO ::= <do> |<does>| <did> | <done>
```

Fig.1 The deductive grammar of a generic sentence in English

The DGE provides not only an essential set of syntactic rules for implementing an English parser for natural language processing, but also the establishment of the general pattern of English sentences based on it any grammatically corrected sentence is an instance of the general pattern. The DGE theory indicates that, instead of analyzing variously infinite instances of sentences in linguistics, a more efficient way is to rigorously specify the general pattern of the language where all sentences fit. This approach can be demonstrated by Table 2 where the higher level structures describe the syntactical models of a general sentence in DGE; while the bottom level structure is the general pattern of sentence, which function as a template for deductively deriving any concrete instance according to DGE. The DGE can also be illustrated by a syntactic diagram for the general sentence pattern as the most general and most complicated structure of English syntax.

A number of examples are demonstrated in Table 2 against the general pattern of sentences in English. Observing Fig. 1 and Table 2, it is noteworthy that the syntactic structure of the DGE pattern is highly recursive, particularly all forms of phrases in the subject, object, and predicate of the sentence. The recursive property of elements in sentences at the levels of parts of speeches, phrases, and clauses is interesting demonstrated in the general pattern, particularly the recursive structure of clauses as shown in Example 3 by clause $\mathcal{C}_{32}$. In order to save space, a few of simple phrases in Table 2 are not completely extended to its terminal lexical elements. The case studies indicate that no matter how simple or complex an instance sentences would be, they can be efficiently and effectively recognized, analyzed, and processed by the general formal model of DGE by human or machines.

Table 2. Application Examples of the Deductive Grammar of English

| Syntactical Structure | | | | | | | Example 1 | Example 2 | Example 3 | | Example 4 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Sentence ($\mathcal{S}$) | Subject ($\mathcal{J}$) | NP+ | NP | | | τ | | The | The | | The |
| | | | | AP | DP | δ | | | | | newly |
| | | | | | | $\mathbb{D}$ | | | strongest | | registered |
| | | | | | | CP | | | | | |
| | | | | | | $\mathbb{A}$ | | dark | | | |
| | | | | | | $\mathbb{N}$ | | clouds | flood | officials | students |
| | | | | PP | DP | δ | | | ever | | |
| | | | | | | $\mathbb{D}$ | | | recorded | | |
| | | | | | | CP | | | | | |
| | | | | | | $\mathbb{P}$ | | on | In | of | of |
| | | | | | | NP | | the horizon | south | UN | Math-30 |
| | | | | CP | | AP | | | | | |
| | | | | | | NP | | | | | |
| | | | | | | γ | | | | | |
| | | | NP | | | … | | | | | |
| | Predicate ($\mathcal{P}$) | VP+ | VP | | | α | | had | | | will |
| | | | | | | ¬ | | | | | not |
| | | | | | | κ | | | | | |
| | | | | DP | | δ | | | | | |
| | | | | | | $\mathbb{D}$ | | suddenly | | | immediately |
| | | | | | | CP | | | | | |
| | | | | | | $\mathbb{V}$ | Look | appeared | caused | said | get |
| | | | | DP | | δ | | | | | |
| | | | | | | $\mathbb{D}$ | | to approach to | | today | |
| | | | | | | CP | | the town | | | |
| | | | | | | γ | | | | | |
| | | | VP | | | … | | | | | |
| | Object ($\mathcal{O}$) | NP+ | NP | | | τ | | | | | the |
| | | | | AP | DP | δ | | | | | previously |
| | | | | | | $\mathbb{D}$ | | | | | expected |
| | | | | | | CP | | | | | |
| | | | | | | $\mathbb{A}$ | | | | | comprehensive |
| | | | | | | $\mathbb{N}$ | | | thousands | | handbook |
| | | | | PP | | $\mathbb{P}$ | | | of | | on |
| | | | | | | NP | | | People | | Monday |
| | | | | CP | | AP | | | | | |
| | | | | | | NP | | | | | , the first day of the class |
| | | | | DP | | δ | | | | | |
| | | | | | | $\mathbb{D}$ | | | homeless | | |
| | | | | | | CP | | | | | |
| | | | | | | γ | | | | | |
| | | | NP | | | … | | | | | |
| γ | | | | | | | | | , | | . |
| $\mathcal{S}$ | … | | | | | | | | $\mathcal{C}_{31}$ | $\mathcal{C}_{32}$ | |
| End | | | | | | | ! | . | | . | . |

## 5. Conclusions

This paper has represented a formal, general, and precise methodology for natural language processing in cognitive and computational linguistics. Denotational mathematics has been adopted to model and specify the Deductive Grammar of English (DGE). The work has demonstrated that the grammar of a natural language can be formally specified by a set of rigorous syntactical structures and rules. Based on them, cognitive systems for machine-enabled language processing and learning may be implemented. A wide range of applications of the formal models of DGE have been identified in natural language processing, computational linguistics, and cognitive informatics in general, and in word processing, web search engines, machine language comprehension, autonomous machine learning, smart cell phones, semantic computing, cognitive computing, and computing with words in particular.


