

A DEVELOPED ANALYSIS of TYPE COERCION based on TYPE THEORY and CONVENTIONALITY

Seohyun Im¹ and Chunngmin Lee²

¹ Seoul National University seohyunim71@gmail.com

² Seoul National University clee@snu.ac.kr

Abstract. This paper aims to propose a developed analysis of the type coercion phenomenon such as *begin the book* by introducing Type Theory and Conventional Non-linguistic Context, making a distinction between linguistic and non-linguistic context. We argue that linguistic and non-linguistic context as well as the lexical meaning of the words are deeply involved in the interpretation of the type-coerced construction. In the lexical semantic level, the type-coerced construction is ambiguous. Although its linguistic context can decrease the number of possible interpretations of the construction, it is still ambiguous until its non-linguistic context disambiguates the meaning of the construction. More importantly, we propose that the lexical meaning of a word is a conventionalized meaning under the assumption of a conventional non-linguistic context linked to the word. The context holds in the compositional process. Therefore, a type-coerced construction has a preferred interpretation derived from its conventional non-linguistic context, if no specific non-linguistic context (situation of utterance) is provided and its linguistic context is neutral. For instance, the preferred interpretation of *begin the book* is *begin reading the book*, because the conventional non-linguistic context of *book* is the situation of reading the book. However, the preference is only a probability and the construction is still ambiguous.

1 Introduction

In this paper, we aim to give a developed analysis of type coercion phenomena such as *begin the book* explained in the early Generative Lexicon theory (GL: [PJ1995], [PJms]). For that purpose, we adopt the type theory based on Type Compositional Logic (TCL: [AN2011]). In addition, we introduce the “conventional non-linguistic context”, distinguishing between linguistic and non-linguistic context.

One of the strong points of the GL is to explain well the polymorphic behavior of argument selection by predicates, as shown in (1).

- (1) a. John **began** *reading the book*.
- b. John **began** *to read the book*.
- c. John **began** *the book*.

The GL provides the methodology - qualia structure and type coercion - to explain the polymorphism of the *begin* construction and recover the missing predicate in a construction such as (1c). However, the type coercion mechanism of the GL brings up the following questions:

- What conditions allow the type coercion of *begin*?
- What is the mechanism of recovering the missing predicate in the type coercion construction with *begin*?
- What are the effects of context on type coercion and its interpretation?
- What is the relation between type coercion and (lexical) semantics and pragmatics?

We go through the following steps to answer the above questions.

1. We propose the semantic type of *begin* and its selection restriction. (section 2)
2. We propose the semantic type of *book* and its distributional constraint. (section 3)
3. We consider the condition which allows the type coercion.(section 4)
4. We discuss the interpretation of the type-coerced construction and the effects of context. (section 5, 6)
5. We conclude our argument . (section 7)

In the next section, we explore the semantic type of the aspectual verb *begin* in English.

2 The Semantic Type of *BEGIN*

According to the GL, the verb *begin* requires an *event*-type object argument, because it is an *aspectual*-type verb. However, this type assignment is not specific enough to explain the type coercion phenomenon of *begin*. In this paper, we divide the aspectual verbs into more specific subtypes, following [IS2013]. The verb *begin* is classified as a *begin*-type, distinguished from other aspectual verbs such as *finish*, *continue*, and *stop*. The *begin*-type verbs select only *process* or *accomplishment*-type expressions as their object arguments³. Consider the following examples:

- (2) a. Kern began **building a house**. (accomplishment)
b. He began **working**. (process)

In (2), *building a house* (2a) is *accomplishment*-type and *working* (2b) is *process*-type (2b). On the contrary, *begin* takes neither an *achievement* nor a *state*-type gerundive construction. The sentence in (3a) is ungrammatical because *buying the modern painting* represents an *achievement*-type eventuality.

³ We do not discuss here the other subtypes in the *aspectual*-type verb class.

- (3) a. *He had begun *buying* the modern **painting**.
 b. He had begun *buying* the modern **paintings**.

In (3b), *begin* allows the gerundive construction in which *buying* takes a plural object, since the gerundive is **process** type, not **achievement** type⁴. The argument selection restriction of the **begin**-type verbs is explained well by its event structure ([IS2013]).

