Lifetime Achievement Award

Natural Language Processing and Computational Linguistics

Junichi Tsujii
Artificial Intelligence Research Center
National Institute for Advanced
Industrial Science and Technology,
Japan
Department of Computer Science
University of Manchester, UK
j-tsujii@aist.go.jp

1. Introduction

As an engineering field, research on natural language processing (NLP) is much more constrained by currently available resources and technologies, compared with theoretical work on computational linguistics (CL). In today's technology-driven society, it is almost impossible to imagine the degree to which computational resources, the capacity of secondary and main storage, and software technologies were restricted when I embarked upon my research career 50 years ago. While these restrictions inevitably shaped my early research into NLP, my subsequent work evolved, according to the significant progress made in associated technologies and related academic fields, particularly CL.

Figure 1 shows the research topics in which I have been engaged. My initial NLP research was concerned with a question answering system, which I worked on during my M.Eng and D.Eng degrees. The research focused on reasoning and language understanding, which I soon found was too ambitious and ill-defined. After receiving my D.Eng., I changed my direction of research, and began to be engaged in processing forms of language expressions, with less commitment to language understanding, machine translation (MT), and parsing. However, I returned to research into reasoning and language understanding in the later stage of my career, with clearer definitions of tasks and relevant knowledge, and equipped with access to more advanced supporting technologies.

In this article, I begin by briefly describing my views on mutual relationships among disciplines related to CL and NLP, and then move on to discussing my own research.

https://doi.org/10.1162/COLI_a_00420

- **Question-Answering System （≒1970s）**
  - Language and Knowledge
  - Reasoning based on linguistic structures
- **Machine Translation（≒1980s~mid90s）**
  - Compositional Translation
  - Transfer based on shallow semantic structures
  - Lexicalized rules
- **Parsing and HPSG（≒1990s-2010）**
  - Computational Equivalence among different grammar formalisms
  - Parsing based on grammar formalisms
- **Structure-based Text Mining for Biology （≒2000s~）**
  - Language and Domain Knowledge
  - Information Extraction using linguistic structures

Figure 1
Research topics.

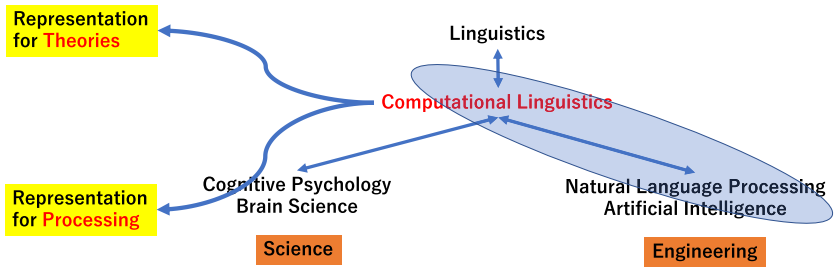


Figure 2
Language-related disciplines.

2. NLP, CL, and Related Disciplines

Language is a complex topic to study, infinitely harder than I first imagined when I began to work in the field of NLP.

There is a whole discipline on the study of language—namely, linguistics. Linguistics is concerned not only with language per se, but must also deal with how humans model the world.[1] The study of semantics, for example, must relate language expressions to their meanings, which reside in the mental models possessed by humans.

Apart from linguistics, there are two fields of science that are concerned with language, that is, brain science and psychology. These are concerned with how humans process language. Then, there are two disciplines in which we are involved—namely, CL and NLP. Natural Language concerns perception of text to understand the true purpose of human language

Figure 2 is a schematic view of these research disciplines. Both of the lower disciplines are concerned with processing language, that is, how language is processed in our minds or our brains, and how computer systems should be designed to process language efficiently and effectively.

1 This statement is a bit of a simplification. Some formal semanticists, like R. Montague, do not assume a "mental" model. They are interested in a universal theory of semantics of language. Language need not be human language (Montague 1970).

The top discipline, linguistics, on the other hand, is concerned with rules that are followed by languages. That is to say, linguists study language as a system. This schematic view is certainly oversimplified, and there are subject fields in which these disciplines overlap. Psycholinguistics, for example, is a subfield of linguistics which is concerned with how the human mind processes language. A broader definition of CL may include NLP as its subfield.

Linguistics is concerned with the rules of language and its proper use

In this article, for the sake of discussion, I adopt narrower definitions of linguistics and CL. In this narrower definition, linguistics is concerned with the rules followed by languages as a system, whereas CL, as a subfield of linguistics, is concerned with the formal or computational description of rules that languages follow.[2]

CL, which focuses on formal/computational description of languages as a system, is expected to bridge broader fields of linguistics with the lower disciplines, which are concerned with processing of language.

Given my involvement in NLP, I would like to address the question of whether the narrowly defined CL is relevant to NLP. The simple answer is yes. However, the answer is not so straightforward, and requires us to examine the degree to which the representations used to describe language as a system are relevant to the representations used for processing language.

Although my colleagues and I have been engaged in diverse research areas, I pick up only on a subset of these, to illustrate how I view the relationships between NLP and CL. Due to the nature of the article, I ignore technical details and focus instead on the motivation of the research and the lessons which I have learned through research.

3. Machine Translation

Mathematical approach towards the rules in linguistics and translation among languages so could be related to Calculus concepts!

Background and Motivation. Following the ALPAC report (Pierce et al. 1966), research into MT had been largely abandoned by academia, with the exception of a small number of institutes (notably, GETA at Grenoble, France, and Kyoto University, Japan). There were only a handful of commercial MT systems, being used for limited purposes. These commercial systems were legacy systems that had been developed over years and had become complicated collections of ad hoc programs. They had become too convoluted to allow for changes and improvements. To re-initiate MT research in academia, we had to have more systematic and disciplined design methodologies.

On the other hand, theoretical linguistics, initiated by Noam Chomsky (Chomsky 1957, 1965) had attracted linguists with a mathematical orientation, who were interested in formal frameworks of describing rules followed by language. Those linguists with interests in formal ways of describing rules were the first generation of computational linguists.

Although computational linguists did not necessarily follow the Chomskyan way of thinking, they shared the general view of treating language as a system of rules. They had developed formal ways of describing rules of language and showed that these rules consisted of different layers, such as morphology, syntax, and semantics, and that each layer required different formal frameworks with different computational powers. Their work had also motivated work on how one could process language by computerizing its rules of language. This work constituted the beginning of NLP research, and resulted in

These rules would really make up the different layers of the text to ultimately understand its meaning as a machine.

2 In this definition, I take research on parsing as part of NLP, since it is concerned with processing of language. Research on parsing algorithms, however, may be quite different in nature from the engineering side of NLP. Thus, the boundary between NLP and CL is not so clear-cut.

the development of parsing algorithms for context-free language, finite-state machines, and so forth.[3] It was natural to use this work as the basis for designing the second generation of MT systems, which was initiated by an MT project (MU project, 1082-1986) led by Prof. M. Nagao (Nagao, Tsujii, and Nakamura 1985).

New technology used today for language translation without need for context

Research Contributions. When I began research into MT in the late 1970s, there was a common view largely shared by the community, which had been advocated by the group of GETA, in France. The view was called the **transfer approach** of MT (Boitet 1987).

The transfer approach viewed translation as a process consisting of three phases: analysis, transfer, and generation. According to linguists, a language is a system of rules. The analysis and generation phases were monolingual phases that were concerned with a set of rules for a single language, the analysis phase using the rules of the source language and the generation phase using the rules of the target language. Only the transfer phase was a bilingual phase.

Hierarchy to create meaning by each level every time and then translate it to the other language to have the smallest difference in meaning

Another view shared by the community was an abstraction hierarchy of representation, called the **triangle of translation**. For example, Figure 3(a)[4] shows the hierarchy of representation used in the Eurotra project, with their definition of each level (Figure 3(b)).

By climbing up such a hierarchy, the differences among languages would become increasingly small, so that the mapping (i.e., the transfer phase) from one language to another would become as simple as possible. Independently of the target language, the goal of the analysis phase was to climb up the hierarchy, while the aim of the generation phase was to climb down the hierarchy to generate surface expressions in the target language. Both phases are concerned only with rules of single languages.

In the extreme view, the top of the hierarchy was taken as the language-independent representation of meaning. Proponents of the interlingual approach claimed that, if the analysis phase reached this level, then no transfer phase would be required. Rather, translation would consist only of the two monolingual phases (i.e., the analysis and generation phases).

