Sept. 20, 2021 (v1)

Hilbert's Epsilon Operator in Linguistic Theory: A Metatheoretical Comparison of Two Applications

Abstract. Hilbert's ε-operator is a valuable tool from the foundations of mathematics. This paper compares two representative linguistic applications thereof, one by the Konstanz School on (in)definite NP semantics and the other by Chomsky on coordination syntax. The goal of this paper is twofold. First, it does a metatheoretical comparison of the two applications and thereby examines the ε-operator's status in linguistic theory. Second, it formally implements Chomsky's ε-based theory Form Sequence. There are three main similarities and two main differences between the two applications. The similarities include their semideterministic usage of the ε-operator, the ancillary role of the ε-operator in the language faculty, and their overlapping in a particular ε-based component. Their differences concern the level where the ε-operator takes effect and factors that make it context-sensitive. The implementation of Form Sequence is in a distributed fashion, partly in syntax and partly in discourse. This study reveals that Chomsky's application of the ε-operator has a broader relevance in linguistic theory beyond coordination and suggests that the ε-mechanism could be a third-factor principle of information processing.

Keywords. Hilbert's epsilon operator, formal syntax, metatheory, Form Sequence, third factor

1. Introduction

Hilbert's epsilon operator or ε-operator, named after the German mathematician David Hilbert, was first introduced in Hilbert (1926) and later established as a fundamental symbol in *Grundlagen der Mathematik* (Hilbert & Bernays 1939).[1] In short, an ε-operator is a constant that forms a term out of a formula, as in (1).[2]

(1) $\varepsilon x.\,\mathrm{F}(x)$

[1]This classic is only available in German till this day, but see Leisenring (1969) for a comprehensive and accessible introduction of the subject matter in English.

[2]There are various alternative notations, such as $\varepsilon x\mathrm{F}x$, $\varepsilon x\mathrm{F}$, and $\varepsilon_x\mathrm{F}(x)$.

Here x is an individual variable, F is a first-order predicate, and the entire string is an ε-term. Roughly speaking, a term corresponds to a noun, and a formula corresponds to a sentence.[3] Therefore, (1) roughly means "an individual x such that F is true for x." That is, the ε-term picks out a particular individual that satisfies F. For example, if F is *apple*, then $\varepsilon x.\mathrm{F}(x)$ is a particular apple. In Hilbert's original conception, ε-terms are nondeterministic, so there is no way to know precisely which individual (e.g., which apple) gets picked by the ε-term $\varepsilon x.\,\mathrm{apple}(x)$.

Hilbert's ε-operator has been applied to linguistics in two areas, one in semantics on (in)definite NPs and intersentential anaphora (by the "Konstanz School" in Winter's 1997 words, most representatively by Klaus von Heusinger) and the other in syntax on the derivation of coordination, especially the unbounded unstructured case thereof (by Noam Chomsky). The former application has a longer history (since the 1980s) while the latter is much more recent (since 2019). The first goal of this paper, as the title suggests, is to compare these two applications on a metatheoretical level and thereby examine the status of the ε-operator tool in linguistic theory—in particular, how it fits into the general architecture of contemporary generative grammar. Since the Konstanz School's use of ε is relatively well established in the literature (see, e.g., Egli & von Heusinger 1995 and von Heusinger 1997a,b, 2000, 2002, 2004, 2013), I will give more space to Chomsky's application.

As I will show, the ε-operator is used as a quite powerful theoretical tool in both applications, which makes a general examination of it highly worthwhile. Most formal tools employed in generative linguistics are module-specific. For instance, the λ-calculus is specific to semantics, and derivation trees are specific to syntax. There are familiar exceptions, such as features, which are used throughout all core linguistic areas. But then an interesting question is whether the status of ε is more like that of features or more like that of trees/λs. If we only consider the Konstanz School's application, we would naturally identify ε as a purely semantic tool on a par with its kin ι and η.[4] However, faced with Chomsky's application, which cannot be given a

[3] Thus, ε is more exactly a *subnector* in the sense of Curry (1963:32–33). In Curry's system, operators map nouns to nouns, while subnectors map sentences to nouns. I will keep calling ε an "operator" since that is the more common term in the literature.

[4] In mathematical logic, ι is used for definite descriptions (e.g., $\iota x.\,\mathrm{apple}(x)$ for *the apple*), and η is sometimes used for indefinite descriptions (e.g., $\eta x.\,\mathrm{apple}(x)$ for *an apple*). Part of Hilbert's idea is that ε alone is enough and that it can replace ι and η. See §2–3 for more detail.

straightforward formal semantic characterization, the above first impression must be revisited.

In a nutshell, my conclusion will be that the ε-operator has mixed status in linguistic theory. On the one hand, it is a task-specific tool because it has a highly specific application scenario, basically when a metalinguistic choice needs to be made. On the other hand, it is not straightforwardly applied at the "material" level of natural language analysis in either syntax or semantics, for it is not explicitly used or strictly needed in phrase-structure derivation or model-theoretic composition. Rather, its role is more like an ancillary tool generally available in the background, hence its cross-modular presence. This is reminiscent of the "third factors" in recent development of the minimalist program, which are identified as "principles not specific to the faculty of language" (Chomsky 2005:6), including principles of data analysis, principles of efficient computation, etc. I will suggest that the ε-operator—or at least the abstract idea it represents (basically the Axiom of Choice)—could be a third-factor principle for information processing that underlies both the design of language and that of mind.

In the remainder of this introduction, I briefly outline the two linguistic applications of the ε-operator mentioned above, which will then be examined in more detail in subsequent sections. The earlier, semantic application by the Konstanz School concerns (in)definite NPs and intersentential anaphora, as exemplified in (2).

(2) a. (In)definite descriptions: the man, a man, ...

b. Intersentential anaphora: A man comes. The man / He smokes.

Dissatisfied with previous approaches, such as the ι-operator approach to *the*, the quantificational approach to *a*, and the E-type pronoun approach to intersentential anaphora, von Heusinger proposed a unified approach to all three phenomena based on an extended theory of the ε-operator (with contextual indexing). On this approach, the utterances in (2) have the logical forms in (3). (3a–b) are based on von Heusinger (2000:254), and (3c) is adapted from von Heusinger (2013:370).[5]

[5] In his 2013 paper, which has a slightly different theoretical context, von Heusinger uses choice functions instead of ε-operators. Since these are just two sides of the same coin (see §2), and my focus here is on ε, I have adjusted von Heusinger's formula accordingly. I have also made some typographical adjustments for expository clarity.

(3) a. $\varepsilon_c x.\,\mathrm{man}(x)$ (a man chosen by a contextually determined, global ε-operator)

b. $\varepsilon_1 x.\,\mathrm{man}(x)$ (a man chosen by a free, local ε-operator)

c. $\mathrm{comes}(\varepsilon_1 x.\,\mathrm{man}(x)) \wedge \mathrm{smokes}(\varepsilon_j x.\left\{\begin{matrix}\mathrm{man} \\ \lambda y.\, y = y\end{matrix}\right\}(x))$, where $\varepsilon_j = \varepsilon_{i\langle [\![\mathrm{Man}]\!]/d, D/d\rangle}$
(The initial context is i. A freely chosen man d comes, and this choice updates the context to j, where *the man* or *he* ends up referring to d.)

I will explain the technicalities in (3) in Section 3. For now let us focus on the general idea. Basically, von Heusinger uses the ε-operator to pick out NP referents. Recall from the beginning of this section that in Hilbert's original conception the ε-operator is nondeterministic. By comparison, when indexing his global ε with a context parameter, von Heusinger grants it a certain level of determinacy. Thus, the ε-choice in his theory is not arbitrarily made but guided by discourse information (more exactly by a salience ranking of referents). This is a major difference between the linguistic usage of ε and its original mathematical usage, and as I will show, the same is true in Chomsky's application. This cross-disciplinary difference is inevitable given the context-sensitive nature of human language.

Compared to von Heusinger's application, Chomsky's application of the ε-operator is much more recent. Until now, it has only been explicitly mentioned in a lecture in 2019 and implicitly used in another lecture in 2020. A few authors have partly cited Chomsky's new idea (e.g., Sigurðsson 2020, Ott 2021, and especially Blümel 2020), but to my knowledge there is not yet any in-depth discussion of it in the literature. Thus, the results in this paper can hopefully lay some groundwork for future research in this direction. Chomsky's (2019) idea is basically that the ε-operator could be part of an extended theory of Pair Merge (Chomsky 2000) designed to derive "unbounded unstructured coordination." The phenomenon is exemplified in (4).[6]

(4) a. I met someone young, happy, eager to go to college, tired of wasting his time, . . .

b. The guy is young, tall, happy, young, eager to go to Harvard, . . . (Chomsky 2019)

Chomsky draws a parallelism between coordination and adjunction, proposing that each coordinated item in (4) is "independently adjoined to the host" and thereby "individually predicated

[6] These are Chomsky's original examples. In particular, he gives (4b), which has two occurrences of *young*, to demontrate that coordinated items in a sequence may repeat.

of what it links to." Chomsky lets each predicate or adjunct S_i pair-merge with a link element L_i[7] and places all the $\langle \mathrm{S}_i, \mathrm{L}_i \rangle$ pairs in a sequence whose first slot is occupied by a conjunction, as in (5).

(5) $\langle \mathrm{CONJ}, \langle \mathrm{S}_1, \mathrm{L}_1 \rangle, \ldots, \langle \mathrm{S}_n, \mathrm{L}_n \rangle \rangle$ (Chomsky 2019)

With some slight modification (see §4), Chomsky (2020) names the operation that gives rise to (5) Form Sequence. A familiar way to pin down such a sequence (out of a set of alternatives), as Chomsky (2019) suggests, is via Hilbert's ε-operator.[8] Note that while the sequence involved in an unbounded unstructured coordination is unstructured from a phrase-structural perspective (i.e., there is no c-command relation between the conjuncts), it is structured from a mathematical perspective, for the sequence itself is a kind of mathematical structure. Moreover, this mathematical structure (i.e., the sequence-defining order relation) could be linguistically significant, as illustrated in (6).

(6) John and Bill saw Tom and Mary respectively. (Chomsky 2019)

There are two coordinate noun phrases in (6), and in the presence of *respectively* their internal orders obviously matter, even though on the Form Sequence approach neither *and*-phrase has a hierarchical configuration. Above I have mentioned that in von Heusinger's application the ε-operator is not entirely nondeterministic. Given cases like (6)—and presumably also cases where the sequence-internal order involves certain speaker agency, such as (7)—we must conclude that the ε-operator is not totally nondeterministic in Chomsky's application either.

(7) As for fruit, I like apples, bananas, oranges, and strawberries—in that order.

Many technical details are left out in Chomsky's lectures. The second goal of this paper is to provide a concrete implementation of Chomsky's idea. My implementation foregrounds three questions that any implementation of Form Sequence must address:

Q1. In which module does the ε-based choice of sequence take place?

[7] Chomsky identifies all links in a coordinate sequence as one and the same. See §4 for more detail.

[8] Chomsky's original words are "[t]here are formal ways of doing it . . . one is David Hilbert's ε operator." This suggests that Form Sequence does not necessarily rely on the ε-operator. However, since Chomsky does mention the ε-based method (and only mentions it), it is reasonable to count Form Sequence as a linguistic application of ε.

Q2. How exactly is ε associated with the Form Sequence operation?

Q3. What linguistic element(s) encode ε?

And in the context of this paper, a fourth question must also be addressed:

Q4. How do von Heusinger's and Chomsky's use of ε coexist in the same grammar?

This is a valid question because the relevant empirical phenomena (i.e., NPs, anaphora, and coordination) coexist in the same grammar. In addition, my implementation of Chomsky's idea highlights an interesting connection between this new component of the minimalist program and an existing idea in the literature (de Vries 2005) that has so far escaped mainstream attention.