- (4) Event Structure of the **begin**-type verbs
 se1: state = not_in_process(e2)
 se2: process = beginning(x, e2)
 se3: process = in_process(e2)

For an event to be in process after its beginning, it should at least belong to **process** or **accomplishment**.

To sum up, we propose the semantic type of *begin* and the argument structure - argument selection restriction - of the **begin**-type verb class as follows:

- (5) *begin*
 a. **semantic type** = **begin**
 b. **argument structure** (selection restriction) of **begin**-type verbs
 arg1 = agent: top
 arg2 = event: {**process**, **accomplishment**}

The argument structure above represents the type selection restriction on the arguments of *begin*. It implies that *begin* cannot take the **achievement** or **state** type ones out of the verbs governing the NP *the book*. In the next section, we explore the semantic type of *book* and its governing predicates.

3 The Semantic Type of the Noun *BOOK*

The semantic type of the noun *book* needs to be more specific than **phys_object-info**, which was proposed in the GL, to cover its distributional behavior, although it is right to consider it as a complex type. In this paper, we propose a tentative semantic type of *book* below.

- The Semantic Type of *book* (tentative proposal)
 [text_info]· [info_container_made_by_binding_papers]· [goods]

The dot type of *book* has many entailments related with its linguistic context - expressions co-occurring with the noun *book*. The semantic type **text.info** shows the entailments below ⁵:

⁴ This kind of type shifting from **achievement** to **process** is a well-known phenomenon.

⁵ In this work, we only consider the verbs which take *the book* as their direct object. We will extend the corpus so that we can define the complete semantic type of *book* in the future work.

- **text_info**

\rightarrow_{entail}

[*write, read, translate, publish, digitize, evaluate, underestimate, criticize, etc.*]

The **text_info** subtype means that a book is text-type information written in a language and thus is readable and translatable⁶. Moreover, it can be published, digitized, or evaluated because it is a kind of information. The **text_info** is a subtype of **info** and thus inherits all the properties of it. The subtype triggers entailments related with the governing predicates of the noun *the book* listed above.

The second subtype **info_container_made_by_binding_papers** inherits from **artifact** (**artifact** \subset **physical_object**). We show entailments triggered by this type as follows.

- **info_container_made_by_binding_papers**

\rightarrow_{entail}

[*take, carry, put, place, position, pack, tear, burn, weight, borrow, own, have, lend, etc.*]

The list of verbs includes motion verbs and possession or change_of_possession verbs. Since it is made by binding papers, entailments such as tearing are allowed.

Finally, the subtype **goods** entails all events related to buying or selling. We show the entailments below:

- **goods**

\rightarrow_{entail}

[*buy, sell, promote, market, advertise, etc.*]

In order to represent the complete semantic type of the noun *book*, we need to analyze its linguistic context in more detail. In this paper, we suggest its tentative dot-type, pointing out that **phys_object·info** is not specific enough to explain the linguistic behavior of the noun *book* and type coercion. Nevertheless, it is a crucial argument that the predicates taking *book* are derived from the dot-type of *book*. Based on the semantic types of *begin* and *book* and their distributional restriction, we argue about the type coercion phenomenon of *begin the book* in the next section.

4 Compositional Meaning and Type Coercion

Relying on the semantic type and selection restriction of *begin* and *book* given by their lexical meaning, they combine with each other by type coercion in the compositional process. Type coercion of **begin**-type verbs has some constraints. A part of the constraints were discussed in [PJPB1996]. Adopting their argument, we propose the constraints on type coercion of *begin the book*.

⁶ In addition, there are different types of book such as an audio book, a video book, and an e-book. However, the distinction will be studied in the future work. we consider a book to represent a paper book here.

• **Constraints on type coercion of *begin the book***

- only in the control construction of *begin*;
- when the subject is animate;
- only when the missing predicate belongs to a **process** or an **accomplishment** type verb class;
- and only when the missing predicate is a two-place verb which takes a subject and an object.

The first two constraints are explained in detail in [PJPB1996]. We focus on the last two in this paper. Given the list of governing verbs of *the book*, *begin* does not take the **state** or **achievement** type ones out of them. Consider the following examples.