However, in Tsujii (1986), I claimed, and still maintain, that this was a mistaken view about the nature of translation. In particular, this view assumed that a translation pair (consisting of the source and target sentences) encodes the same "information". This assumption does not hold, in particular, for a language pair such as Japanese and English, that belong to very different language families. Although a good translation should preserve the information conveyed by the source sentence as much as possible in the target sentence, translation may lose some information or add extra information.[5]

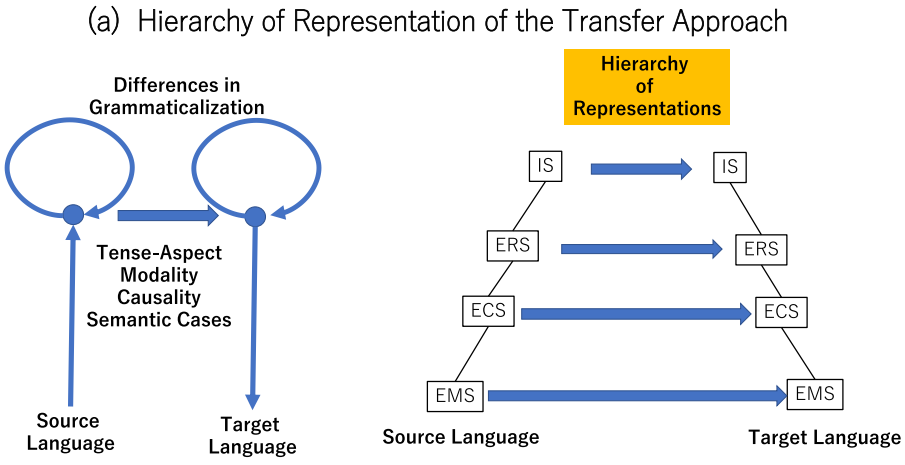
Furthermore, the goal of translation may not be to preserve information but to convey the same pragmatic effects to readers of the translation.

However, there is also some degree of complexity towards translation and it may not be as straightforward as it seems to be

3 The explanation here is simplified. The formal theory of language was not necessarily concerned with human language. They revealed the relationship between classes of language and the computational power of their recognizers. Parsing algorithms of formal languages were studied not necessarily for human languages. Strictly speaking, this research was not conducted as NLP research.

4 The left side of this figure is the transfer with a pre- and post-cycle of adjustment. These two cycles are required to treat language pairs like Japanese and English. The cycles adjust differences in grammaticalization of the two languages. Differences are abundant when we treat languages that belong to very different language families (Tsujii 1982).

5 Apart from the equality of information, the interlingual approach assumed that the language-independent representation consists only of language-independent lexemes. This involved implausible work of defining a set of language-independent concepts.



Transfer Hierarchy for converting to different languages using natural language processing.

Figure 3
(a) Hierarchy of representation of the transfer approach. (b) Hierarchy of representation (Eurotra).

Many things such as empathy, distinction between the chronology of events, emphasis of phrases were lost due to the hierarchy and passing of information without context

More seriously, the abstract level of representation such as Interface Structure[6] in Eurotra focused only on the propositional content encoded in language, and tended to abstract away other aspects of information, such as the speaker's empathy, distinction of old/new information, emphasis, and so on.

To climb up the hierarchy led to loss of information in lower levels of representation. In Tsujii (1986), instead of mapping at the abstract level, I proposed "transfer based on a bundle of features of all the levels", in which the transfer would refer to all levels of representation in the source language to produce a corresponding representation in the target language (Figure 4). Because different levels of representation require different geometrical structures (i.e., different tree structures), the realization of this proposal had to wait for development of a clear mathematical formulation of feature-based

6 IS (Interface Structure) is dependent on a specific language. In particular, unlike the interlingual approach, Eurotra did not assume language-independent leximemes in ISs so that the transfer phase between the two ISs (source and target ISs) was indispensable. See footnote 5.

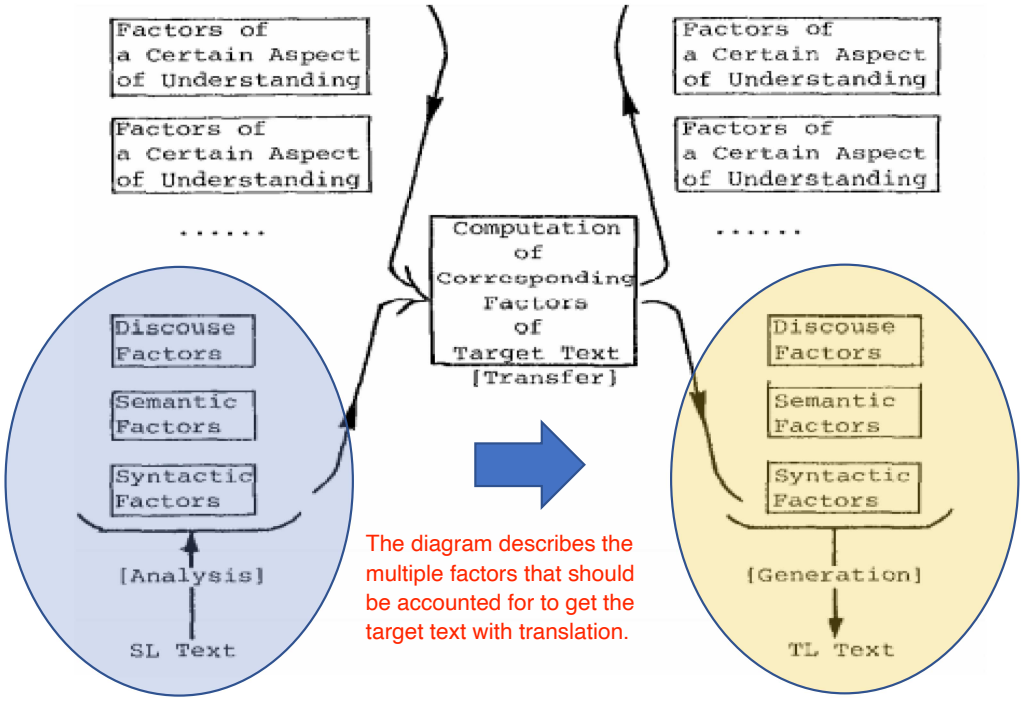


Figure 4
Description-based transfer (Tsujii 1986).

Compositional Semantics would lead to recursive transfer, or use multiple trees to combine the meaning of subphrases to get the overall meaning of the text and use it to translate and convey the full meaning of a text.

representation with reentrancy, which allowed multiple levels (i.e., multiple trees) to be represented with their mutual relationships (see the next section).

Another idea we adopted to systematize the transfer phase was **recursive transfer** (Nagao and Tsujii 1986), which was inspired by the idea of compositional semantics in CL. According to the views of linguists at the time, a language is an infinite set of expressions which, in turn, is defined by a finite set of rules. By applying this finite number of rules, one can generate infinitely many grammatical sentences of the language. Compositional semantics claimed that the meaning of a phrase was determined by combining the meanings of its subphrases, using the rules that generated the phrase. Compositional translation applied the same idea to translation. That is, the translation of a phrase was determined by combining the translations of its subphrases. In this way, translations of infinitely many sentences of the source language could be generated.

Using the compositional translation approach, the translation of a sentence would be undertaken by recursively tracing a tree structure of a source sentence. The translation of a phrase would then be formulated by combining the translations of its subphrases. That is, translation would be constructed in a bottom up manner, from smaller units of translation to larger units. Continuously check for the previous phrases and understand the context.

Furthermore, because the mapping of a phrase from the source to the target would be determined by the lexical head of the phrase, the lexical entry for the head word specified how to map a phrase to the target. In the MU project, we called this **lexicon-driven, recursive transfer** (Nagao and Tsujii 1986) (Figure 5).

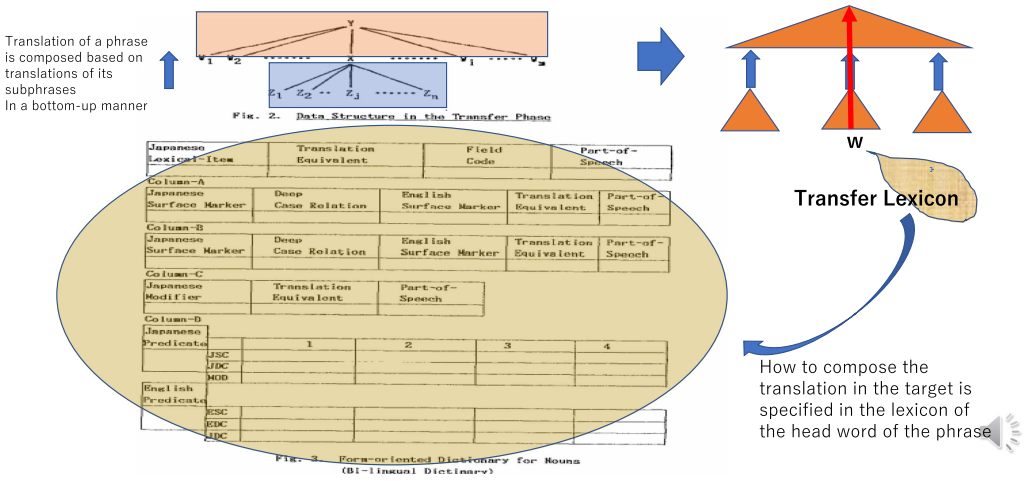


Figure 5
Lexicon-driven recursive structure transfer (Nagao and Tsujii 1986).