This paper is organized as follows. In Section 2, I introduce the mathematical background of Hilbert's ε-operator, which will facilitate my subsequent discussion. In Section 3, I review the Konstanz School's application of ε and summarize its main features. In Section 4, I review Chomsky's application of ε in comparison with the Konstanz School's application and implement Form Sequence within current minimalist syntax. In Section 5, I summarize the main similarities and differences between the two applications and further show that in fact Chomsky's usage of ε also underlies the Konstanz School's usage of it. Section 6 concludes.

2. Mathematical background

The content in this section is mainly based on Avigad & Zach (2020), Chatzikyriakidis, Pasquali & Retoré (2017), Leisenring (1969), and Slater's entry[9] in the *Internet Encyclopedia of Philosophy*. Partly inspired by Russell's iota operator for definite descriptions (Whitehead & Russell 1910),[10] Hilbert proposed two generic element symbols in the 1920s—first τ (1923) and then ε (1926). See (8) for an informal unfolding of the three symbols.

(8) a. $\iota x.\mathrm{F}(x)$: the unique x that satisfies F

b. $\tau x.\mathrm{F}(x)$: an x that satisfies F when every individual does so

c. $\varepsilon x.\mathrm{F}(x)$: an x that satisfies F when some individual does so

[9] https://iep.utm.edu/ep-calc/ (retrieved in Aug 2020)

[10] The original shape of the iota symbol in *Principia Mathematica* is an inverted ι. But the upright ι is also common in the literature in both logic and linguistics. I use ι to be typographically consistent with the functionally similar ε, τ, and η. These are all subnectors à la Curry (see note 3).

Unlike ι, which basically says *the*, both τ and ε return generic elements, and in principle there is no way to know exactly which individual is chosen. For this reason, Hilbert's two operators are said to be nondeterministic (or indeterminate). Moreover, τ and ε are closely related to the two quantifiers $\forall$ and $\exists$ in predicate logic. Indeed, (8b) and (8c) are respectively a universal and an existential generic object with regard to F (Chatzikyriakidis et al. 2017), as in (9).

(9) a. $\mathrm{F}(\tau x.\mathrm{F}(x)) \equiv \forall x.\mathrm{F}(x)$

b. $\mathrm{F}(\varepsilon x.\mathrm{F}(x)) \equiv \exists x.\mathrm{F}(x)$

While τ and ε had started their lives as different symbols, they are in fact mutually definable (see, e.g., Retoré 2014 and Abrusci 2017), so in the end Hilbert only kept ε.

As pointed out in sources like Leisenring (1969) and Chatzikyriakidis et al. (2017), Hilbert's original motivation with the ε-operator was to find a consistent and complete axiom set for mathematics, first and foremost for arithmetic. The ε-operator was convenient for this purpose in that it eliminated the two quantifiers and could also replace the Axiom of Choice (see, e.g., Bourbaki 1970/1954 [in terms of τ] and Bernays 1991/1958 [in terms of ε]). Hilbert's program eventually failed due to Gödel's (1931) incompleteness theorems, according to which there is no consistent axiomatic system for arithmetic, and no system can prove its own consistency. Nevertheless, Hilbert's endeavor left us with a number of valuable results, including the ε-operator.

The ε-operator, as a logical symbol, should be equipped with an ambient syntax and a corresponding semantics. Its syntax is known as the ε-calculus, which is a minimal extension of predicate calculus, with ε being the only new symbol. The ε-calculus defines an ε-term for each and every predicate, and for each ε-term there is a corresponding axiom known as Axiom ε.

(10) Axiom ε: $\mathrm{F}(t) \rightarrow \mathrm{F}(\varepsilon x.\mathrm{F}(x))$

"If any term t has the property F at all, then $\varepsilon x.\mathrm{F}(x)$ has it."

What this axiom says is essentially that for any nonempty subset of the domain of discourse, we can choose a representative element from it, but that is basically the Axiom of Choice.[11] Indeed, Hilbert's ε-operator is also known as the choice operator.

Hilbert did not give ε any semantics at the time of its proposal but merely used it as a syntactic tool to facilitate proof construction. The first semantics for ε was proposed in Asser (1957), where it was interpreted by a choice function. Asser's model for the ε-calculus is defined as

(11) $\mathscr{M} := \langle \mathbf{J}, \mathbf{I} \rangle$ (based on Asser 1957:33–34)

where $\mathscr{M}$ is the model, $\mathbf{J}$ is its domain, and $\mathbf{I}$ is its constant-interpretation function. Asser sets $\mathbf{I}(\varepsilon)$ to be a choice function Φ.[12] Specifically, Φ is a function that chooses an arbitrary element from each subset of $\mathbf{J}$. Asser also took into consideration the empty set—namely, the case where the *if*-clause in Axiom ε is false. He suggested two possible solutions, one with a total choice function and the other with a partial one. On the total function solution, $\Phi(\emptyset)$ returns an arbitrary element ξ_0 of $\mathbf{J}$—namely, an arbitrarily chosen individual in the whole world—and on the partial function solution it is undefined. Thus, we have

(12) $$[\![\varepsilon x.\,\mathrm{F}(x)]\!] = \Phi([\![\mathrm{F}]\!]) = \Phi(A \subseteq \mathbf{J}) = \begin{cases} a \in A, & \text{if } A \neq \emptyset \\ \xi_0 \in \mathbf{J} \text{ or undefined}, & \text{if } A = \emptyset \end{cases}$$

As Leisenring (1969) points out, the total function solution suits Hilbert's original purpose better. This is also the sentiment in many later works,[13] including those on the philosophy of language. Among others, Slater (2017:278) explicates that if there is no such x that satisfies $\mathrm{F}(x)$, then the denotation of $\varepsilon x.\,\mathrm{F}(x)$ "is a fiction, which means it is simply a pragmatically chosen individual in the whole world at large." Slater illustrates this point with phrases like *that mouse in the room* and *that man with martini in his glass*. When no such mouses or men exist, these could just refer to "a shadow on the carpet" and "a man with water in his glass" instead (e.g., under illusion).

[11] In Jech's (1973: 1) formulation, the Axiom of Choice says "for every family $\mathscr{F}$ of nonempty sets, there is a function f such that $f(S) \in S$ for each set S in the family $\mathscr{F}$."

[12] In Asser's (1957: 33) words, the ε symbol is a variable for choice functions of the individual domain (*das Zeichen ε ... ist eine Variable für Auswahlfunktionen des Individuenbereiches* $\mathbf{J}$).

[13] Though see Muskens (1989) for a further development of the partial function solution.

In sum, Hilbert's ε-operator had originally been proposed as part of a program to completely axiomatize mathematics. Despite the unfortunate fate of Hilbert's program, ε has survived as a useful tool. The formal system it lives in is the ε-calculus, whose corresponding semantics is a model equipped with a choice function. With the above mathematical background, next I will examine the two linguistic applications of ε mentioned in Section 1 in more detail.

3. Linguistic application I: the Konstanz School

As mentioned in Section 1, Hilbert's ε-operator was first applied to linguistics in semantics, most representatively in a series of work by Klaus von Heusinger (built on earlier works like Slater 1986 and Egli 1991). Winter (1997:410, note 13) refers to this line of research as the Konstanz School. The relevant empirical phenomena are repeated in (13).

(13) a. (In)definite NPs: the man, a man, ... (=(2))

b. Intersentential anaphora: A man comes. The man / He smokes.

In this section, I will first briefly go through the conventional analyses of these phenomena, with a focus on their problems, which have motivated the Konstanz School's application of the ε-operator (§3.1). Then, I will lay out the key components of this application (§3.2). Finally, I will discuss its design features on a metatheoretical level, with a focus on how it differs from the original conception of ε in mathematics (§3.3).

3.1 Motivation

In formal semantics, the mainstream logical tools used to translate indefinite and definite NPs are the existential quantifier and the ι-operator, as in (14).

(14) a. A man walks.

$\exists x[\mathrm{man}(x) \wedge \mathrm{walks}(x)]$ (adapted from von Heusinger 2000:247)

"There exists an individual x such that x is a man and x walks."

b. The man comes.

$\mathrm{comes}(\iota x.\,\mathrm{man}(x))$ (adapted from von Heusinger 1997a:68)

"The unique x such that x is a man comes."

Konstanz School scholars were dissatisfied with these tools for two main reasons. First, they do not reflect the syntactic constituency of the natural language phrases or the intuition that NPs are referring expressions. Indeed, no part in the formula in (14a) precisely corresponds to *a man*, and if we unfold the ι symbol a bit, we can see that no part in it corresponds to *the man* either, as in (15).

(15) $\text{comes}(\iota x.\,\text{man}(x)) \equiv \exists x[\text{man}(x) \wedge \forall y[\text{man}(y) \rightarrow x = y] \wedge \text{comes}(x)]$

"There exists an individual x such that x is a man, for any individual y such that y is a man we have $x = y$, and x comes."

Since both the conventional treatment of indefinite NPs and that of definite NPs involve quantification, henceforth I will collectively refer to them as the quantificational approach. The broader issue here, in Retoré (2014:16) words, is that the generalized quantifier treatment of determiners "does not provide [them] with a proper logical form that can be interpreted on its own."

Second, the quantificational approach crucially relies on two presuppositions—the uniqueness presupposition (for *the*) and the existence presupposition (for both *the* and *a*)—which makes it "too restrictive to deal with natural language phenomena" (Egli & von Heusinger 1995:121). Consider the following examples.

(16) a. The town on Lake Constance is famous.

b. A wet nappy is not wet because no nappy is wet.

(Egli & von Heusinger 1995:124–125)

On a quantificational analysis, (16a) would presuppose that there exists one and only one town on Lake Constance, which is inconsistent with our world knowledge. Similarly, the first half of (16b), *a wet nappy is not wet*, would presuppose that there exists some nappy that is both wet and not wet, which is impossible. The inflexibility of the quantificational approach is also evident in the analysis of intersentential anaphoric pronouns (or E-type pronouns à la Evans 1977), such as (17).[14]

[14] There are still other influential approaches to intersentential anaphora (e.g., Kamp 1981, Heim 1982). See von Heusinger (1997a) for a critical overview. Since my purpose is not to investigate the empirical phenomenon per se but merely to contextualize the Konstanz School application of ε, I will not go into further detail.

(17) A man comes. He smokes.

$\exists x[\text{man}(x) \wedge \text{comes}(x)] \wedge \text{smokes}(\iota x[\text{man}(x) \wedge \text{comes}(x)])$

(adapted from von Heusinger 1997a:67)

Due to the presuppositions of ι, (17b) entails that there exists one and only one man who comes that smokes. While this could be true for the sentence above, it is more problematic for sentences like the following.

(18) a. A wine glass broke last night. It had been very expensive.

b. Every farmer who owns a donkey beats it. (von Heusinger 1997a:68)

A quantificational analysis of the two *it*s in (18) would strictly limit the number of existent wine glasses and that of the donkeys a farmer may own to one and only one, which is once again inconsistent with our world knowledge.

3.2 Application

Dissatisfied with the conventional semantic analyses of (in)definite NPs and intersentential anaphora, Konstanz School scholars proposed an alternative, ε-based analysis. In Egli & von Heusinger's (1995) words, the ε-operator is a "generalized iota operator without existence presupposition and uniqueness conditions" (p.129), and compared to conventional analyses, the ε-based analysis "is always very close to the surface of natural language expressions" (p.133). As such, the ε tool avoids both fundamental flaws in the conventional (i.e., quantificational) tools mentioned above, which is perfectly in line with Hilbert's intention to replace quantifiers with epsilons (see §2). In what follows, I will provide more details on how Konstanz School scholars approached the above-mentioned grammatical phenomena, starting with definite NPs.