- (7) a. *John began **having** the book. (**state**)
b. *John began **losing** the book. (**achievement**)

The verb *have* cannot combine with *begin* because it is a **state** type and thus violates the semantic type restriction on its theme argument. In the same way, *losing* also cannot head the gerundive phrase, the object of *begin*. Therefore, these verbs are not candidates for the missing predicate in type coercion. The following examples show that type coercion is allowed when the head of the gerundive construction is **accomplishment** or **process** type.

- (8) a. John began **reading** the book. (**accomplishment**)
b. John began (reading) the book.

In the above example, *reading the book* can combine with *begin* because it is an **accomplishment** eventuality. As shown in (8b), *reading* can be ellipsed and recovered.

Even when a gerundive satisfies the selection restriction of *begin* aspectually, it cannot be ellipsed if it is not a two-place predicate. See the following example:

- (9) a. John began **giving** me books.
b. *John began me books.

Since *books* is plural, *giving me books* is a **process** aspectual type expression, although *give* is **achievement** type. Therefore, the sentence in (9a) is grammatical but *begin* does not allow the ellipsis of *giving*, as shown in (9b). This example shows that type coercion is allowed only when the missing predicate takes a subject and a direct object.

In the above sections, we explored which verbs can be the head of the gerundive construction with *the book* in the *begin* construction. However, there still is the problem of recovering an appropriate predicate in the interpretation of *begin the book*. The next section discusses it.

5 Interpretation: Ambiguity and Linguistic and Non-linguistic Context

Regarding type coercion, one of the biggest issues is to clarify how to recover the missing predicate in the interpretation of a type-coerced construction. As argued above, the constraints on type coercion exclude many predicates from the list of potential recoverable predicates. Nevertheless, there still remain many verbs as candidates. For instance, the following verbs can be recovered in the interpretation of *begin the book*:

- *read, write, publish, print, review, translate, promote, market, use, etc.*

In order to solve this issue, we suggest the distinction between linguistic and non-linguistic context. A linguistic context means the distributional behavior of an expression: that is, the other expressions co-occurring with the expression in a sentence. On the other hand, a non-linguistic context is defined as the situation in which the utterance occurs.

The linguistic context of *begin the book* (e.g., the subject) can lead us to prefer a specific interpretation of the type coerced construction. Moreover, it sometimes excludes some predicates from the candidate list for recovering the missing predicate. Consider the following example:

- (10) a. **The author** began (*{writing} / {reading, promoting, translating, using, ...}*) the book.
b. The author began (*?*{writing, promoting, ...} / {reading, translating, using, ...}*) the **borrowed** book.

Because of the lexical meaning of *author*, the sentence in (10a) is predominantly interpreted as ‘The author began *writing* the book’. However, the sentence is still ambiguous. The NP *the book* in (10a) can be interpreted as reading the book or other events, even though the subject of the sentence is *the author*. The NP *the borrowed book* in (10b) presupposes there exists the book. The presupposition implies that the author cannot write the book. The writer cannot promote or advertise the borrowed book on the purpose of selling it, either. Given that the linguistic context of *begin the book* narrows the range of candidates for the position of a missing predicate, its non-linguistic context, the utterance situation, finally disambiguates the sentence with *begin the book*. Consider the following examples:

- (11) a. In the situation in which the author has been writing a new book,
*The author began (**writing**) the book.*
b. In the situation in which the author should read some books to prepare for writing,
*The author began (**reading**) the book.*

In the situation as in (11), it is natural that the missing predicates are recovered as *writing* and *reading*, respectively.

To sum up, both linguistic and non-linguistic contexts contribute to disambiguation in the interpretation process of type coercion construction. A type-coerced construction such as *begin the book* can have various interpretations including *reading* and *writing*, depending on linguistic and non-linguistic context. The disambiguation process of the type coercion construction goes from the lexical meaning of words to non-linguistic context (situation) via linguistic context (composition). Nevertheless, we prefer interpretation such as *reading* or *writing* to others in the interpretation of *begin the book*. Why do we prefer some interpretation to the others? In the next section, we introduce “conventional non-linguistic context” to answer the question.