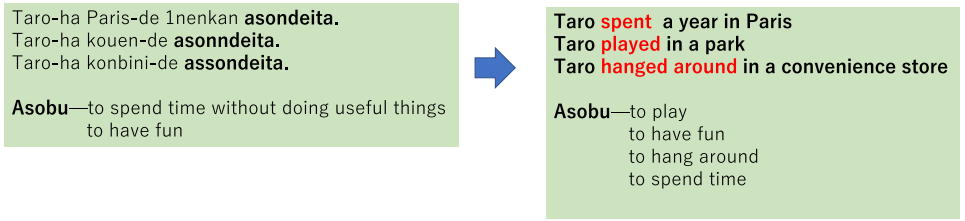


Figure 6
Disambiguation at lexical transfer.

Compared with the first-generation MT systems, which replaced source expressions with target ones in an undisciplined and ad hoc order, the order of transfer in the MU project was clearly defined and systematically performed.

Lessons. Research and development of the second-generation MT systems benefitted from research into CL, allowing more clearly defined architectures and design principles than first-generation MT systems. The MU project successfully delivered English-Japanese and Japanese-English MT systems within the space of four years. Without these CL-driven design principles, we could not have delivered these results in such a short period of time. All types of different information must be handled in translation

However, the differences between the objectives of the two disciplines also became clear. Whereas CL theories tend to focus on specific aspects of language (such as morphology, syntax, semantics, discourse, etc.), MT systems must be able to handle all aspects of information conveyed by language. As discussed, climbing up a hierarchy that focuses on propositional content alone does not result in good translation.

Understanding depends on the details for any project in NLP!

A more serious discrepancy between CL and NLP is the treatment of ambiguities of various kinds. Disambiguation is the single most significant challenge in most NLP tasks; it requires the context in which expressions to be disambiguated occur to be processed. In other words, it requires understanding of context.

The concept of disambiguation relates to my project for sentiment analysis on iMDb move reviews from last year!

Typical examples of disambiguation are shown in Figure 6. The Japanese word *asobu* has a core meaning of "spend time without engaging in any specific useful tasks", and would be translated into "to play", "to have fun", "to spend time", "to hang around", and so on, depending on the context (Tsujii 1986).

Considering context for disambiguation contradicts with recursive transfer, since it requires larger units to be handled (i.e., the context in which a unit to be translated occurs). The nature of disambiguation made the process of recursive transfer clumsy. Disambiguation was also a major problem in the analysis phase, which I discuss in the next section.

Importance of the full picture and context in NLP!

The major (although hidden) general limitation of CL or linguistics is that it tends to view language as an independent, closed system and avoids the problem of understanding, which requires reference to knowledge or non-linguistic context.[7] However, many NLP tasks, including MT, require an understanding or interpretation of language expressions in terms of knowledge and context, which may involve other input modalities, such as visual stimuli, sound, and so forth. I discuss this in the section on the future of research.

End of annotations

4. Grammar Formalisms and Parsing

Background and Motivation. At the time I was engaged in MT research, new developments took place in CL, namely, feature-based grammar formalisms (Kriege 1993).

At its early stage, transformational grammar in theoretical linguistics by N. Chomsky assumed that sequential stages of application of tree transformation rules linked the two levels of structures, that is, deep and surface structures. A similar way of thinking was also shared by the MT community. They assumed that climbing up the hierarchy would involve sequential stages of rule application, which map from the representation at one level to another representation at the next adjacent level.[8] Because each level of the hierarchy required its own geometrical structure, it was not considered possible to have a unified non-procedural representation, in which representations of all the levels co-exist.

This view was changed by the emergence of feature-based formalisms that used directed acyclic graphs (DAGs) to allow reentrancy. Instead of mappings from one level to another, it described mutual relationships among different levels of representation in a declarative manner. This view was in line with our idea of description-based transfer, which used a bundle of features of different levels for transfer. Moreover, some grammar formalisms at the time emphasized the importance of lexical heads. That is, local structures of all the levels are constrained by the lexical head of a phrase, and these constraints are encoded in lexicon. This was also in line with our lexicon-driven transfer.

A further significant development in CL took place at the same time. Namely, a number of sizable tree bank projects, most notably the Penn Treebank and the Lancaster/IBM Treebank, had reinvigorated corpus linguistics and started to have significant impacts on research into CL and NLP (Marcus et al. 1994). From the NLP point of

7 This is overgeneralization. Theoretical linguistics by N. Chomsky explicitly avoided problems related with interpretation and treated language as a closed system. Other linguistic traditions have had more relaxed, open attitudes.

8 Note that transformational grammar considered a set of rules applied to generate surface structures from the deep structure. On the hand, the "climbing-up the hierarchy" model of analysis considered a set of rules to reveal the abstract level of representation from the surface level of representation. The directions are opposite. Ambiguities did not cause problems in transformational grammar.

view, the emergence of large tree banks led to the development of powerful tools (i.e., probabilistic models) for disambiguation.[9]

We started research that would combine these two trends to systematize the analysis phase—that is, parsing based on feature-based grammar formalisms.

Research Contributions. It is often claimed that ambiguities occur because of insufficient constraints. In the analysis phase of the "climbing up the hierarchy" model, lower levels of processing could not refer to constraints in higher levels of representation. This was considered the main cause of the combinatorial explosion of ambiguities at the early stages of climbing up the hierarchy. Syntactic analysis could not refer to semantic constraints, meaning that ambiguities in syntactic analysis would explode.

On the other hand, because the feature-based formalisms could describe constraints at all levels in a single unified framework, it was possible to refer to constraints at all levels, to narrow down the set of possible interpretations.

However, in practice, the actual grammar was still vastly underconstrained. This was partly because we do not have effective ways of expressing semantic and pragmatic constraints. Computational linguists were interested in formal declarative ways for relating syntactic and semantic levels of representation, but not so much in how semantic constraints are to be expressed. To specify semantic or pragmatic constraints, one may have to refer to the mental models of the world (i.e., how humans see the world), or discourse structures beyond single sentences, and so on. These fell outside of the scope of CL research at the time, whose main focus is on grammar formalisms.

Furthermore, it is questionable whether semantics or pragmatics can be used as constraints. They may be more concerned with the plausibility of an interpretation than the constraints which an interpretation should satisfy (for example, see the discussion in Wilks [1975]).

Therefore, even for parsing using feature-based formalisms, issues of disambiguation and how to handle the explosion of ambiguities remained major issues for NLP.

Probabilistic models were one of the most powerful tools for disambiguation and handling the plausibility of an interpretation. However, probabilistic models for simpler formalisms, such as regular and context-free grammars, had to be changed for more complex grammar formalisms. Techniques for handling combinatorial explosion, such as packing, had to be reformulated for feature-based formalisms.

Furthermore, although feature-based formalisms were neat in terms of describing constraints in a declarative manner, the unification operation, which was a basic operation for treating feature-based descriptions, was computationally very expensive. To deliver practical NLP systems, we had to develop efficient implementation technologies and processing architectures for feature-based formalisms.

The team at the University of Tokyo started to study how we could transform a feature-based grammar (we chose HPSG) into effective and efficient representations for parsing. The research included:

1. Design of an abstract machine for processing of typed-feature structures and development of a logic programming system—LiLFeS (Makino et al. 1998; Miyao et al. 2000).

9 There had been attempts to construct probabilistic models without supervision of annotated tree banks (for example, Fujisaki [1984]). They used EM algorithms such as the inside-outside algorithm. However, they could not have brought significant results on their own. Significant progresses were made possible by large treebanks (for example, Fujisaki (1984)).