A first point to note is that they did not stick to Hilbert's original theory but gave the ε-operator a linguistically oriented modification instead, mainly concerning how exactly the ε-choices were made. In the case of definite NPs, they added a certain level of determinacy to the originally nondeterministic choice operator by making it context-dependent—via Lewis's (1979) notion of salience. On this new conception, the descriptive material in a definite NP (e.g., *apple* in *the apple*) denotes a set as usual, but this set is furthermore equipped with a salience-based ranking of its members, which we can view as a total order.[15] Following Lewis, Konstanz School scholars viewed this total order as a hierarchy of salience and assumed that the specific ordering was determined by pragmatics or discourse. This context-dependence is formally encoded by an index. Thus, given a context c, the classical ε-term $\varepsilon x.\mathrm{F}(x)$ becomes $\varepsilon_c x.\mathrm{F}(x)$, which picks out the most salient element in $[\![\mathrm{F}]\!]$ under c. On the semantic side, ε_c is interpreted by an indexed choice function Φ_c. As such, there is "not *one single* choice function but *a whole family* of them indexed with situations" (Egli & von Heusinger 1995:134). This echoes Asser's (1957) assertion that the ε symbol is a variable for choice functions (see note 12).

Egli & von Heusinger use the sentence in (19a) for illustration. The relevant world knowledge is that there are ten islands on Lake Constance that are bigger than 2,000 m^2, among which the biggest three are Reichenau, Lindau, and Mainau.[16] Suppose the sentence is uttered in three different contexts: *i*) by a Reichenau fisherman, *ii*) by a Lindau tour guide, and *iii*) by the Earl who owns Mainau. Then *the island* would have three different extensions, as in (19b) (adapted from von Heusinger 2004:314–315).

(19) a. The island on Lake Constance is nice.

b. $[\![\text{the island}]\!] = [\![\varepsilon_c x.\,\text{island}(x)]\!] =$

$$\Phi_c([\![\text{island}]\!]) = \begin{cases} \Phi_c(\langle \text{Reichenau}, \ldots \rangle) = \text{Reichenau, if } c = i \\ \Phi_c(\langle \text{Lindau}, \ldots \rangle) = \text{Lindau, if } c = ii \\ \Phi_c(\langle \text{Mainau}, \ldots \rangle) = \text{Mainau, if } c = iii \end{cases}$$

According to Egli & von Heusinger (1995:133), *the island* denotes "the most salient, the most prominent, the most conspicuous island that can be talked about in a given situation."

As for indefinite NPs, Konstanz School scholars proposed a freely introduced, locally effective ε-operator to analyze them, which is also indexed, though not by the context (so this index is merely distinctive). I repeat the exemplary ε-term from Section 1 below.

[15] In mathematical order theory, an ordered set is a *total order* (aka *chain* or *linear order*) iff every pair of elements in it is part of the order relation. As Schröder (2016:23) remarks, "[w]hen people talk about ranking objects, they typically are talking about a chain."

[16] Source: `https://en.wikipedia.org/wiki/Lake_Constance#Islands` (retrieved in Aug 2021)

(20) $\varepsilon_1 x.\,\mathrm{man}(x)$ (=(3b))

There are two key differences between the global and the local ε-operator, both concerning the nature of the choice functions they denote. First, a global choice function is defined for all predicates (i.e., all subsets of the domain of discourse), whereas a local one is only defined for a single predicate (i.e., a single subset)—the one that is introduced by the local ε-term (von Heusinger 2004:316–317). To illustrate, the choice function that interprets *the man* not only picks out a man but also picks out a dog, a table, etc. By comparison, the choice function interpreting *a man* only picks out a man.

Second, a global choice function is contextually determined, which means that all the choices it makes are the most salient elements in the contextually imposed salience hierarchies. Meanwhile, a local choice function is not contextually bound but only existentially closed at the text level. Therefore, the single choice it makes is not determined by any context-induced salience hierarchy. In other words, the index in a local ε-term is merely distinctive and imposes no structure on the extension of its descriptive predicate. Thus, (20) denotes an arbitrary or unknown element in the set $[\![\mathrm{man}]\!]$, which is exactly what *a man* intuitively means. Below is a more formal representation of the above differences (adapted from von Heusinger 2004:317).[17]

(21) a. $[\![\text{the F}]\!] = [\![\varepsilon_c x.\,\mathrm{F}(x)]\!] = \Phi_c([\![\mathrm{F}]\!]^c)$

(where $c \in \mathbb{C}$ [a set of contexts] and $[\![\mathrm{F}]\!]^c$ is an ordered set)

b. $[\![\text{an F}]\!] = [\![\varepsilon_n x.\,\mathrm{F}(x)]\!] = \phi_n([\![\mathrm{F}]\!])$

(where $n \in \mathbb{N}$ [the set of natural numbers] and $[\![F]\!]$ is a plain set)

Although a local choice function is only defined for a single predicate, it simultaneously updates the existing global context—by promoting the arbitrarily chosen element to be the most salient element in its local set—and by assumption also in its supersets (von Heusinger

[17]For expository clarity, I use Φ_c for global, model-level choice functions and ϕ_n for local, predicate-level choice functions. As for the indices, von Heusinger's symbolism is not always consistent, especially for local choice functions, for which he uses i, j, etc. in von Heusinger (2000, 2004), x, y, etc. in von Heusinger (2013), and η terms (instead of ε ones) in von Heusinger (1997a,b), presumably following an older tradition in the pre-ε era (see note 4). Sometimes he further distinguishes syntactic and semantic indices (e.g., i, j, etc. for ε and random natural numbers for Φ in von Heusinger 1997b). I use i, j, etc. to index global choice functions and use $1, 2$, etc. to index local ones.

2013:370)). This reflects the intuition that after an indefinite NP has been mentioned, its referent can later be referred to by a definite NP or a pronoun. See (22) for an illustration.

(22) A man comes. The man / He smokes. (=(2b)/(3b))

$$\mathrm{comes}(\varepsilon_1 x.\,\mathrm{man}(x)) \wedge \mathrm{smokes}(\varepsilon_j x.\begin{Bmatrix}\mathrm{man} \\ \lambda y.\, y = y\end{Bmatrix}(x)),\ \text{where } \varepsilon_j = \varepsilon_{i\langle [\![\mathrm{Man}]\!]/d, D/d\rangle}$$

Suppose the initial context is i, which determines a global ε_i. After $\varepsilon_1 x.\,\mathrm{man}(x)$ picks out an arbitrary man—call him d—that man becomes the most salient individual in $[\![\mathrm{man}]\!]$ and thereby creates a different context j, where the value of $[\![\varepsilon_i x.\,\mathrm{man}(x)]\!]$ is updated to that of $[\![\varepsilon_1 x.\,\mathrm{man}(x)]\!]$ (von Heusinger uses a slash notation to indicate this update). This update is in theory the only difference between contexts i and j, but in order to handle anaphoric pronouns von Heusinger exploited the salience-changing power of the local ε a bit further and let it update the salience-based choice of certain supersets as well[18]—all the way up to the entire domain D (see also von Heusinger 1997a:79). Thus, when $\varepsilon_1 x.\,\mathrm{man}(x)$ promotes a man to the top of the salience hierarchy in $[\![\mathrm{man}]\!]$, it also promotes him to the top in $[\![\text{male objects}]\!]$ and $[\![\text{objects}]\!]$. Probably for syntax/semantics distinction purposes, von Heusinger uses an identity predicate $\lambda y.\, y = y$ instead of D in the maximally general choice operator $\varepsilon_j x.\,[\lambda y.\, y = y](x)$.

3.3 Discussion

Above I introduced the key components in the Konstanz School's linguistic application of Hilbert's ε-operator—namely, a contextually indexed global ε-operator, a contextually free (yet still indexed) local ε-operator, the salience-ranking power of the context, and the context-updating power of the local ε-operator. Next I will make a few metatheoretical remarks on these components, with a focus on how they make the Konstanz School's ε-theory differ from the original mathematical theory.

First, choice functions are taken to be partial in the Konstanz School's application, presumably to facilitate salience-modeling (see von Heusinger 2004:327, notes 2–3). This is clearly different from the situation in mathematics, where choice functions are more naturally defined

[18] In fact, he further expands the notion of salience-change potential to cover definite NPs (see von Heusinger 2004:322), letting them carry out trivial context updates where the pre- and post-update global choice function are the same. The purpose of von Heusinger (2004) is to define dynamic semantics in a formally unified way, but due to my more restricted scope I abstract away from this aspect of von Heusinger's theory.

as total functions (see §2). That said, Konstanz School scholars' position on the empty set is not always consistent. For instance, contrary to the sentiment in von Heusinger (2004), Egli & von Heusinger (1995:132–133) state that "the definition of the epsilon operator for empty sets allows us to analyze sentences that could not be represented within more classical formats," such as nonexistent or impossible objects. Clearly, the decision between a total function interpretation of the ε-operator and a partial function one is an empirical issue in linguistics.

Second, as Egli & von Heusinger (1995:134) point out, the global ε-operator does two jobs at once—it ranks the elements in $[\![\text{F}]\!]$ and chooses the most salient element from it. This also deviates from Hilbert's original conception, where ε is no more and no less than a choice operator. The deviation is easy to eliminate, though. We just need to induce a division of labor by letting the context c impose the salience ranking and letting ε perform the choice.

Third, although Egli & von Heusinger define the salience ranking as a total order, the totality is not really needed in their theory. To interpret *the F*, we only need to know what the most salient element of $[\![\text{F}]\!]$ is, and the rest elements are simply collectively less salient. Thus, to interpret *the island* in (19a) there is no need to also specify what the second, third, . . . , *n*th most salient island is—and there is probably no way to do so either, for when a fisherman/guide/Earl utters (19a), they most likely do not care about how the remaining ten-ish islands on Lake Constance are ranked against each other. One potential way to adjust Egli & von Heusinger's theory to better reflect this aspect of the discourse is to allow some flexibility in the salience ranking. Depending on what the context is, the ranking could be a full-fledged total order, a simple bipartition (i.e., most salient vs. everything else), or anything in between. Mathematically, all these scenarios fall under the definition of a *partial order*, where certain elements can stay mutually incomparable. Thus, in the scenario of a bipartition, only the most salient island is ordered with regard to all other islands on Lake Constance, and in the in-between scenarios, only those contextually relevant islands are in the order relation. I illustrate these with the toy examples in (23).

(23) Islands on Lake Constance: Reichenau, Lindau, Mainau, Werd, and Hoy

a. $[\![\text{island}]\!]^i = \langle [\![\text{island}]\!], >_i \rangle = \{(\text{R},\text{L}), (\text{R},\text{M}), (\text{R},\text{W}), (\text{R},\text{H})\}_>$

(where the order relation $>$ is more-salient-than; same below)

b. $[\![\text{island}]\!]^{iv} = \langle[\![\text{island}]\!], >_{iv} z\rangle = \{(\text{R},\text{L}), (\text{R},\text{M}), (\text{R},\text{W}), (\text{R},\text{H}), (\text{L},\text{M}), (\text{L},\text{W}), (\text{L},\text{H})\}_{>}$ (context *iv* is when (19a) is uttered by a tour guide who is in charge of R and L and is currently talking about R)

As in (23), when the speaker does not care about the ranking between two elements (e.g., Werd and Hoy), they, as a pair, do not need to be in the order relation imposed by the context c (though they are still in the plain set characterizing $[\![\text{island}]\!]$). The ε tool works fine as long as *one* element is ranked above everything else. The takeaway here is that when applying the ε-operator in natural language analysis in the way the Konstanz School does—namely, with a context parameter—there can be much complication and subtlety in how the context influences the ε-choice, but that is a separate thing from the functionality of ε itself, which is (and should be) still just a choice operator. Teasing apart the context-free and the context-sensitive components in the formalism not only makes Egli & von Heusinger's particular analysis more accurate but also makes it easier for us to compare different domain-specific applications of the ε-operator. I will return to this point when reviewing Chomsky application in Section 4.