6 Conventional Non-linguistic Context

Usually, the most preferable interpretation of *begin the book* is to begin reading the book, with no special non-linguistic context. We argue that it is because there is a “conventional non-linguistic context” linked to an expression which is commonly assumed by most normal people in a language culture. In other words, the lexical meaning of a word is a conventionalized meaning which assumes a conventional non-linguistic context linked to the word (cf. [GDnd]). In addition, the context holds in the compositional process of the words. For example, the noun *book* is conventionally linked to the situation of reading it, because reading is the most common activity which people do with a book. That argument is proved by the frequency of predicates combining with *the book* collected from various corpora. The verb *read* has the highest frequency and the second is *write* in most corpora. If we have no information about the non-linguistic context and the linguistic context is neutral, we usually interpret the sentence *John began the book* as *John began reading the book*.

However, the interpretation under the assumption of “conventional non-linguistic context” is only the most probable one (cf. [PC2015]). The ambiguity of *begin the book* is not dissolved yet. The specific non-linguistic context of the utterance finally disambiguates the type coerced construction *begin the book*. We mention here that the Qualia Structure in the GL is attractive in that it gives insight about the preferred interpretation of a type coercion construction by postulating telic and agentive qualia, in spite of the risk that qualia can be extended too much. The conventional activity of normal people with a book is closely related to the original function of the artifact and how to create it.

7 Conclusion

In this paper, we tried to rethink the condition which allows aspectual coercion of *begin* and the interpretation of the type-coerced construction (*begin the book*), introducing Asher’s type-theoretical point-of-view and the “conventional non-linguistic context” based on Conventional Semantics ([GDnd]). The lexical semantic type of *begin* triggers some constraints on type coercion. In addition, we distinguished linguistic and non-linguistic context. Basically, the constraints

of aspectual coercion of *begin* are dependent on its lexical meaning. Regarding the interpretation of the type-coerced construction *begin the book*, we summarize our argument as follows:

I. lexical meaning

The semantic type and selection restriction of *begin* and *book* under the conventional non-linguistic context linked to the words

II. compositional meaning

The set of possible interpretations of the type-coerced construction *begin the book* with support of linguistic context

cf. *begin reading the book* is “preferred” as the interpretation of *begin the book* by its conventional non-linguistic context.

III. contextual meaning

The non-linguistic context finally dissolves the ambiguity of the type coerced construction *begin the book*.

A type-coerced construction is basically ambiguous until the non-linguistic context disambiguates it. We showed the three steps of interpreting a type-coerced construction above. Without any specific linguistic or non-linguistic context, conventionally-assumed situation linked to words works for disambiguation. Reading interpretation is the most probable one for *begin the book*, although it is still ambiguous.

There remain many important and interesting issues related with semantic type of words and type coercion. First, we need to define the concept of “conventional non-linguistic context” more strictly, considering its relation to the lexical meaning of words. Second, we need to decide what a dot-type is theoretically and propose the semantic type of *book* specific enough to explain its linguistic behavior (cf. [AN2011]). In the future research, we will explore the semantic type of words, including *book* and *begin*, and their semantic relation. In addition, we will develop the formalism of representing the interpretation mechanism which puts together lexical meaning and linguistic and non-linguistic context.

References

- [AN2011] Asher, N.: *Lexical Meaning in Context: A Web of Words*. Cambridge: Cambridge University Press. (2011)
- [GDnd] Gutzmann, D.: *Semantics vs. Pragmatics* L. Matthewson, C. Meier, H. Rullman, and E. Zimmerman, *The Semantic Companion*. (n.d.)
- [IS2013] Im, Seohyun: *The Generator of the Event Structure Lexicon (GESL): Automatic Annotation of Event Structure for Textual Inference Tasks*. PhD Dissertation. Brandeis University. (2013)
- [PC2015] Potts, C., Lassiter, D., Levy, R. and Frank, M. C.: *Embedded Implicatures as Pragmatic Inferences under Compositional Lexical Uncertainty*. Manuscript. (2015)
- [PJ1995] Pustejovsky, J.: *The Generative Lexicon*. The MIT Press. (1995)
- [PJms] Pustejovsky, J.: *Lexical Semantics*. Manuscript.

[PJPB1996] Pustejovsky, J. and P. Bouillon: Aspectual Coercion and Logical Polysemy. Pustejovsky, J. and B. Bogurave (eds.) *Lexical Semantics: The Problem of Polysemy*. Clarendon Press: Oxford. (1996)