### Acknowledgement

The authors would like to acknowledge the Natural Science and Engineering Council of Canada (NSERC) for its partial support to this work. We would like to thank the anonymous reviewers for their valuable comments and suggestions.


### References


[1] Chomsky, N. (1956), Three Models for the Description of Languages, *I.R.E. Transactions on Information Theory*, IT-2(3), 113-124.

[2] Chomsky, N. (1957), *Syntactic Structures*, Mouton, the Hague.

[3] Chomsky, N. (1959), On Certain Formal Properties of Grammars, *Information and Control*, 2, 137-167.

[4] Chomsky, N. (1962), Context-Free Grammar and Pushdown Storage, Quarterly Progress Report, *MIT Research Laboratory*, 65, 187-194.

[5] Chomsky, N. (1965), *Aspects of the Theory of Syntax*, MIT Press, Cambridge, MA.

[6] Chomsky, N. (1982), *Some Concepts and Consequences of the Theory of Government and Binding*, MIT Press, Cambridge, MA.

[7] Chomsky, N. (2007), Approaching UG from Below. Interfaces + Recursion = Language? *Chomsky's Minimalism and the View from Syntax-Semantics*, Mouton, Berlin.

[8] Crystal, D. (1987), *The Cambridge Encyclopedia of Language*, Cambridge University Press, NY.

[9] Evans, V. & M. Green (2006). *Cognitive Linguistics: An Introduction*. Edinburgh: Edinburgh University Press.

[10] Gibbs (1996), The Expansion of a New Paradigm in Linguistics: Cognitive Linguistic Research, in Casad ed. *Cognitive Linguistics in the Redwoods*, Mouton De Gruyter, June.

[11] Hodges, J.C. and M.E. Whitten (1982), *Harbrace College Handbook*, 9th ed., HBJ Inc., NY.

[12] Langlotz, Andreas (2006), Idiomatic Creativity: A Cognitive-Linguistic Model of Idiom-Representation and Idiom Variation in English. Amsterdam: John Benjamins.

[13] O'Grady, W. and J. Archibald (2000), *Contemporary Linguistic Analysis: An Introduction*, 4th ed., Pearson Education Canada Inc., Toronto.

[14] Pullman, S. (1997), *Computational Linguistics*, Cambridge University Press, Cambridge, UK.

[15] Huddleston, R. and G.K. Pullum (2002), *The Cambridge Grammar of the English Language*, Cambridge University Press, UK.

[16] Tarski, A. (1944), The Semantic Conception of Truth, *Philosophic Phenomenological Research*, 4, 13-47.

[17] Taylor, J.R. (2002), *Cognitive Grammar*, Oxford University Press.

[18] Wang, Y. (2002), *Keynote: On Cognitive Informatics*, Proceedings of the First IEEE International Conference on Cognitive Informatics (ICCI'02), Calgary, AB., Canada, IEEE CS Press, August, pp.34-42.

[19] Wang, Y. (2003), On Cognitive Informatics, *Brain and Mind: A Transdisciplinary Journal of Neuroscience and Neurophilosophy*, 4(2), 151-167.

[20] Wang, Y. (2007a), *Software Engineering Foundations: A Software Science Perspective*, CRC Series in Software Engineering, Vol. II, Auerbach Publications, USA, July.

[21] Wang, Y. (2007b), The Theoretical Framework of Cognitive Informatics, *International Journal of Cognitive Informatics and Natural Intelligence*, 1(1), 1-27.