Fourth, recall from Section 2 that in Hilbert's original conception the ε-operator was a purely syntactic symbol. Similarly, the Konstanz School's explicit use of the ε-operator has also helped them keep syntax and semantics separate, which sets their analysis of (in)definite NPs apart from other analyses that also use choice functions but do not use the ε-operator (e.g., Winter 1997, Kratzer 1998). As von Heusinger (2002:266, note 13) points out, he uses the ε-operator "as the syntactic representation of the indefinite article, while the choice function is the corresponding semantic function." This explicit separation of syntax and semantics is significant for our discussion, because recall from Section 1 that one of the main concerns of this paper is precisely about the grammatical module or representation level of ε. Note that von Heusinger's meticulous separation of syntax and semantics is also manifested in his adoption of different indices for ε-operators and choice functions (see note 17) and his adoption of $[\lambda y.\, y = y](x)$ instead of D as the generic predicate in (22).

Fifth, and as a further elaboration of the fourth point, while von Heusinger treats ε as part of syntax rather than part of semantics, the "syntax" here does not refer to natural language grammar but refers to the mediating logical language (i.e., the ε-calculus) between natural language and its semantic model. This mediating role suggests that the ε-operator is not really indispensable in the analysis of natural language grammar. Indeed, von Heusinger (2013) presents the same idea from his earlier works in an ε-free fashion, solely in terms of choice functions. In a word, even though the Konstanz School emphasizes the syntactic status of the ε-operator, it is still more of an ancillary analytical tool rather than an integral part of the syntax-semantics interface of human language (the same is true for the λ-calculus). In Section 4, I will show that Chomsky's application of the ε-operator faces a somewhat similar situation.

4. Linguistic application II: Chomsky

Compared to the Konstanz School's application of the ε-operator, Chomsky's application is not only much more recent but also more open-ended. In fact, it is only mentioned in passing in a 2019 lecture. However, since the operation it conceptually supports—Form Sequence—is a key new addition to minimalist syntax, I deem it worthwhile to examine the role of ε in Chomsky's new theory with some meticulosity. In this section, I will first review Chomsky's new theory (§4.1), then make some metatheoretical remarks on Chomsky's application of the ε-operator in comparison with the Konstanz School's application (§4.2), and finally provide a concrete implementation of Form Sequence within current syntactic theory (§4.3). Due to lack of relevant information in the literature, much of the content in this section is based on my own understanding, which I hope can serve as a point of departure for future research.

4.1 Form Sequence with ε

Chomsky's application of the ε-operator is part of an extended theory of Pair Merge. The new theory, called Form Sequence,[19] is designed to generate unbounded unstructured coordination. The examples below are from Chomsky (2019).

(24) a. I met someone young, happy, eager to go to college, tired of wasting his time, . . .

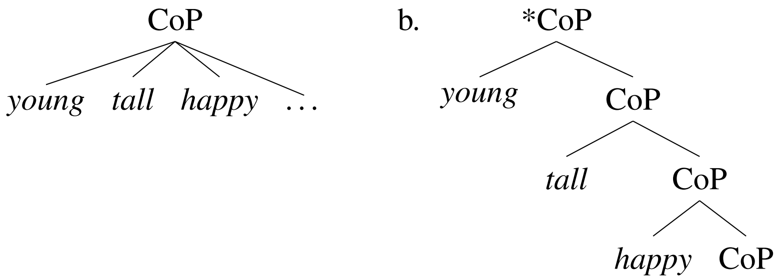
b. The guy is young, tall, happy, young, eager to go to Harvard, . . .

c. John, Bill, Tom, the young man. . . read the book, walked to the store, . . .

[19] This term first appears in Chomsky (2020).

The sequences in (24) can all go on ad infinitum. In fact, Chomsky suggests that the simple phrases we normally see—such as *John saw Bill* and *John ran*—are "just limiting cases of sequences" (Chomsky 2020). Thus, the notion sequence plays an important role in Chomsky's new theory, perhaps even in the entire minimalist system, as Chomsky (2020) generalizes it to the extent that "wherever there is an XP, there would be a sequence." The coordinate phrases in (24) are unstructured in the sense that none of the conjuncts is in the scope of any other (i.e., there is no asymmetrical c-command). So, their structural relationship is in a sense "flat" rather than (strictly) configurational, as illustrated in (25).[20]

(25) a. b.



In Chomsky's (2019) words, each conjunct in an unbounded unstructured coordination is "independently adjoined to the host" and "individually predicated of what it links to." So, Chomsky is making a connection between coordination and adjunction, and the connection is not merely due to the fact that the motivating example he uses happens to be an adjectival phrase, because he clearly endows the host component with a more abstract theoretical status. Thus, although neither of the two sequences in (24c) has an overt item to cling to, they still both have a host—or more exactly a "link"—in the underlying syntax, which Chomsky takes to be the phasal heads *n* and *v*.[21]

In post-2000 minimalism, adjunction is derived by Pair Merge, which takes two syntactic

[20]Here I merely use the label CoP for expository convenience, but in §4.3 I will give it a more substantive place in my implementation. As in §3, here too I will not aim to review literature on the empirical phenomenon (i.e., coordination) but just present enough background to contextualize Chomsky's application of the ε-operator.

[21]There is some inconsistency between Chomsky's 2019 and 2020 conceptions of *n*/*v*. First, while the starred n^*/v^* and the plain *n*/*v* are both treated as phase heads in Chomsky (2019), only the starred ones are treated so in Chomsky (2020). Second, while the phase-marking *n*/*v* and the categorizing *n*/*v* (from distributed morphology) are treated as "totally different notions" in Chomsky (2019), they are identified as the same heads in Chomsky (2020). Building on this identification, Chomsky (2020) furthermore takes "the v/v^* distinction" (and presumably also the n/n^* distinction) to be "eliminable" and attributes their different phase-creating abilities to "the lexical content of the roots." Such inconsistency brings about much theoretical uncertainty. For current purposes, I assume that *n*/*v*, either starred or not, can serve as links in coordination. Perhaps the star is just an ancillary notation and has no virtual-conceptual substance in the grammar.

objects α, β as input and returns an ordered pair $\langle\alpha,\beta\rangle$ as output, with α being the adjunct and β being the host.[22] Pair Merge constitutes a different kind of Merge than Set Merge, as in (26).

(26) a. Set Merge$(\alpha,\beta) = \{\alpha,\beta\} = \{\beta,\alpha\}$

b. Pair Merge$(\alpha,\beta) = \langle\alpha,\beta\rangle \neq \langle\beta,\alpha\rangle$

Set elements are unordered, whereas pair components are ordered. Chomsky (2004:117–118) further likens adjunction to multidimensional structure building, suggesting that "we might intuitively think of α as attached to β on a separate plane, with β retaining all its properties on the 'primary plane,' the simple structure." This multidimensionality is made even clearer in Chomsky (2019):

> The unbounded unstructured cases show [that] there are unboundedly many dimensions to what's going on up there [in our minds]. [It's] not two-dimensional like a blackboard. You can add any number of adjuncts at any point.

I illustrate this dimension-expanding capacity of Pair Merge in (27), where n adjuncts are attached to a single host.[23]

(27) Pair Merge$(\alpha_1,\beta) = \langle\alpha_1,\beta\rangle$

Pair Merge$(\alpha_2,\beta) = \langle\alpha_2,\beta\rangle$

...

Pair Merge$(\alpha_n,\beta) = \langle\alpha_n,\beta\rangle$

Chomsky assumes that each conjunct-link unit constitutes such a pair and that all links within the same CoP are identifiable, because "[we] are attaching everything to the same point" (Chomsky 2019). In the foregoing context, this means that the link element is just the (abstract) host (e.g., β in (27)). But I will slightly deviate from this in my implementation in Section 4.3 and make a distinction between the notions link, host, and attachment point. Specifically, I

[22] The identification of the first component of $\langle\alpha,\beta\rangle$ as the adjunct and the second component as the host seems to be a stipulation, which is fine as long as the designation is consistent.

[23] Chomsky's shift from plane-talk to dimension-talk may have more substantive consequences than just a change of terminology. That is because an n-plane space is usually not the same thing as an n-dimensional space. For instance, a 1-plane space is 2-dimensional, and a 3-dimensional space may have an infinite number of planes. The purely formal notation in (27) may be interpreted as merely multiplanary or fully multidimensional.

will reserve the term *link* to the identifiable elements within the conjuncts (e.g., *n*/*v*), stick to the more conventional usage of *host* to refer to the actual second component in a pair-merged product (i.e., β in $\langle\alpha,\beta\rangle$), and use *attachment point* to refer to the structural point (i.e., tree node) occupied by the entire CoP from the perspective of the main structure (e.g., the complement/specifier position of *v*P).

Moreover, Chomsky assumes that the structure of CoP itself is an ordered tuple—namely, a sequence—rather than a set, hence the term Form Sequence. I will deviate from this in my implementation too, contending that the sequence structure is only available in the derivational environment/background but not in narrow-syntactic derivation (i.e., tree-building) per se. That is, I take the following structures from Chomsky (2019, 2020) to be high-level declarations rather than concrete syntactic objects.

(28) a. $\langle \text{CONJ}, \langle \text{S}_1, \text{L}_1 \rangle, \ldots, \langle \text{S}_n, \text{L}_n \rangle \rangle$ (Chomsky 2019)

(where CONJ is a conjunction, $\text{S}_{i\in\mathbb{N}}$ are conjuncts, and $\text{L}_{i\in\mathbb{N}}$ are links)

b. $\langle (\&), \text{X}_1, \ldots, \text{X}_n \rangle$ (Chomsky 2020)

(where & is an optional conjunction and each X_i is a conjunct)

The link component is neither explicitly used nor mentioned in Chomsky (2020), but since it plays a vital role in Chomsky (2019), being what holds the multidimensional object together and what connects it to the main derivation tree, I take the liberty to assume that it still implicitly exists in Chomsky (2020). Nevertheless, in my implementation I will not treat the pair $\langle \text{S}_i, \text{L}_i \rangle$ in the above notation as a product of Pair Merge, because the link element in Chomsky's conception (e.g., the phase-marking *n*/*v*) is unlikely to be pair-merged with its sister node in current minimalism. This is well reflected in the following example taken from Chomsky (2020).

(29) a. John arrived and met Bill.

b. {C, {John_3, {INFL, $\underline{\langle \&, \{_1\ \{_2\ v, \{\text{arrive John}_1\}\}\}, \{_3\ \text{John}_2, \{v^*, \{\text{meet B}\}\}\}\rangle}$}}}

The underlined coordinate sequence in (29b) contains no conjunct-link *pairs*—the $\langle \text{S}_i, \text{L}_i \rangle$ slots in (28a) are filled by ordinary sets instead. To reconcile this discrepancy, I treat the pair notation in (28a) not as a pair-merged syntactic object but as a high-level declaration specifying that a link element can be identified for each conjunct. In formal terms, this amounts to defining a function $\lambda \mathrm{S}_i.\mathrm{L}_i$ that assigns to each conjunct term one of its subterms, which in set talk is exactly a set of pairs $\{\langle \mathrm{S}_1, \mathrm{L}_1\rangle, \langle \mathrm{S}_2, \mathrm{L}_2\rangle, \dots, \langle \mathrm{S}_n, \mathrm{L}_n\rangle\}$. To avoid confusion with Pair Merge—which, as I will contend, plays a separate role in Form Sequence—and to reduce clutter, hereafter I write $\langle \mathrm{S}_i, \mathrm{L}_i\rangle$ as $\mathrm{S}_i^{\mathrm{L}_i}$.