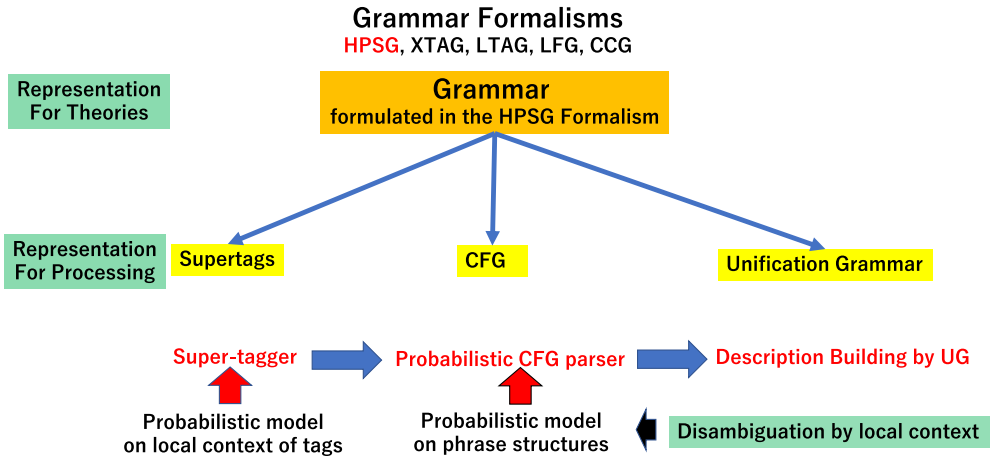


Figure 7
Staged architecture of parsing.

2. Transforming HPSG grammar into a more processing-oriented representation, such as extracting CFG skeletons (Torisawa and Tsujii 1996; Torisawa et al. 2000) and supertags from original HPSG.
3. Packing of feature structures (feature forest) and long-linear probabilistic models (Miyao and Tsujii 2003, 2005, 2008).
4. A staged architecture of parsing based on transformation of grammar formalisms and their probabilistic modeling (Matsuzaki, Miyao and Tsujii 2007; Ninomiya et al. 2010).

A simplified representation of our parsing model is shown in Figure 7. Given a sentence, its representation of all the levels was constructed at the final stage by using the HPSG grammar. Disambiguation took place mainly in the first two phases. The first phase was a supertagger that would disambiguate supertags assigned to words in a sentence. Supertags were derived from the original HPSG grammar and a set of supertags were attached to a word in its lexicon. A suppertagger would choose the most probable sequence of supertags for the given sequence of words. The task was a sequence labeling task, which could be carried out in a very efficient manner (Zhang, Matsuzaki, and Tsujii 2009). This means that the surface local context (i.e., local sequences of supertags) was used for disambiguation, without constructing actual DAGs of features.

The second phase was CFG filtering. A CFG skeleton, which also was derived from the HPSG grammar, was used to check whether sequences of supertags chosen by the first phase could reach a successful derivation tree. The supertagger did not build actual parse trees explicitly to check whether a chosen sequence could reach legitimate derivation trees or not. The second phase of CFG filtering would filter out supertag sequences that could not reach legitimate trees.

The final phase not only built the final representation of all the levels, but it also checked extra constraints specified in the original grammar. Because the first two phases only use partial constraints specified in the HPSG grammar, the final phase would reject results produced by the first two phases if they failed to satisfy these extra constraints.

In this case, the system would backtrack to the previous phases to obtain the next candidate.

All of these research efforts collectively produced a practical efficient parser based on HPSG (Enju).

Lessons. As in MT, CL theories were effective for the systematic development of NLP systems. Feature-based grammar formalisms drastically changed the view of parsing as "climbing up the hierarchy". Moreover, mathematically well-defined formalisms helped the systematic implementation of efficient implementations of unification, transformation of grammar into supertags, CFG skeletons, and so forth. These formalisms also provided solid ground for operations in NLP such as packing of feature structures, and so on, which are essential for treating combinatorial explosion.

On the other hand, direct application of CL theories to NLP did not work, since this would result in extremely slow processing. We had to transform them into more processing-oriented formats, which required significant efforts and time on the NLP side. For example, we had to transform the original HPSG grammar into processing-oriented forms, such as supertags, CFG skeletons, and so on. It is worth noting that, while the resultant architecture was similar to the climbing-up hierarchy processing, each stage in the final architecture was clearly defined and related to each other through the single declarative grammar.

I also note that advances in the fields of computer science/engineering significantly changed what was possible to achieve in NLP. For example, the design of an abstract machine and its efficient implementation for unification in LiLFeS (Makino et al. 1998), effective support systems for maintaining large banks of parsed trees (Ninomiya, Makino, and Tsujii 2002; Ninomiya, Tsujii, and Miyao 2004), and so forth, would be impossible without advances in the broader fields of computer science/engineering and without much improved computational power (Taura et al. 2010).

On the other hand, disambiguation remained the major issue in NLP. Probabilistic models enabled major breakthroughs in terms of solving the problem. Compared with the fairly clumsy rule-based disambiguation that we adopted for the MU project,[10] probabilistic modeling provided the NLP community with systematic ways of handling ambiguities. Combined with large tree banks, objective quantitative comparison of different models also became feasible, which made systematic development of NLP systems possible. However, the error rate in parsing remained (and still remains) high.

While reported error rates are getting lower, measuring the error rate in terms of the number of incorrectly recognized dependency relations was misleading. At the sentence level, the error rate remains high. That is, a sentence in which all dependency relations are correctly recognized remains very rare. Because most of dependency relations are trivial (i.e., pairs of adjacent words or pairs of close neighbors), errors in semantically critical dependencies, such as PP-attachments, scopes of conjunction, etc., remain abundant (Hara, Miyao, and Tsujii 2009).

10 The MU project, like the GETA group in France, took the approach of "procedural grammar" in which rules for disambiguation check the context explicitly to choose plausible interpretation. Without probabilistic models, the approach would be the only option for delivering working systems. However, the analysis phase in this approach becomes clumsy and convoluted (Tsujii, Nakamura, and Nagao 1984; Tsujii et al. 1988).

Even using probabilistic models, there are obvious limits to disambiguation performance, unless a deeper understanding is taken into account. This leads me to the next research topic: language and knowledge.

5. Text Mining for Biomedicine

Background and Motivation. I was interested in the topic of how to relate language with knowledge at the very beginning of my career. At the time, my naiveté led me to believe initially that a large collection of text could be used as a knowledge base and was engaged in research of a question-answering system based on a large text base (Nagao and Tsujii 1973, 1979). However, resources such as a large collection of text, storage capacity, processing speed of computer systems, and basic NLP technologies, such as parsing, were not available at the time.

I soon realized, however, that the research would involve a whole range of difficult research topics in artificial intelligence, such as representation of common sense, human ways of reasoning, and so on. Moreover, the topics had to deal with uncertainty and peculiarities of individual humans. Knowledge or the world models that individual humans have may differ from one person to another. I felt that the research target was ill-defined.

However, through research in MT and parsing in the later stages of my career, I started to realize that NLP research is incomplete if it ignores how knowledge is involved in processing, and that challenging NLP problems are all related to issues of understanding and knowledge. At the same time, considering NLP as an engineering field, I took it to be essential to have a clear definition of knowledge or information with which language is to be related. I would like to avoid too much vagueness of research into commonsense knowledge and reasoning and to restrict our research focus to the relationship between language and knowledge. As a research strategy, I chose to focus on the biomedicine as the application domain. There were two reasons for the choice.

One reason was that microbiology colleagues at the two universities with which I was affiliated told me that, in order to understand life-related phenomena, it had become increasingly important for them to organize pieces of information scattered in a large collection of published papers in diverse subject fields such as microbiology, medical sciences, chemistry, and agriculture. In addition to the large collection of papers, they also had diverse databases that had to be linked with each other. In other words, they had a solid body of knowledge shared by domain specialists that was to be linked with information in text.

The other reason was that there were colleagues at the University of Manchester who were interested in sublanguages. According to the discussion on information formats in a medical sublanguage by the NYU group (Sager 1978) and research into medical terminology at the University of Manchester, focusing on relations between terms and concepts (Ananiadou 1994; Frantzi and Ananiadou 1996; Mima et al. 2002), the biomedical domain had been a natural choice of sublanguage research. The important point here was that information formats in a sublanguage and terminology concepts were defined by the target domain, and not by NLP researchers. Furthermore, domain experts had actual needs and concrete requirements to help solve their own problems in the target domains.

Research Contributions. Although there had been quite a large amount of research into information retrieval and text mining for the biomedical domain, there had been no serious efforts to apply structure-based NLP techniques to text mining in the domain.