[22] Wang, Y. (2007c), The OAR Model of Neural Informatics for Internal Knowledge Representation in the Brain, *International Journal of Cognitive Informatics and Natural Intelligence*, USA, 1(3), 64-75.

[23] Wang, Y. (2008a), On Contemporary Denotational Mathematics for Computational Intelligence, *Transactions of Computational Science*, Springer, August, 2, 6-29.

[24] Wang, Y. (2008b), On Concept Algebra: A Denotational Mathematical Structure for Knowledge and Software Modeling, *International Journal of Cognitive Informatics and Natural Intelligence*, USA, 2(2), 1-19.

[25] Wang, Y. (2009a), A Formal Syntax of Natural Languages and the Deductive Grammar, *Fundamenta Informaticae*, 90(4), 353-368.

[26] Wang, Y. (2009b), On Cognitive Computing, *International Journal of Software Science and Computational Intelligence*, 1(3), 1-15.

[27] Wang, Y. (2009c), On Abstract Intelligence: Toward a Unified Theory of Natural, Artificial, Machinable, and Computational Intelligence, *International Journal of Software Science and Computational Intelligence*, 1(1), 1-17.

[28] Wang, Y. (2010a), Cognitive Robots: A Reference Model towards Intelligent Authentication, *IEEE Robotics and Automation*, 17(4), 54-62.

[29] Wang, Y. (2010b), On Formal and Cognitive Semantics for Semantic Computing, *International Journal of Semantic Computing*, 4(2), 203–237.

[30] Wang, Y. (2010c), On Concept Algebra for Computing with Words (CWW), *International Journal of Semantic Computing*, 4(3), 331-356.

[31] Wang, Y. (2011), Inference Algebra (IA): A Denotational Mathematics for Cognitive Computing and Machine Reasoning (I), *International Journal of Cognitive Informatics and Natural Intelligence*, 5(4), 62-83.

[32] Wang, Y. (2012a), Cognitive Linguistic Perspectives on the Chinese Language, *International Journal of New Mathematics and Applications*, in press.

[33] Wang, Y. (2012b), Formal Ontology and Concept Algebra for Machine Learning and Cognitive Computing, *Journal of Advanced Mathematics and Applications*, 1(2), in press.

[34] Wang, Y. and Y. Wang (2006), Cognitive Informatics Models of the Brain, *IEEE Transactions on Systems, Man, and Cybernetics (*Part C*)*, 36(2), March, 203-207.

[35] Wang, Y., Y. Wang, S. Patel, and D. Patel (2006), A Layered Reference Model of the Brain (LRMB), *IEEE Transactions on Systems, Man, and Cybernetics (*Part C*)*, 36(2), March, 124-133.

[36] Wang, Y., W. Kinsner, and D. Zhang (2009a), Contemporary Cybernetics and its Faces of Cognitive Informatics and Computational Intelligence, *IEEE Transactions on System, Man, and Cybernetics* (Part B), 39(4), 823-833.

[37] Wang, Y., W. Kinsner, J.A. Anderson, D. Zhang, Y. Yao, P. Sheu, J. Tsai, W. Pedrycz, J.-C. Latombe, L.A. Zadeh, D. Patel, and C. Chan (2009b), A Doctrine of Cognitive Informatics, *Fundamenta Informaticae*, 90(3), 203-228.

[38] Wang, Y., Y. Tian, and K. Hu (2011a), Semantic Manipulations and Formal Ontology for Machine Learning Based on Concept Algebra, *International Journal of Cognitive Informatics and Natural Intelligence*, 5(3), 1-29.

[39] Wang, Wang, Y., R.C. Berwick, S. Haykin, W. Pedrycz, W. Kinsner, G. Baciu, D. Zhang, V.C. Bhavsar, and M. Gavrilova (2011b), Cognitive Informatics and Cognitive Computing in Year 10 and Beyond, *International Journal of Cognitive Informatics and Natural Intelligence*, 5(4), 1-21.

[40] Wang, Y. and R.C. Berwick (2012), Formal Models of the Relational Syntactic Rules of the Deductive Grammar of English (DGE), *Journal of Advanced Mathematics and Applications*, 1(3), to appear.

[41] Wardhaugh, R. (2006), *An Introduction to Sociolinguistics*, Wiley-Blackwell, NY.