Note that the Form Sequence procedure is highly compact, with a number of intermediate stages. As Chomsky (2019) notes:

> [I]n order to generate these objects [i.e., the CoPs], you generate a finite set, and then you form from that set a sequence. It could be any sequence of elements, and there's in fact infinitely many possible sequences. You pick one out of those, and that sequence, call it S, is the thing that you are then going to merge into the construction. This operation of picking a particular element out of the set of sequences is David Hilbert's epsilon operator, which picks a single thing out of a set. It was part of his work on the foundations of mathematics—[a] basic operation. It's a straightforward operation, but it does have the property of being indeterminate.

This is the only place in Chomsky's two lectures where the ε-operator is explicitly mentioned. Nevertheless, it clearly shows that the ε-operator plays a fundamental role in Form Sequence. See (30) for example.

(30) young, tall, and happy

a. $\{\text{young}^{\mathrm{L}}, \text{tall}^{\mathrm{L}}, \text{happy}^{\mathrm{L}}\}$ (a set of conjuncts with links)

b. $\{\langle \&, \text{young}^{\mathrm{L}}, \text{tall}^{\mathrm{L}}, \text{happy}^{\mathrm{L}}\rangle,$ (a set of potential sequences)
$\langle \&, \text{tall}^{\mathrm{L}}, \text{young}^{\mathrm{L}}, \text{happy}^{\mathrm{L}}\rangle, \langle \&, \text{happy}^{\mathrm{L}}, \text{tall}^{\mathrm{L}}, \text{young}^{\mathrm{L}}\rangle, \dots\}$

c. $\langle \&, \text{young}^{\mathrm{L}}, \text{tall}^{\mathrm{L}}, \text{happy}^{\mathrm{L}}\rangle$ (a particular sequence)

Abstracting away from the identity of L for now (which varies with the grammatical function of the adjectives), we can see several unusual traits in (30). First, the set in (30a) is multimembered and cannot be formed by Set Merge, which is binary by definition. This is a real curiosity, which I will temporarily ignore but return to in Section 4.2. Second, there is a conjunction & in the sequences in (30b–c) but none in the set in (30a), which leads us to ask where that conjunction comes from. In my implementation, I will remove & from (30b–c)—which I take to be part of the discourse—and relocate it to the numeration used in the actual derivation.

In fact, we can give (28) a bit more formal foundation. A sequence is like a list in computer languages (or a string in formal language theory), and the set of all possible sequences generated from a given set A is just the free list monoid A^* on A, whose identity element is the empty list $[]$ and whose monoid operation is list concatenation (++). See (31) for the definitions of monoid and free monoid.

(31) a. A monoid $\langle M, \cdot, e\rangle$ is a set M equipped with an associative binary operation $\cdot$ and an identity element e such that $\forall m \in M. e \cdot m = m \cdot e = m$.

b. The free monoid on a set has as elements all finite sequences (aka strings or lists) generated from zero or more elements of that set by concatenation.

Thus, the free list monoid on the set $\{a, b\}$ is the set $\{[], [a], [b], [a, b], [a, a], [a, a, b], \ldots\}$, where singleton entries like $[a]$ are generated by vacuous concatenations like $[a]{+}{+}[]$. Accordingly, what the ε-operator does in the Form Sequence procedure is pick a particular item out of the free monoid A^* on an initializing set A. I give the ε-term in (32a) and define its corresponding choice function in (32b).

(32) a. $\varepsilon x. \mathrm{seq}_A(x)$

b. $\Phi_{\mathrm{seq}_A} = \lambda A^*. a \in A^*$

As mentioned in Section 1, there are certain constraints on how the ε-operator picks out particular sequences in Chomsky's application (just as in von Heusinger's application in §3.2). Thus, sentences like the following crucially rely on the interconjunct ordering for correct interpretation. Accordingly, that ordering cannot be totally arbitrary but crucially reflects the speaker's volition.

(33) a. John and Bill saw Tom and Mary respectively. =(7)

b. As for fruit, I like apples, bananas, oranges, and strawberries—in that order.

The above constraint can be attributed to the semanticopragmatic interface or the discourse, which I will come back to in §4.2. In addition, there is a more fundamental and algorithmic constraint on the ε-choice in Form Sequence.

(34) A linguistically significant sequence (call it a *proper sequence* or *p-seq* for short) over a set A should have at least all members of A as components.

While the underspecification problem does not arise in Chomsky's discussion, it does in our formalization here, since a free list monoid on a set contains *all* concatenations of zero or more of its elements, including the empty list $[\,]$. I deem this formal perspective advantageous not only because it makes the Form Sequence theory more precise, but also because it suits Chomsky's pansequential view better. Recall from the beginning of this section that Chomsky assumes that "wherever there is an XP, there would be a sequence." With the list monoid, a noncoordinate XP can be viewed as a trivial list produced by concatenation with $[\,]$. That said, the underspecification problem does force us to accept the constraint in (34). Take (30) for example, which I repeat below with an updated presentation based on our discussion so far.

(35) young, tall, and happy

a. $\{\mathrm{young}^{\mathrm{L}}, \mathrm{tall}^{\mathrm{L}}, \mathrm{happy}^{\mathrm{L}}\}$ (a set of conjuncts with links)

b. $\{\langle\rangle, \langle \mathrm{young}^{\mathrm{L}}\rangle, \langle \mathrm{tall}^{\mathrm{L}}\rangle, \langle \mathrm{happy}^{\mathrm{L}}\rangle,$ (a set of potential sequences)
$\langle \mathrm{young}^{\mathrm{L}}, \mathrm{tall}^{\mathrm{L}}\rangle, \langle \mathrm{young}^{\mathrm{L}}, \mathrm{happy}^{\mathrm{L}}\rangle, \ldots,$
$\langle \mathrm{young}^{\mathrm{L}}, \mathrm{tall}^{\mathrm{L}}, \mathrm{happy}^{\mathrm{L}}\rangle, \langle \mathrm{young}^{\mathrm{L}}, \mathrm{happy}^{\mathrm{L}}, \mathrm{tall}^{\mathrm{L}}\rangle, \ldots,$
$\langle \mathrm{young}^{\mathrm{L}}, \mathrm{tall}^{\mathrm{L}}, \mathrm{happy}^{\mathrm{L}}, \mathrm{young}^{\mathrm{L}}\rangle, \ldots\}$

c. $\langle \mathrm{young}^{\mathrm{L}}, \mathrm{tall}^{\mathrm{L}}, \mathrm{happy}^{\mathrm{L}}\rangle$ (a particular sequence)

The constraint in (34) rules out underspecified sequences like $\langle\rangle$ and $\langle \mathrm{young}^{\mathrm{L}}\rangle$—not because those are ill-formed sequences, but because they do not meet communicative requirements and so have little linguistic significance. Underspecification amounts to the effect that, say, while a speaker has planned to convey the three ideas "young," "tall," and "happy" (in whichever order), they end up only conveying "young." That is an unlikely scenario under normal circumstances. In view of the constraint in (34), we can update the ε-term in (32) to the following version.

(36) $\varepsilon x.\, \text{p-seq}_A(x)$, where

$\lambda x.\, \text{p-seq}_A(x) \equiv \lambda x.\, \text{seq}_A(x) \wedge \forall y \in A. \exists i \in \mathbb{N}.\, \pi_i(x) = y$

($\pi_i(x)$ is the ith component of x)

In this way, we can guarantee that Form Sequence only forms sufficiently specified sequences even in cases where the interconjunct ordering has no interpretative significance (i.e., when the ε-choice is arbitrary). In sum, on a suitably abstract level, Form Sequence is the following procedure.

(37) a. Prepare conjuncts: $\mathrm{S}_i^{\mathrm{L}_i} (i \in \mathbb{N})$

b. Form initializing set: $A = \{\mathrm{S}_i^{\mathrm{L}_i}\}$

c. Form free monoid: $A^* := \langle A, {+}{+}, \langle\rangle\rangle$

d. Choose sequence: $[\varepsilon x.\, \text{p-seq}_A(x)] \in A^*$

4.2 Discussion

In the previous section, we have seen that the ε-operator is crucial in the Form Sequence procedure because it serves to pin down the final sequence. We have also seen that the ε-operator is not entirely indeterminate in Chomsky's application either, just like in the Konstanz School's application. Specifically, the ε-choice in Form Sequence references two types of discourse information: *i*) cross-sequence matching, and *ii*) speaker's intention. Such information may be either overtly signaled by expressions like *respectively* and *in that order* or silently understood. As such, speaker's intention (or agency) may qualify as an overarching label for the kind of discourse information the ε-operator is sensitive to. Recall from Section 3.2 that this is also the case in the Konstanz School's application, where, for instance, *the island* means different things for different speakers depending on their identities and communicative purposes. Overall, we can conclude that the ε-operator is semideterministic in both linguistic applications examined in this paper. I present this semideterminism in a comparative fashion in Table 1.

	Konstanz School	**Chomsky**
Deterministic	global ε (sensitive to discourse salience)	ε for CoP with significant interconjunct ordering
Nondeterministic	local ε	ε for CoP with random ordering

Table 1: Semideterminism of the ε-operator in two linguistic applications

There is a further crucial difference between the two applications of the ε-operator. Regardless of determinism, the Konstanz School's ε only takes effect at the syntax-semantics interface, as its choice only affects semantic interpretation but does not affect syntactic structure building or phonological spell-out—determiner phrases always have the structure [$_{\text{DP}}$ D NP] in syntax and get spelled out as *the/a NP* in English. As such, we can safely regard the Konstanz School's ε as a semantic tool. By comparison, Chomsky's ε takes effect at an earlier stage, since its choice affects both semantic interpretation and phonological spell-out. A most obvious piece of evidence for the latter is that each sequence determines a specific linear string. In Table 2 below, I illustrate the ε-operator's modular effects in the two linguistic applications.

Module	**Konstanz School's ε**	**Chomsky's ε**
Syntax	–	?
Semantics	choice function	CoP interpretation
Phonology	–	CoP pronunciation

Table 2: Modular effects of the ε-operator in two linguistic applications

Comparing the interface effects of ε in the two linguistic applications, we can see a striking difference. The interface effect of the Konstanz School's ε is a direct semantic interpretation of the ε symbol (as a choice function), whereas the interface effects of Chomsky's ε are only indirect effects of the ε symbol—they are direct effects of the semantic/phonological interpretation of the entire CoP in question instead. This indirectness makes the modularity of ε less clear in Chomsky's application. Assuming that each (underlying) syntactic object in natural language has a semantic interpretation (due to Frege's Principle), the fact that Chomsky's ε has no direct interpretation of its own becomes a suspicious symptom—and its lack of pronunciation adds to this suspicion. These two symptoms combined seem to suggest that Chomsky's ε—and thereby his Form Sequence—is somehow outside narrow-syntactic derivation despite its postsyntactic effects. But how could that be? Let us reexamine the Form Sequence procedure in (37), which is repeated in abbreviatory terms in Table 3.

Step	**Event**	**Locus**
1	conjunct preparation	syntax
2	initializing set formation	?
3	free monoid formation	?
4	sequence selection	?

Table 3: Four steps of Form Sequence

Step 1 in Table 3 is just the derivation of the individual conjuncts, which is no doubt a syntactic process. As for the remaining three steps, however, things are much less clear. As already mentioned in Section 4.1, the initializing set in Step 2 cannot be the product of Set Merge due to its multimembered nature, which means it cannot be formed in syntax. Step 3 is even less likely to be in syntax, since each syntactic derivation can only prove the well-formedness of a single expression; it cannot generate a whole set of potential expressions, let alone an infinite set thereof. Besides, nor do we have any derivation rule that can generate a monoid from a set. Step 4 is unlikely to be within syntactic derivation either, again due to the lack of a suitable derivation rule. Note that the ε-rule (i.e., the choice function) is not a syntactic rule, for it neither builds up syntactic objects nor manipulates them in the way that familiar syntactic rules (e.g., Merge, Agree, Move) do.