To address this, the teams at the University of Manchester and the University of Tokyo jointly launched a new research program in this direction.

Because this was a novel research program, we first had to define concrete tasks to solve, to prepare resources, and to involve not only NLP researchers, but also experts in the target domains.

Regarding the involvement of NLP researchers and domain experts, we found that a few groups in the world also began to be interested in similar research topics. In response to this, we organized a number of research gatherings in collaboration with colleagues around the world, which led to establishment of a SIG (SIGBIOMED) at ACL. The first workshop took place in 2002, collocated with the ACL conference (Workshop 2002). The SIG now organizes annual workshops and co-located shared tasks. It has been expanding rapidly and has become one of the most active SIGs in NLP applications. The research field of application of structure-based NLP to text-mining is broadening to cover clinical/medical domains (Xu et al. 2012; Sohrab et al. 2020), chemistry, and material science domains (Kuniyoshi et al. 2019).

Research contributions by the two teams include the GENIA corpus (Kim et al. 2003; Thompson, Ananiadou, Tsujii 2017), a large repository of acronyms with their original terms (Okazaki, Ananiadou, and Tsujii 2008, 2010), the GENIA POS tagger (Tsuruoka et al. 2005), the BRAT annotation tool (Stenetorp et al. 2012), a workflow design tool for information extraction (Kano et al. 2011), an intelligent search system based on entity association (Tsuruoka, Tsujii, Ananiadou 2008), and a system for pathway construction (Kemper et al. 2010).

The GENIA annotated corpus is one of the most frequently used corpora in the biomedical domain. To see what information domain experts considered important in text and how it was encoded in language, we annotated 2000 abstracts, not only from the linguistic point of view but also from the viewpoint of domain experts. Two types of annotations, namely, linguistic annotations (POS, and syntactic trees) and domain-specific annotations (biological entities, relations, and events) were added to the corpus (Ohta et al. 2006).

Domain-specific annotations were linked with ontologies of the target domain (GENE ontology, anatomy ontology, etc.) which had been constructed by the target domain communities to share information in diverse databases.

To involve domain experts in annotation, we developed a user-friendly annotation tool with intuitive visualization (BRAT), which is now used widely by the NLP community. In close cooperation with domain experts, we defined a set of NLP tasks (Hirshman et al. 2002; Ananiadou, Friedman, and Tsujii 2004; Ananiadou, Kell, and Tsujii 2006; Ananiadou et al. 2020), and developed a set of basic IE tools (Nobata et al. 2008; Miwa et al. 2009; Pyysalo et al. 2012) for solving them, which were to be combined into workflows to meet specific needs of individual groups of domain experts (Kano et al. 2011; Rak et al. 2012).

As a result of this work, we recognized large discrepancies between linguistic units such as words, phrases, and clauses, and domain-specific semantic units, such as named entities, and relations and events that link them together (Figure 8). The mapping between linguistic structures and the semantic ones defined by domain specialists was far more complex than the mapping assumed by computational semanticists.

We soon realized that, as an NLP task, information extraction (IE) was very different from MT. In particular, there were considerable differences between the information that the authors intended to convey and encode in text and the information that the readers (i.e., a group of domain experts) wanted to identify and extract from text. Regardless

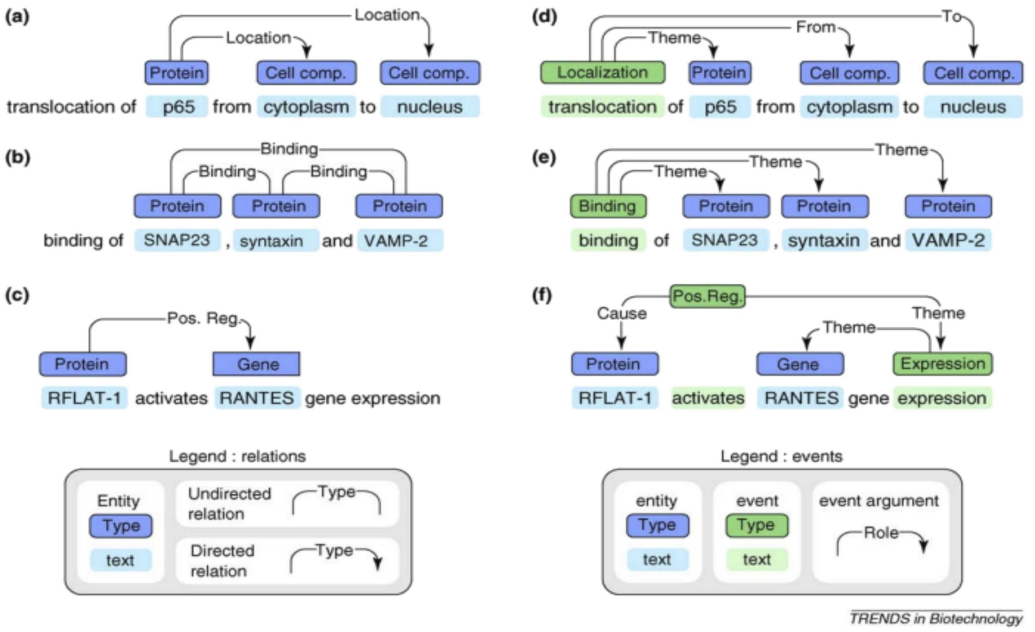


Figure 8
Event and Relation Extraction (Ananiadou et al. 2020).

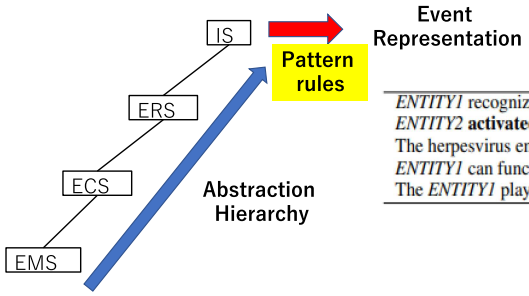
of the information that the authors intended to convey, the reader would identify the information that *they* were interested in.[11]

These characteristics of IE as an NLP task made the mapping from language to information very different from the transfer phase in MT, which attempts to covey the same information in the source and target languages. While the task of named entity recognition (NER) benefited from linguistic structures (i.e., noun phrases and their coordination), linguistic structures would only give cues for the automatic recognition of relations and events, and these cues were to be combined with other cues. The mapping became similar to the transfer based on a bundle of features.

Like the transfer in the higher level of representation, we first used the HPSG parser to climb up the hierarchy to the IS (which we called PAS [predicate-argument structure]), from which we tried to identify a set of pattern-rules to extract events (Figure 9) (Yakushiji 2001, 2006). We assumed that, although extraction patterns based on surface sequences of words may be diverse,[12] this diversity would reduce at a higher level of abstraction—that is, the same approach to simple transfer at the abstract level. Although this approach initially achieved reasonable performance, it soon reached its limit; extracted patterns became increasingly clumsy and convoluted.

11 This means that the reader identifies information in text that may not be the main information the writer intends to convey. The task of IE is essentially concerned with interpretation of text by the reader, and the reader infers diverse sorts of information from text.

12 The information format in String Grammar by the NYU group tried to use surface patterns as rules of sublanguage grammar. I thought this approach would not work on language in published papers. Compared with language in medical records, language in published papers is not so restricted and intertwined with rules of general language.



ENTITY1 recognizes and **activates** *ENTITY2*.
ENTITY2 **activated** by *ENTITY1* are not well characterized.
The herpesvirus encodes a functional *ENTITY1* that **activates** human *ENTITY2*.
ENTITY1 can functionally cooperate to synergistically **activate** *ENTITY2*.
The *ENTITY1* plays key roles by **activating** *ENTITY2*.

Figure 1: Syntactical Variations of "activate"

Figure 9
Event recognition of the climbing-up model (Yakushiji 2006).

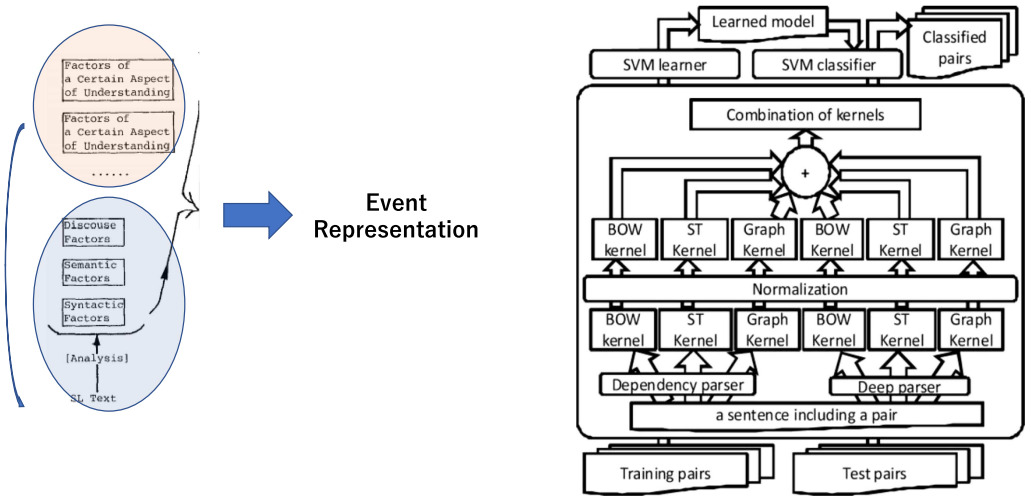


Figure 10
Event recognizer using diverse cues including parse trees (Miwa et al. 2009).

As discussed above, we realized that this was because of the nature of IE tasks, and switched to the approach based on a bundle of features (Figure 10) (Miwa et al. 2009). This shift continued further to the ongoing research, which uses a large language model (BioBERT). In this recent work, linguistic information is assumed to be implicitly embedded in the language model. The information is not represented explicitly in IE systems (Figure 11) (Ju, Miwa, and Ananiadou 2018; Trieu et al. 2020).

On the other hand, our interest in biomedical text mining extended beyond the traditional IE tasks and moved toward coherent integration of extracted information. In this integration, it became apparent that linguistic structures play more significant roles.

For example, claims about an event extracted from different articles often contradict each other. As such, techniques for measuring the credibility or reliability of claims are crucial.

In scientific fields such as biology and medical sciences, claims about an event can be made affirmatively or speculatively, with different degrees of confidence. To measure the degree of confidence of a claim, we have to examine the type of linguistic structure in which the claim is embedded (Zerva et al. 2017; Zerva and Ananiadou 2018). Additionally, depending on the position of an extracted event in a sentence, it may

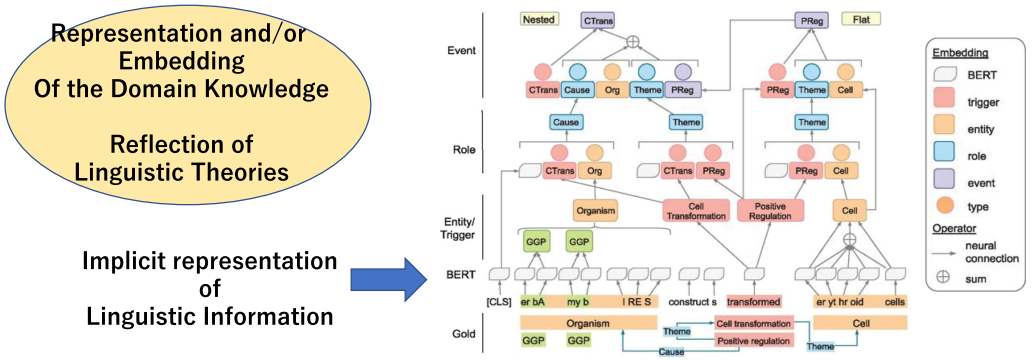


Figure 11
Event extraction using BERT (Trieu 2020).

be considered as a pre-supposed fact, hypothesis, and so forth. The manner in which structural information recognized by a parser can be utilized to detect and integrate contradicting claims remains an important research issue.

Another typical example of an integration problem is the automatic curation of pathways, in which an NLP system is used to combine a set of different events extracted from different articles to build a coherent network of events (Kemper et al. 2010). In this task, a system must be able to decide whether two events reported in different articles can be treated as the same event in a pathway. To do this, the system must be able to detect the biological environments in which two reported events take place, by considering the surrounding contexts of the events. This may also require linguistic structures to be taken into account.

Lessons. By focusing on the biomedical domain, we introduced concrete forms of extra-linguistic knowledge (i.e., domain ontologies built by the target domain communities) and diverse databases, which include manually curated pathway networks. The task of linking information in text with these resources helped to define concrete research topics focusing on the relation between language and knowledge of the target domains. Because scientific communities such as microbiologists have agreed views on which pieces of information constitute their domain knowledge, we can avoid the uncertainty and individuality of knowledge that may have hampered research in the general domain.

Linguistic structures, with which NLP technologies such as parsing have previously been concerned, play less important roles than we initially expected. Nevertheless, ongoing research into the integration of extracted information has started to reassess the importance of linguistic structures.

Another important finding is the nature of human reasoning. The CL and NLP communities tend to consider reasoning as a kind of logical process based on symbolic knowledge. However, the actual reasoning that the experts in the biomedical domain perform may not be so symbolic in nature.

For example, we found that the reasoning carried out by domain experts on pathways is based on similarities between entities. For example, they infer that a protein A is likely to be involved in an event by observing a reported fact that a protein B, which is similar to protein A, is involved in the same type of event. The similarity between protein A and B is based on the similarities between their 3D structures. Because such similarities among proteins are scarcely manifested in their occurrences

in text, large language models trained on a large collection of papers would be unable to capture their similarities. Symbolic domain ontologies (i.e., a classification scheme of biological entities) also fail to capture such fine-grained similarities. Accordingly, it may be necessary to use heterogenous sources of information, such as databases of protein structures, large collections of pathways, and so on, to capture such semantic similarities among entities and to carry out reasoning based on them.

6. Future – Conclusions

We have witnessed the rapid progress and significant changes that neural network (NN) models and deep learning (DL) have brought to the field of NLP. This is a typical example in which advances in broader fields of computer science/engineering open up new opportunities to change and enhance the NLP field.

The changes brought by NN and DL are broad and have had a profound impact not only on NLP but also visual/image processing, speech/signal processing, and many other areas of artificial intelligence. It is a paradigm shift.

The new paradigm has significantly improved the performance of diverse NLP tasks. Furthermore, I expect it will contribute significantly toward solving the most challenging NLP problems, by integrating NLP with the processing of other information modalities (images, sounds, haptics, etc.), and with knowledge processing, and so on.

In technological fields such as image and speech processing, reasoning based on knowledge traditionally used different modeling and processing techniques. They now share the same technological basis of NN and DL. It is becoming much easier to integrate heterogeneous forms of processing, meaning that carrying out NLP in multimodal contexts and NLP with knowledge bases are far more feasible than we previously thought. The research teams of the institutions with which I am affiliated are now working on these directions (Kumar Sahu et al. 2019; Iso et al. 2020; Christopoulou, Miwa, and Ananiadou 2021).

On the negative side, NLP based on large language models is increasingly separating itself from other research disciplines that involve the study of language. The black box nature of NN and DL also makes the analytical methods way of assessing NLP systems difficult.

Although the characteristics are very different, I fear that the paradigm may encounter similar difficulties to those suffered by first-generation MT systems. One could improve the overall performance by tweaking computational models, but without rational and systematic analysis of problems, this failed to solve real difficulties and recognize the limit of the technology.

As revealed through detailed analysis of parsing errors, even when the overall quantitative performance improved, semantically crucial errors of specific types remained unsolved.

Without analysis based on theories provided by other language-related disciplines, erratic and unexpected behaviors of NN-based NLP systems will remain and limit potential applications.

On the other hand, CL tends to treat language as a closed system or focus on study on specific aspects of regularities that language show. By examining what takes place in NLP systems, together with NLP practitioners, CL researchers would be able to enrich the scope of their theories and to provide a theoretical basis for analytic assessment of NLP systems.

It is the time to re-connect NLP and CL.

Acknowledgments

I was introduced to the field of NLP by my long-time mentor, Professor Makoto Nagao, who was a recipient of the Lifetime Achievement Award (2003). He passed away last May. It is unfortunate that I could not share my honor and happiness with him. In my career of almost 50 years, I have conducted research into NLP at several institutes worldwide, including Kyoto University; CNRS (GETA, Grenoble), France; University of Manchester, UK; the University of Tokyo; and Microsoft Research, China. I receive the honor on behalf of the colleagues, research fellows, and students who worked with me at these institutions. Without them, my research could not have progressed in the way that it has. I deeply appreciate their support.