As far as I am concerned, Steps 2–4 in Table 3 are more like "backstage" processes rather than processes in syntactic derivation proper. That is, they serve to generate certain extra information (i.e., the sequence intended by the speaker) that may be referenced by (broad) syntax but is not part of it. Abusing the term "discourse" a bit, we can identify this speaker-intended sequential information as a type of discourse information. And if that is indeed the case, then all our puzzles above are solved. None of Steps 2–4 needs to have a corresponding derivation rule anymore, and the ε symbol no longer needs to have any direct semantic/phonological interpretation, for it is no longer part of natural language syntax (but merely part of a general-purpose formal language). Based on the above line of thought, I tentatively propose the theoretical context for Form Sequence in Table 4, where I distribute Chomsky's original conception in three big steps (Steps 1–3), and the discourse-level step (Step 2) is where the ε-operator takes effect.

As Table 4 shows, Form Sequence is not a purely syntactic operation but comprises both

Step	**Event**	**Locus**
1	conjunct preparation	syntax
2	sequence information generation (i) initializing set formation (ii) free monoid formation (iii) sequence selection	discourse
3	CoP formation (by multidimensional Pair Merge)	syntax
4	CoP interpretation (based on planned sequence)	interface

Table 4: A distributed conception of Form Sequence

syntactic and discourse-level computational processes. The discourse serves as a generative backstage or environment and accommodates all sorts of information that may guide interface interpretation but do not play a direct role in syntactic structure building, phonological externalization, or semantic composition. Many familiar types of information are held in this space, such as presupposition, deictic reference, as well as the salience ranking used by von Heusinger.[24] The semideterministic sequential information in coordination—which is essentially an order relation just like salience ranking—makes another good inhabitant of the discourse information stack.

The discourse as a level of representation evidently has access to some domain-general computational resources. For instance, von Heusinger's salience ranking and its dynamic update both require some algorithmic processing.[25] Considering this, the equally algorithmic generation of the sequential information in Step 2 should be handleable by domain-general computation as well. Indeed, the three substeps in Step 2 each corresponds to a quite general mental capacity of human beings—namely, information grouping (formally set formation), information concatenating (formally monoid formation), and information choosing (formally the Axiom of Choice). Moreover, this last capacity may be a particular manifestation of the even more general capacity of decision making. Since none of these capacities is specific to the language faculty, yet all of them can be accessed by it, they might qualify as aspects of a "third

[24]It might be fruitful to embed the current discussion in a dynamic theory of discourse. I leave that to future exploration.

[25]In fact, as I will show in §5, the salience ranking process itself can be treated as an instance of Form Sequence.

factor" (e.g., information processing principles) in the sense of Chomsky (2005).

4.3 Implementation

In this section, I provide a concrete implementation of Form Sequence in the distributed form in Table 4, beginning with the two syntactic steps. Step 1 generates individual conjuncts. Since no conjunct's derivation relies on any other's, I take Step 1 to be made up of a series of parallel derivations. This is also the assumption in Chomsky (2020). Thus, when analyzing the sentence in (29), which is repeated in (38), Chomsky remarks that "there are two parallel things generated separately. One of them is *arrive John*, the other is *John meet Bill*."

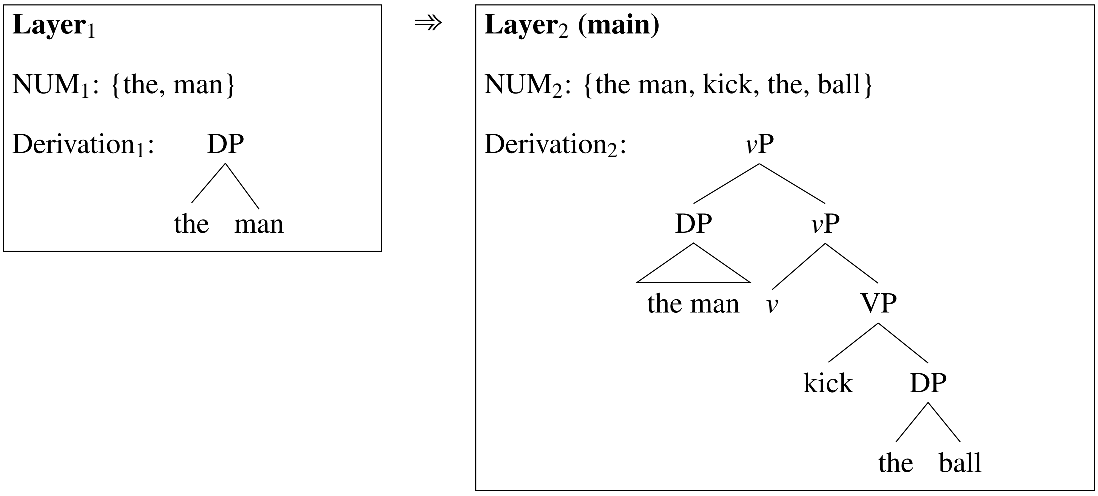
(38) a. John arrived and met Bill. =(29)

b. {C, {John_3, {INFL, $\langle$&, {$_1$ {$_2$ *v*, {arrive John_1}}}, {$_3$ John_2, {v^*, {meet B}}}$\rangle$}}}

To implement this parallel derivation, I adopt Zwart's (2007, 2009, 2011) theory of layered derivation and assume that each conjunct is derived in a separate layer. On Zwart's theory, each derivational layer is defined by a separate numeration (NUM),[26] and complex noncomplements like subjects are constructed in separate layers before they join the main layer. Specifically, one layer's output may be included in another layer's input (i.e., its numeration). Johnson (2003) calls this mechanism *renumeration*. See (39) for an illustration, where I use ⇛ to indicate a sequential relationship between derivational layers. I have omitted projections above *v*P for expository convenience.

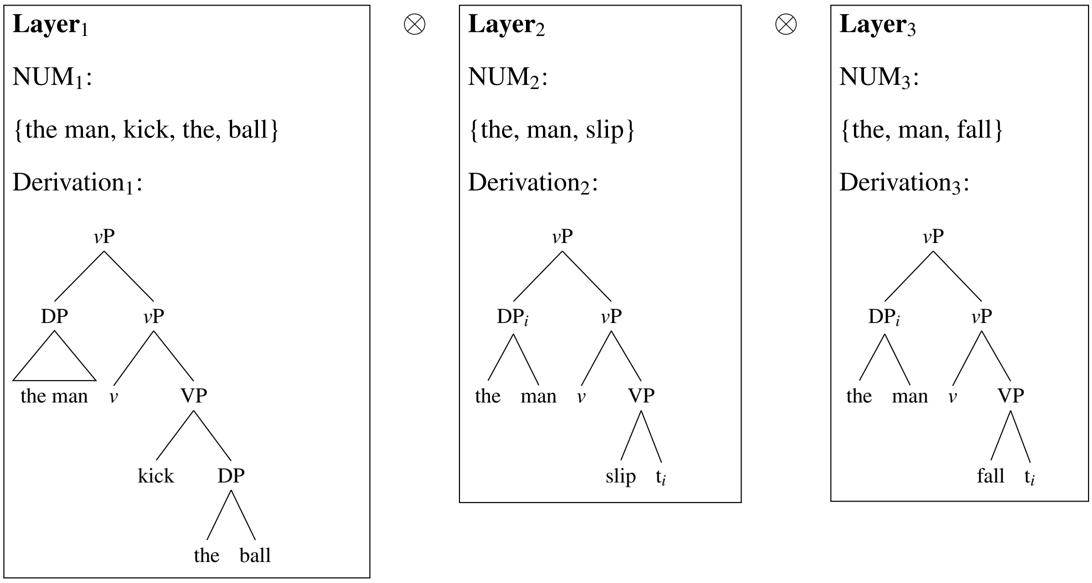
[26]For current purposes, I do not distinguish the two notions numeration (Chomsky 1995) and lexical array (Chomsky 2000). The numeration or lexical array for an entire derivation may be divided into subparts (called subarrays in Chomsky 2000), each defining a self-contained portion of the derivation. It might be interesting to explore the mereological relationship between numerations and their subparts in Zwart's theory.

(39) The man kicked the ball. (adapted from Zwart 2011:48)



Apart from complex subjects, Zwart suggests that several other types of syntactic objects are amenable to a layered-derivation analysis, including coordination. From the viewpoint of the current layer, elements derived in previous layers "have a dual nature," since they are "complex in the sense that they have been derived in a previous derivation [but] single items in that they are listed as atoms in the numeration for a subsequent derivation" (Zwart 2009:173). I illustrate the layered derivation of conjuncts in (40), where I use $\otimes$ to indicate a parallel relationship between derivational layers.[27]

(40) The man kicked the ball, slipped, and fell.



[27] The GB-style trace notation here is merely for expository convenience (to save space). I am not committed to such an analysis of movement. The different derivational status of subject and object DPs here is due to Zwart's particular view of argument structure. On a different view (e.g., that in Pylkkänen 2008 or Siddiqi 2009), both kinds of DPs may be treated as prederived objects.

I follow Chomsky in treating coordinate verbal predicates as full-fledged *v*Ps.[28] This calls for some careful handling of movement copies—and also a relaxation of the definition of copy—because while each layer in (40) contains its own "copy" of *the man*, there is only one copy of it in the final sentence. In the spirit of Chomsky (2020), any of the three occurrences of *the man*—which have identical interpretations—may raise to Spec-IP, while occurrences that do not raise are indistinguishable from copies of the raised occurrence and therefore deleted across the board.[29]

Next, the three conjuncts in (40) are renumerated into a new layer, and the new numeration additionally contains a functional category Co (whose semantic interpretation is a logical constant). This prepares the ground for Step 3 in Table 4. Subsequently, each of the conjunct pair-merges with Co in a separate plane, which yields a multidimensional object. I use dotted lines to indicate different planes.

(41)

Layer$_1$
Layer$_2$
Layer$_3$

$\Rrightarrow$

Layer$_4$

NUM$_4$: {Co,

[$_{vP_1}$ the man v_1 kick the ball]],

[$_{vP_2}$ the man v_2 slip]],

[$_{vP_3}$ the man v_3 fall]]}

Derivation$_4$: Co

vP$_1$ vP$_2$ vP$_3$

A multidimensional derivation of CoP was already proposed in de Vries (2004, 2005), and my dotted-line presentation above is borrowed from de Vries's work. In fact, the question de Vries was concerned about is not too different from our concern here:

> [H]ow can we represent the intuitive symmetry of coordination, and in particular, how can we prevent the first conjunct from c-commanding the second? (de Vries 2005:92)

[28] Chomsky (2020) assumes that tense is encoded in *v* instead of T, hence his returning to the older label INFL. I remain agnostic about this point, and nothing in my implementation hinges on it.

[29] Chomsky's proposal here relies on the assumption that these parallel conjuncts are not islands.

And de Vries's way of constructing multidimensional objects is not that different from Chomsky's multidimensional Pair Merge either. He proposed a "behindance" relation in addition to dominance based on the conception that "conjuncts are behind each other in a three-dimensional structure" (ibid.) and, accordingly, a "b-Merge" operation (i.e., Merge by behindance). Noticing the striking similarity between b-Merge and Pair Merge, Song (2017:24) practically identifies the two. Despite their different details, de Vries's idea and Chomsky's idea on coordination are almost the same.[30]

Recall from §4.1 that the conjunction is optional in Chomsky's theory. Deviating from this position but in line with de Vries (2005), I take Co to be always present in the underlying syntax of coordination—what is optional is its phonological exponence. My rationale is as follows. Based on Chomsky (2019, 2020), each conjunct is added from a different dimension to a pivot point, and to get the dimension-expanding effect we need Pair Merge between the pivot and the conjunct, as in (27). This, then, rules out the possibility that the pivot is just the link element shared by all conjuncts in a CoP, because the link is subject to Set Merge within its ambient phrase marker, as I have highlighted with Chomsky's example in (29), where the link is the phase head *v* and its ambient phrase *v*P is derived by set-merging *v* and VP. But if the link is not the pivot we need, then what is? What we know is that the pivot serves to hold the conjuncts together, so it lies at the intersection of all the dimensions in a CoP—or from another angle, it lies outside the local domain of any dimension. Such categorial independence exempts it from the dimension-internal, run-of-the-mill derivational relations (i.e., relations engendered by Set Merge), such as c-command, head-specifier/complement, and the like. Therefore, an ideal pivot is a syncategorematic functor with flexible arity (i.e., one that accepts any number of arguments). The logical connectives AND/OR are the obvious candidates, which I notate by the umbrella label Co.

As mentioned in Section 4.1, I make a distinction between the three terms link, host, and attachment point. Adding in the above-discussed pivot, we now have four separate notions associated with the multidimensional CoP (which further confirms that multidimensional syntactic objects are really complex creatures). Following Chomsky (2019), the link in (41) is just

[30]Except for the potential technical difference that a 3-dimensional space is enough for de Vries's theory, whereas Chomsky's theory may require a truly *n*-dimensional space. See note 23.

the phase head *v*. I assume a well-formedness constraint requiring that all links in the same CoP be the same, at least in terms of interpretation. I take this constraint to be what underlies Chomsky's idea that all the links in a CoP are identical.[31] Compared to the link, the host component in (41) is much harder to satisfactorily pin down. In a pair-merged object $\langle \alpha, \beta \rangle$, α is the adjunct and β is the host, and the category of the entire object is the same as that of β (i.e., β "projects" in traditional parlance). For instance, the category of *young man* is the same as that of *man*. However, there is a crucial difference between this classical scenario of Pair Merge and the multidimensional Pair Merge involved in (41), and we encounter nontrivial trouble if we simply take the pivot of the multidimensional object (i.e., Co) to be the host, even though coordinate phrases are often conveniently labeled as CoPs as if Co were the head,[32] including in this paper. The trouble is that while the host in classical Pair Merge (e.g., *man* in *young man*) can be used on its own, Co cannot—hence its syncategorematicity—nor are the conjuncts its modifiers in any sense. Intuitively, if anything in a CoP projects at all, it should be the conjuncts' (shared) category—namely, the link—rather than Co. Therefore, if we reserve the term host for the labeling component as in standard Pair Merge theory, then the host of each Co-XP pair should be XP instead of Co. This brings us to the somewhat peculiar conclusion that the multidimensional CoP is multi-hosted—though probably zero-headed (see note 32)—with as many hosts as its dimensions. That being said, these "hosts" are still not the same as those in classical Pair Merge, for Co is not a modifier of XP in any sense either, though it is formally "adjoined" to XP, just as XP is not a modifier of Co. In any case, this state of affairs suggests that we cannot understand the projection/labeling nature of a multidimensional object in the same way as we understand that of an ordinary, set-merged object. Henceforth, I will use $\langle \mathrm{Co}, \mathrm{XP} \rangle$ to notate the Co-XP pair and call XP the host, though this designation is more formal than substantive. To wrap up this paragraph, the link component in (41) is *v*, the pivot is Co, and the hosts are the three *v*Ps. I will turn to address the term attachment point shortly. I use Figure 1 to illustrate the internal relationships of a multidimensional syntactic object.

[31] Satisfaction of this constraint may also be what makes a multidimensional CoP labelable. Thus, if the link element is *v*, then the CoP's real label is *v*P, which I will write as Co*v*P to make it clear that this is a coordinate phrase. This scenario constitutes a special instance of the XP-YP case in Chomsky's (2013, 2015) labeling algorithm, where the label is provided by some shared feature(s).

[32] As discussed above, due to the non-set-merged nature of multidimensional objects, terms like head and complement may not be applicable to them at all.

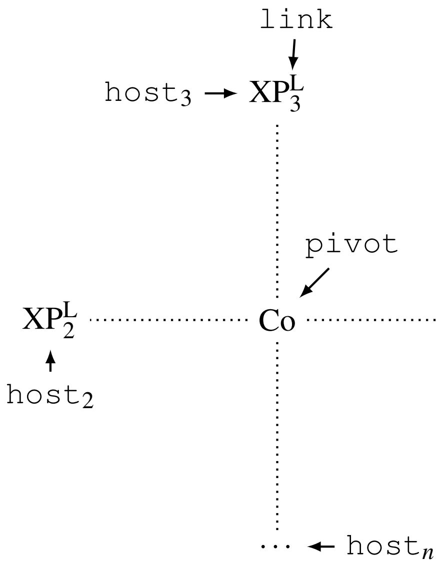


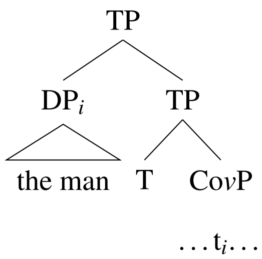
Figure 1: Relationships in a multidimensional CoP

After the multidimensional CoP is derived, the next step is to merge it into the main structure, again in a new derivation layer—this time the main layer. In the above example, the coordinate *v*P merges with T (or some other inflectional category), and this is normal Set Merge, as in (42).

(42) **Layer$_4$** ⇛ **Layer$_5$ (main)**

NUM$_5$: {Co*v*P, T, …}

Derivation$_5$:



The complement position of TP is occupied by the coordinate *v*P from layer$_4$, which I have labeled as Co*v*P to be a bit more informative (see note 31). One of the several occurrences of *the man* raises from *v*P to Spec-TP as aforementioned.[33] Finally, let us put the five derivational layers above together, which gives us the procedure below. Note that ⊗ binds more tightly than

[33]Chomsky does not specify how exactly this cross-dimensional raising takes place, though it is clearly needed in his theory, as shown in (29). In my implementation, I tentatively suggest that the pivot Co, being the connection between each of the dimensions in Co*v*P and the main dimension (where T lives in), may serve as a bridge or "edge" for cross-dimensional movement. The detailed motivation and systematic ramifications of this proposal must be left to future research.

$\Rrightarrow$.

(43) $\textbf{Layer}_1 \otimes \textbf{Layer}_2 \otimes \textbf{Layer}_3 \Rrightarrow \textbf{Layer}_4 \Rrightarrow \textbf{Layer}_5$ **(main)**

The above derivation is difficult to illustrate with conventional tree diagrams due to its multi-layeredness and multidimensionality, but it can be easily illustrated by a proof tree, as in Figure 2. The proof tree also shows the complex subject layer (i.e., that for *the man* in *the man kick the ball*) that I have glossed over above. Note that the final line of the proof (i.e., its "root") highly resembles Chomsky's (2020) IP structure in (29)/(38), except that for reasons mentioned in Section 4.1 I do not treat the multidimensional CoP itself as a tuple/sequence in syntax but treat it as a set of pairs with a shared component (i.e., Co). Mathematically, this meets the definition of a partial order, more exactly one where one element is ranked above everything else. Recall that we have seen such a partial order in Section 3.3 when commenting on von Heusinger's salience ranking procedure. In sum, we can view multidimensional Pair Merge as an operation that takes a certain kind of numeration (i.e., a set with a pivot and some syntactic objects with a common link) as input and yields a partially ordered set as output. This is exactly what happens in the step labeled "s" in Figure 2.

The proof tree in Figure 2 is a comprehensive representation of my implementation of Steps 1 and 3 in Table 4. Next I turn to associate Step 2 (Form Sequence) and Step 4 (interface interpretation) with the syntactic derivation. Both steps are straightforward now that we have had the derivational layers laid out. Step 2 kicks off in the backstage as soon as all the individual conjuncts are derived—namely, at the end of Step 1—because that is when the initializing set of Form Sequence becomes available. Step 2 and Step 3 are two mutually independent processes, with the former going on in the discourse[34] and the latter going on in narrow syntax. When the product of Step 3 is sent to the interfaces for interpretation, the phonological and semantic modules can both access the sequential information in the discourse environment, whereby the multidimensional CoP gets a linear order on the one hand and a reading with event-sequencing on the other. That is, in this example the ε-choice in Form Sequence is determined by the temporal order in which the three events happened. See Figure 3 for a vivid illustration of the

[34]Recall from §4.2 that I hold a more generalized view of the notion discourse than that usually assumed in semantic works. I assume that there is a discourse module in the human mind for all kinds of backstage resources and processes supporting language production and comprehension.

$$
\dfrac{\dfrac{}{\text{T}}\,\text{a} \qquad \dfrac{\dfrac{}{\text{Co}}\,\text{a} \quad \dfrac{\dfrac{\dfrac{}{\text{the}}\,\text{a} \quad \dfrac{}{\text{man}}\,\text{a}}{\text{the man}}\,\text{s} \quad \dfrac{\dfrac{}{v}\,\text{a} \quad \dfrac{\dfrac{}{\text{kick}}\,\text{a} \quad \dfrac{\dfrac{}{\text{the}}\,\text{a} \quad \dfrac{}{\text{ball}}\,\text{a}}{\text{the ball}}\,\text{s}}{\text{kick the ball}}\,\text{s}}{v\text{ kick the ball}}\,\text{s}}{\text{the man } v \text{ kick the ball}}\,\text{s} \quad \dfrac{\dfrac{\dfrac{}{v}\,\text{a} \quad \dfrac{\dfrac{}{\text{slip}}\,\text{a} \quad \dfrac{\dfrac{}{\text{the}}\,\text{a} \quad \dfrac{}{\text{man}}\,\text{a}}{\text{the man}}\,\text{s}}{\text{slip the man}}\,\text{s}}{v\text{ slip the man}}\,\text{s}}{\text{the man } v \text{ slip}}\,\text{i} \quad \dfrac{\dfrac{\dfrac{}{v}\,\text{a} \quad \dfrac{\dfrac{}{\text{fall}}\,\text{a} \quad \dfrac{\dfrac{}{\text{the}}\,\text{a} \quad \dfrac{}{\text{man}}\,\text{a}}{\text{the man}}\,\text{s}}{\text{fall the man}}\,\text{s}}{v\text{ fall the man}}\,\text{s}}{\text{the man } v \text{ fall}}\,\text{i}}{\{\langle \text{Co}, \{\text{the man } v \text{ kick the ball}\}\rangle, \langle \text{Co}, \{\text{the man } v \text{ slip}\}\rangle, \langle \text{Co}, \{\text{the man } v \text{ fall}\}\rangle\}}\,\text{pp}}{\dfrac{\{\text{T}, \{\langle \text{Co}, \{\text{the man } v \text{ kick the ball}\}\rangle, \langle \text{Co}, \{\text{the man } v \text{ slip}\}\rangle, \langle \text{Co}, \{\text{the man } v \text{ fall}\}\rangle\}\}}{\{\text{the man}, \{\text{T}, \{\langle \text{Co}, \{\text{the man } v \text{ kick the ball}\}\rangle, \langle \text{Co}, \{\text{the man } v \text{ slip}\}\rangle, \langle \text{Co}, \{\text{the man } v \text{ fall}\}\rangle\}\}\}}\,\text{i}}\,\text{s}
$$

Figure 2: An example proof tree for multilayered, multidimensional derivation
a: axiom s: set merge
i: internal merge pp: multidimensional pair merge

distributed Form Sequence procedure in Table 4.