References

Ananiadou, Sophia, Carol Friedman, and Jun'ichi Tsujii. 2004. Introduction: Named entity recognition in biomedicine. *Journal of Biomedical Informatics*, 37(6):393–395. https://doi.org/10.1016/j.jbi.2004.08.011

Ananiadou, Sophia. 1994. A methodology for automatic term recognition. In *Proceedings of COLING*, pages 1034–1038. https://doi.org/10.3115/991250.991317

Ananiadou, Sophia, Douglas B. Kell, and Junichi Tsujii. 2006. Text mining and its potential applications in systems biology. *Trends in Biotechnology*, 24(12):571–579. https://doi.org/10.1016/j.tibtech.2006.10.002, PubMed: 17045684

Ananiadou, Sophia, Sampo Pyysalo, Junichi Tsujii, and Douglas B. Kell. 2010. Event extraction for systems biology by text mining the literature. *Trends in Biotechnology*, 28(7):381–390. https://doi.org/10.1016/j.tibtech.2010.04.005, PubMed: 20570001

Boitet, Christian. 1987. Current state and future outlook of the research at GETA. In *Machine Translation Summit*, Hakone.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton. https://doi.org/10.1515/9783112316009

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. MIT Press. https://doi.org/10.21236/AD0616323, PubMed: 14125365

Christopoulou, Fenia, Makoto Miwa, Sophia Ananiadou. 2021. Distantly supervised relation extraction with sentence reconstruction and knowledge base priors. In *Proceedings of NAACL-HLT*, pages 11–26. https://doi.org/10.18653/v1/2021.naacl-main.2

Enju. Fast, accurate, deep parser for English. https://mynlp.is.s.u-tokyo.ac.jp/enju/references.html

Frantzi, Katerina T. and Sophia Ananiadou. 1996. Extracting nested collocations. In *Proceedings of COLING*, pages 41–46. https://doi.org/10.3115/992628.992639

Fujisaki, Tesunosuke. 1984. A stochastic approach to sentence parsing. In *Proceedings of COLING/ACL84*, pages 16–19. https://doi.org/10.3115/980491.980496

Hara, Tadayoshi, Yusuke Miyao, Jun'ichi Tsujii. 2009. Descriptive and empirical approaches to capturing underlying dependencies among parsing errors. In *Proceedings of EMNLP*, pages 1162–1171. https://doi.org/10.3115/1699648.1699662

Hirschman, Lynette, Jong C. Park, Jun'ichi Tsujii, Limsoon Wong, Cathy H. Wu. 2002. Accomplishments and challenges in literature data mining for biology. *Bioinformatics*, 18(12):1553–1561. https://doi.org/10.1093/bioinformatics/18.12.1553, PubMed: 12490438

Iso, Hayate, Yui Uehara, Tatsuya Ishigaki, Hiroshi Noji, Eiji Aramaki, Ichiro Kobayashi, Yusuke Miyao, Naoaki Okazaki, and Hiroya Takamura. 2020. Learning to select, track, and generate for data-to-text. *Journal of Natural Language Processing*, 27(3):599–626. https://doi.org/10.5715/jnlp.27.599

Kano, Yoshinobu, Makoto Miwa, K. Bretonnel Cohen, Lawrence Hunter, Sophia Ananiadou, and Jun'ichi Tsujii. 2011. U-Compare: A modular NLP workflow construction and evaluation system. *IBM Journal of Research and Development*, 55(3):11. https://doi.org/10.1147/JRD.2011.2105691

Kemper, Brian, Takuya Matsuzaki, Yukiko Matsuoka, Yoshimasa Tsuruoka, Hiroaki Kitano, Sophia Ananiadou, and Jun'ichi Tsujii. 2010. PathText: a text mining integrator for biological pathway visualizations. *Bioinformatics*, 26(12):374–381. https://doi.org/10.1093/bioinformatics/btq221, PubMed: 20529930

Kim, Jin-Dong, Tomoko Ohta, Yuka Tateisi, Jun'ichi Tsujii. 2003. GENIA corpus—A semantically annotated corpus for bio-textmining. In *Proceedings of ISMB (Supplement of Bioinformatics)*, pages 180–182. https://doi.org/10

.1093/bioinformatics/btg1023, PubMed: 12855455

Kriege, Hans-Ulrich. 1993. Typed feature formalisms as a common basis for linguistic specification. In *Workshop on Machine Translation and Lexicon (WMTL)* 1993: Machine Translation and the Lexicon, pages 99–119. https://doi.org/10.1007/3-540-59040-4_23

Kumar Sahu, Sunil, Fenia Christopoulou, Makoto Miwa, Sophia Ananiadou. 2019. Inter-sentence relation extraction with document-level graph convolutional neural network. *arXiv preprint arXiv:1906.04684*.

Kuniyoshi, Fusataka, Jun Ozawa, Mikiya Fujii, Koji Morikawa, Toru Nakata, Takehiro Tanaka, Emiko Igaki, Junichi Hibino, Masaki Kiyono, and Makoto Miwa. 2019. Graph representation for synthesis process extraction from inorganic material literature. IEICE Technical Report 119(212), pages 7–12.

Makino, Takaki, Minoru Yoshida, Kentaro Torisawa, and Jun'ichi Tsujii. 1998. LiLFes—Towards a practical HPSG parser. In *Proceedings of COLING-ACL*, pages 807–811. https://doi.org/10.3115/980432.980702

Marcus, Michell P., Grace Kim, Mary Ann Marcinkiewicz, Robert MacIntyre, Ann Bies, Mark Ferguson, Karen Katz, and Britta Schasberger. 1994. The Penn Treebank: Annotating predicate-argument structure. In *ARPA Human Language Technology Workshop*. https://doi.org/10.3115/1075812.1075835

Matsuzaki, Takuya, Yusuke Miyao, and Jun'ichi Tsujii. 2007. Efficient HPSG parsing with supertagging and CFG-filtering. In *Proceedings of IJCAI*, pages 1671–1676.

Ju, Meizhi, Makoto Miwa, and Sophia Ananiadou. 2018. A neural layered model for nested named entity recognition. In *Proceedings of NAACL-HLT*, pages 1446–1459. https://doi.org/10.18653/v1/N18-1131

Mima, Hideki, Sophia Ananiadou, Goran Nenadic, and Jun'ichi Tsujii. 2002. A methodology for terminology-based knowledge acquisition and integration. In *Proceedings of COLING*. https://doi.org/10.3115/1072228.1072311

Miyao, Yusuke, Takaki Makino, Kentaro Torisawa, and Jun'ichi Tsujii. 2000. The LiLFeS abstract machine and its evaluation with the LinGO grammar. *Natural Language Engineering*, 6(1):47–61. https://doi.org/10.1017/S1351324900002400

Miyao, Yusuke and Jun'ichi Tsujii. 2003. A model of syntactic disambiguation based on lexicalized grammars. In *Proceedings of CoNLL*, pages 1–8. https://doi.org/10.3115/1119176.1119177

Miyao, Yusuke and Jun'ichi Tsujii. 2005. Probabilistic disambiguation models for wide-coverage HPSG parsing. In *Proceedings of ACL*, pages 83–90. https://doi.org/10.3115/1219840.1219851

Miyao, Yusuke and Jun'ichi Tsujii. 2008. Feature forest models for probabilistic HPSG parsing. *Computational Linguistics*, 34(1):35–80. https://doi.org/10.1162/coli.2008.34.1.35

Miwa, Makoto, Rune Sætre, Yusuke Miyao, and Jun'ichi Tsujii. 2009. A rich feature vector for protein-protein interaction extraction from multiple corpora. In *Proceedings of EMNLP*, pages 121–130. https://doi.org/10.3115/1699510.1699527

Montague, Richard. 1970. Universal grammar. *Theoria*, 36:373–398. https://doi.org/10.1111/j.1755-2567.1970.tb00434.x

Nagao, Makoto and Jun'ichi Tsujii. 1973. Mechanism of deduction in a question-answering system with natural language input. In *Proceedings of IJCAI*, pages 285–290.

Nagao, Makoto and Jun'ichi Tsujii. 1979. S-Net: A foundation for knowledge representation languages. In *Proceedings of IJCAI*, pages 617–624.

Nagao, Makoto, Jun'ichi Tsujii, and Jun-ichi Nakamura. 1985. The Japanese government project for machine translation. *Computational Linguistics*, 11(2–31):91–110.

Nagao, Makoto and Jun'ichi Tsujii. 1986. The transfer phase of the MU machine translation system. In *Proceedings of COLING*, pages 97–103. https://doi.org/10.3115/991365.991392

Ninomiya, Takashi, Takaki Makino, and Jun'ichi Tsujii. 2002. An indexing scheme for typed feature structures. In *Proceedings of COLING*. https://doi.org/10.3115/1071884.1071908

Ninomiya, Takashi, Takuya Matsuzaki, Yusuke Miyao, Yoshimasa Tsuruoka, and Jun'ichi Tsujii. 2010. HPSG parsing with a supertagger. In *Trends in Parsing Technology*, pages 243–256. https://doi.org/10.1007/978-90-481-9352-3_14

Ninomiya, Takashi, Jun'ichi Tsujii, and Yusuke Miyao. 2004. A persistent feature-object database for intelligent text archive systems. In *Proceedings of IJCNLP*,

pages 197–205. https://doi.org/10.1007/978-3-540-30211-7_21

Nobata, Chikashi, Philip Cotter, Naoaki Okazaki, Brian Rea, Yutaka Sasaki, Yoshimasa Tsuruoka, Jun'ichi Tsujii, and Sophia Ananiadou. 2008. Kleio: A knowledge-enriched information retrieval system for biology. In *Proceedings of SIGIR*, pages 787–788.

Ohta, Tomoko, Yuka Tateisi, Jin-Dong Kim, Akane Yakushiji, and Jun'ichi Tsujii. 2006. Linguistic and biological annotations of biological interaction events. In *Proceedings of LREC*, pages 1402–1405.

Pierce, John R. and John B. Carroll, et al. 1966. Language and Machines — Computers in Translation and Linguistics. ALPAC report, National Academy of Sciences, National Research Council, Washington, DC.

Pyysalo, Sampo, Tomoko Ohta, Makoto Miwa, HC Cho, Junichi Tsujii, and Sophia Ananiadou. 2012. Event extraction across multiple levels of biological organization. *Bioinformatics*, 28(18):i575–i581. https://doi.org/10.1093/bioinformatics/bts407, PubMed: 22962484

Okazaki, Naoaki, Sophia Ananiadou, and Jun'ichi Tsujii. 2008. A discriminative alignment model for abbreviation recognition. In *Proceedings of COLING*, pages 657–664. https://doi.org/10.3115/1599081.1599164

Okazaki, Naoaki, Sophia Ananiadou, and Jun'ichi Tsujii. 2010. Building a high-quality sense inventory for improved abbreviation disambiguation. *Bioinformatics*, 26(9):1246–1253. https://doi.org/10.1093/bioinformatics/btq129, PubMed: 20360059

Rak, Rafal, Andrew Rowley, William Black, and Sophia Ananiadouak. 2012. Argo: An integrative, interactive, text mining-based workbench supporting curation. *Database*, 2012:bas010. https://doi.org/10.1093/database/bas010, PubMed: 22434844

Sager, Naomi. 1978. Natural language information formatting: The automatic conversion of texts to a structured data base. *Advances in Computers*, Volume: 89–162. https://doi.org/10.1016/S0065-2458(08)60391-5

Sohrab, Mohammad Golam, Khoa Duong, Makoto Miwa, Goran Topić, Ikeda Masami, and Takamura Hiroya. 2020. BENNERD: A neural named entity linking system for COVID-19. In *Proceedings of EMNLP*, pages 182–188. https://doi.org/10.18653/v1/2020.emnlp-demos.24

Stenetorp, Pontus, Sampo Pyysalo, Goran Topic, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: A web-based tool for NLP-assisted text annotation. In *Proceedings of EACL*, pages 102–107.

Taura, Kenjiro, Takuya Matsuzaki, Makoto Miwa, Yoshikazu Kamoshida, Daisaku Yokoyama, Nan Dun, Takeshi Shibata, Choi Sung Jun, and Jun'ichi Tsujii. 2010. Design and implementation of GXP Make—A workflow system based on Make. *eScience*, 2010:214–221. https://doi.org/10.1109/eScience.2010.43

Thompson, Paul, Sophia Ananiadou, and Jun'ichi Tsujii. 2017. The GENIA corpus: Annotation levels and applications. In *Handbook of Linguistic Annotation*, pages 1395–1432. Springer. https://doi.org/10.1007/978-94-024-0881-2_54

Torisawa, Kentaro and Jun'ichi Tsujii. 1996. Computing phrasal-signs in HPSG prior to parsing. In *Proceedings of COLING*, pages 949–955. https://doi.org/10.3115/993268.993332

Torisawa, Kentaro, Kenji Nishida, Yusuke Miyao, and Jun'ichi Tsujii. 2000. An HPSG parser with CFG filtering. *Natural Language Engineering*, 6(1):63–80. https://doi.org/10.1017/S1351324900002412

Trieu, Hai-Long, Thy Thy Tran, Anh-Khoa Duong Nguyen, Anh Nguyen, Makoto Miwa, and Sophia Ananiadou. 2020. DeepEventMine: End-to-end neural nested event extraction from biomedical texts. *Bioinformatics*, 36(19):4910–4917. https://doi.org/10.1093/bioinformatics/btaa540, PubMed: 33141147

Tsuruoka, Yoshimasa, Yuka Tateishi, Jin-Dong Kim, Tomoko Ohta, John McNaught, Sophia Ananiadou, and Jun'ichi Tsujii. 2005. Developing a robust part-of-speech tagger for biomedical text. In *Proceedings of Panhellenic Conference on Informatics*, pages 382–392. https://doi.org/10.1007/11573036_36

Tsuruoka, Yoshimasa, Jun'ichi Tsujii, and Sophia Ananiadou. 2008. FACTA: A text search engine for finding associated biomedical concepts. *Bioinformatics*, 24(21):2559–2560. https://doi.org/10.1093/bioinformatics/btn469, PubMed: 18772154

Tsujii, Jun'ichi. 1982. The transfer phase in an English-Japanese translation system. In *Proceedings of COLING*, pages 383–390. https://doi.org/10.3115/991813.991875

Tsujii, Jun'ichi, Jun-ichi Nakamura, and Makoto Nagao. 1984. Analysis grammar of

Japanese in the MU-Project: A procedural approach to analysis grammar. In *Proceedings of COLING*, pages 267–274. https://doi.org/10.3115/980431.980548

Tsujii, Jun'ichi. 1986. Future directions of machine translation. In *Proceedings of COLING*, pages 655–668. https://doi.org/10.3115/991365.991558

Tsujii, Jun'ichi, Yukiyoshi Muto, Yuuji Ikeda, and Makoto Nagao. 1988. How to get preferred readings in natural language analysis. In *Proceedings of COLING*, pages 683–687. https://doi.org/10.3115/991719.991777

Wilks, Yorick. 1975. A preferential, pattern-seeking, semantics for natural language inference. *Artificial Intelligence*, 6:53–74. https://doi.org/10.1016/0004-3702(75)90016-8

Workshop. *Proceedings of the ACL-02 Workshop on Natural Language Processing in the Biomedical Domain*. 2002. Anthology ID:W02-03.

Xu, Yan, Yining Wang, Tianren Liu, Junichi Tsujii, and Eric I-Chao Chang. 2012. An end-to-end system to identify temporal relation in discharge summaries: 2012 i2b2 challenge. *Journal of the American Medical Informatics Association*, 20(5):849–858. https://doi.org/10.1136/amiajnl-2012-001607, PubMed: 23467472

Yakushiji, Akane, Yuka Tateisi, Yusuke Miyao, and Jun'ichi Tsujii. 2001. Event extraction from biomedical papers using a full parser. In *Proceedings of Pacific Symposium on Biocomputing*, pages 408–419. https://doi.org/10.1142/9789814447362_0040, PubMed: 11262959

Yakushiji, Akane, Yusuke Miyao, Tomoko Ohta, Yuka Tateisi, and Jun'ichi Tsujii. 2006. Automatic construction of predicate-argument structure patterns for biomedical information extraction. In *Proceedings of EMNLP*, pages 284–292. https://doi.org/10.3115/1610075.1610116

Zerva, Chrysoula, Riza Batista-Navarro, Philip Day, and Sophia Ananiadou. 2017. Using uncertainty to link and rank evidence from biomedical literature for model curation. *Bioinformatics*, 33(23):3784–3792. https://doi.org/10.1093/bioinformatics/btx466, PubMed: 29036627

Zerva, Chrysoula and Sophia Ananiadou. 2018. Paths for uncertainty: Exploring the intricacies of uncertainty identification for news. In *Proceedings of Workshop on Computational Semantics beyond Events and Roles (SemBEaR)*, pages 6–20. https://doi.org/10.18653/v1/W18-1302

Zhang, Yao-zhong, Takuya Matsuzaki, Jun'ichi Tsujii. 2009. HPSG supertagging: A sequence labeling view. In *Proceedings of IWPT*, pages 210–213. https://doi.org/10.3115/1697236.1697277