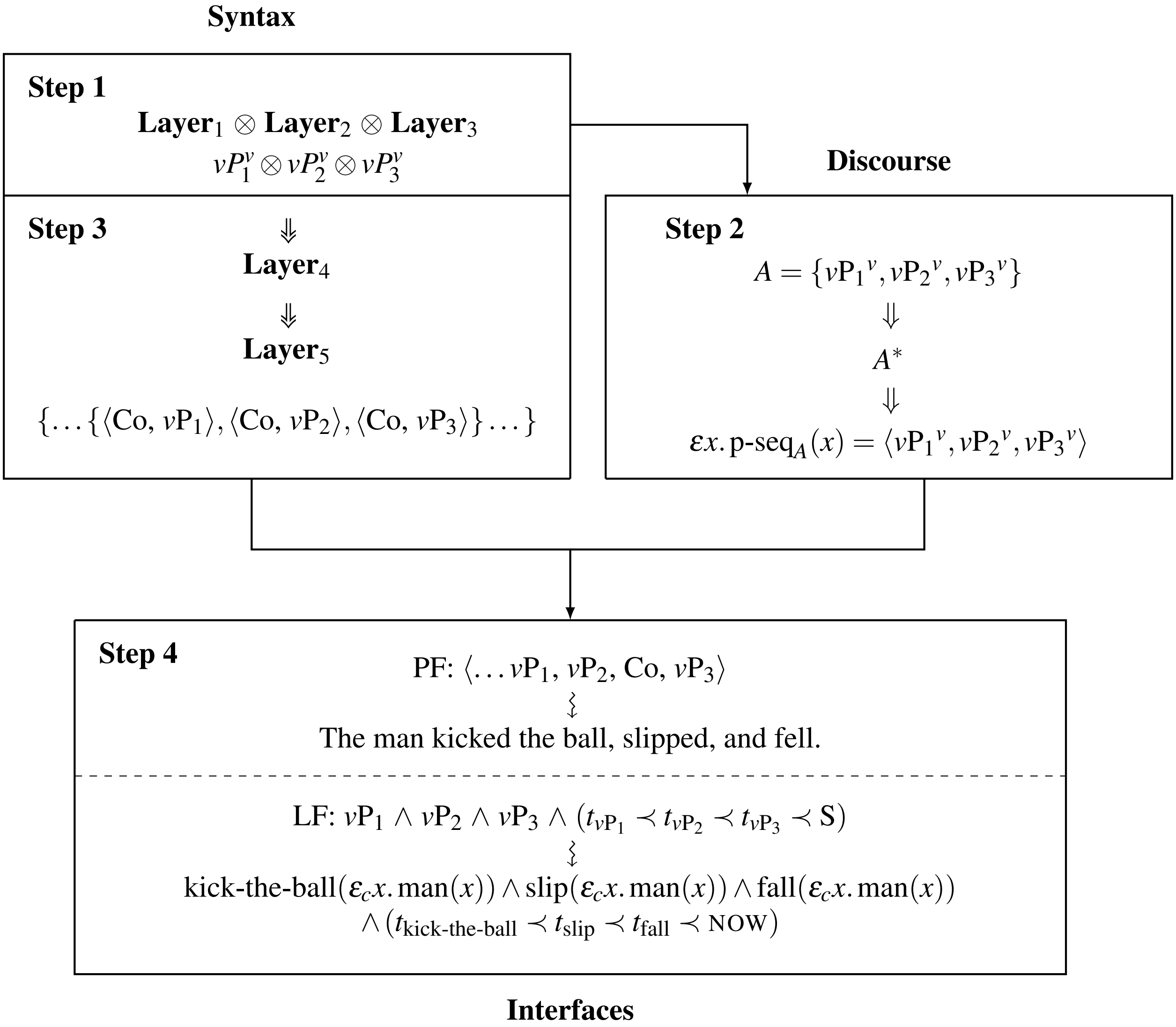


Figure 3: An illustration of the distributed conception of Form Sequence

5. Similarities and differences

In Sections 3 and 4, I examined two linguistic applications of Hilbert's ε-operator, one by the Konstanz School (most representatively by von Heusinger) and the other by Chomsky. The former is in the area of semantics, while the latter is in that of syntax. In this section, I summarize the key similarities (§5.1) and differences (§5.2) between the two applications revealed in the foregoing discussion.

5.1 Similarities

Despite their disparate research domains and major concerns, the two applications manifest some striking similarities, which I will summarize below.

I. Semideterminism The most noteworthy similarity between the two applications is that the ε-operator is semideterministic in both of them—namely, the choice it makes can be either determinate or indeterminate depending on the context of language use. By contrast, the original mathematical conception of ε is fully nondeterministic. As I pointed out in Section 1, the semideterminism of the ε-operator in linguistics is due to the context-sensitive nature of human language.

The Konstanz School's deterministic ε—namely, their indexed global ε—makes its fixed choice by referencing the ranking of a predicate's set-theoretic members. By comparison, Chomsky's deterministic ε fixes its choice based on world knowledge or the speaker's volition. See (44) for an illustration.

(44) Deterministic ε

a. The man smokes.

$[\![\text{the man}]\!]^c = [\![\varepsilon_c x.\,\text{man}(x)]\!] =$ the most salient man in context c

b. I like apples, bananas, oranges, and strawberries—in that order.

Form Sequence($\{$s, a, o, b$\}^c$) = $\varepsilon_c x.\,\text{p-seq}_{\{\text{s, a, o, b}\}^*}(x) = \langle \text{a, b, o, s} \rangle$

In (44a), the man chosen by ε_c is the most prominent male individual in context c, probably because he has just been mentioned. In (44b), the ordering of the four kinds of fruit is also relativized to a specific context—namely, that where the sentence is uttered. As for the nondeterministic version of ε, it is just the normal choice operator from mathematics, which does not rely on contextual information when making its choice. See (45) for example.

(45) Nondeterministic ε

a. A man spoke.

$[\![\text{a man}]\!] = [\![\varepsilon x.\,\text{man}(x)]\!] =$ an arbitrary man

b. There are apples, bananas, oranges, and strawberries in the picture.

Form Sequence($\{$s, a, o, b$\}$) = $\varepsilon x.\,$p-seq$_{\{\text{s, a, o, b}\}^*}(x) = \langle$a, b, o, s$\rangle$

In (45a), the man chosen by ε is unknown, as this is the first time he has been mentioned. Similarly, the particular fruit sequence chosen in (45b) by the context-free ε is random, which implies no significant ordering effect in the reading of the utterance.

II. Ancillary tool A second prominent similarity between the two applications is that in both of them the ε-operator is more of an ancillary tool than an integral part of the linguistic module under investigation. Thus, even though the ε-calculus (and also the λ-calculus for that matter) is a useful tool to mediate between natural language syntax and model-theoretic semantics, it is not part of the model itself and so is in principle dispensable. Indeed, von Heusinger (2013) dispenses with the ε-operator and reformulates his semantic theory solely in terms of choice functions. Similarly, even though the ε-operator in Chomsky's application crucially serves to pin down the interconjunct ordering of a coordinate phrase, that process is unlikely to take place in narrow syntax but more likely takes place in the discourse as I have argued in Section 4.1. Consequently, the same multidimentional syntactic object may get different surface orders in different contexts. This distributed conception of Form Sequence is advantageous in that it keeps narrow syntax simple, which is especially desirable in Chomsky's new theory since the multidimensional CoP itself is already complicated enough to implement, as I have demonstrated in Section 4.3.

III. Salience ranking via Form Sequence Apart from the above two obvious similarities, there is still a third, less obvious similarity between the two applications. Recall from Section 3.3 that the salience ranking in von Heusinger's theory is essentially a total order relation.[35] Similarly, the sequence in Chomsky's Form Sequence theory is also a total order. Thus, both salience ranking and Form Sequence can be viewed as procedures that generate a total or-

[35] In §3.3, I suggested a partial-order-based modification of von Heusinger's theory. Under that modification, to maintain the validity of the discussion here we would need to further generalize Form Sequence into something like Form Poset. Since that mission would take us too far afield, I restrict my comparison here to one between von Heusinger's original theory and Chomsky's theory.

der on a set. This formal similarity opens up the possibility to apply Form Sequence to von Heusinger's theory—namely, to view salience ranking as an instance of Form Sequence. I illustrate this possibility in (46).

(46) Form Sequence($[\![\text{man}]\!]^{\langle \text{salience}, c\rangle}$) = $\varepsilon_{\langle \text{salience}, c\rangle} x.\, \text{p-seq}_{[\![\text{man}]\!]^*}(x)$
(a salience-based ordering of the elements in $[\![\text{man}]\!]$ in context c)

While von Heusinger does not specify how the salience ranking comes about in the discourse but simply assumes its existence, Chomsky's theory in a sense helps us fill that gap. This confirms my suggestion in Section 4.2 that Form Sequence is not a syntax-specific procedure but a more general tool for information processing. On the above perspective, we can now reformulate von Heusinger's analysis of definite NPs with two steps of ε-choice, as in (47).

(47) *the F*, context c

a. Form salience ranking: $\langle \text{F}, >_c \rangle = \varepsilon_{\langle \text{salience}, c\rangle} x.\, \text{p-seq}_{[\![\text{F}]\!]^*}(x)$
(choose a salience-based sequence from all proper sequences on the set $[\![\text{F}]\!]$)

b. Choose element: $a = \varepsilon_c x.\, [\![\text{F}]\!]_{>_c}(x)$
(choose a particular element out of the salience-ranked set $[\![\text{F}]\!]$)

Both steps have a deterministic ε, and they in fact "agree" in the context parameter c. Clearly, here too the Form Sequence procedure takes place in the discourse. Recall from Section 1 that one of the questions we asked was how the two applications of the ε-operator reviewed in this paper could coexist in the same grammar. The above illustration provides a possible answer. They can coexist in the same grammar because the ε-operator is a general-purpose tool for choice-making, and it so happens that there are multiple places in the analysis of any grammar where certain metalinguistic choices need to be made in the discourse. The choice of referents and the choice of sequences are but two particular examples.

5.2 Differences

I. Modularity Of course, the two linguistic applications of ε have significant differences too, one of which has already been mentioned in Section 4.2 and also been repeated above. That is, they differ in the module where the ε-operator takes effect. It takes effect in the semantic module in the Konstanz School's application and in the discourse in Chomsky's application. Thus, in von Heusinger's theory the semantics of the ε-operator is also the semantics of some types of natural language expressions ([in]definite NPs), whereas in Chomsky's theory the semantics of the ε-operator is not the semantics of any natural language expression. In other words, the ε-operator is used as part of a mediating formal language between natural language grammar and its semantic models in the Konstanz School's application but not in Chomsky's application. This is unsurprising, because the same formal language can serve different purposes in different scientific areas, and it just so happens that in Chomsky's theory the ε-calculus serves not to represent natural language grammar but to represent the "grammar" of some more general thought pattern.

II. Choice-determining factors Apart from the above fundamental difference, there is another more subtle difference between the two applications, again in relation to how the ε-choice is made. Above we have seen that both applications make use of a deterministic version of the ε-operator, but that determinism is much more regular in von Heusinger's global choice function than in Chomsky's Form Sequence. All global choices for definite NPs are fully determined by the same contextual factor (i.e., salience) in the same way, whereas factors affecting the choices made by Form Sequence are much more diverse. Above we have seen two such factors: speaker's volition (e.g., in sentences with an *in that order* reading) and world knowledge (e.g., in sentences with an event-sequencing reading). In addition, in my monoid-based understanding of Form Sequence in Section 4.2, I further included a choice-determining constraint in the definition of Form Sequence, ruling out underspecified sequences right from the beginning.

All the choice-determining factors we have seen in this paper are semantically or pragmatically oriented, but there may well be other types of factors as well. For instance, in Author (2021) I investigate a case where the most crucial choice-determining factor is phonological. Faced with such a variety of factors, we can ask two further questions: *i*) How exactly are those different types of information accessed by Form Sequence? *ii*) Are there any constraints on the types of factors that may influence ε-choices?

Answers to both questions can be found in my implementation in Section 4.3. For the first question, since the ε-based step of Form Sequence (i.e., Step 2 in Figure 3) takes place in the discourse, the choice-making process has access to any information that is available in the discourse, such as the relevant world knowledge, the communicative context, the speaker's volition, etc. In addition, for the Form Sequence procedure to successfully proceed, we must also assume that numerations—both initial ones directly formed from the lexicon and renumerated ones formed in the course of layered derivation—are available in the discourse. This is a reasonable assumption because numeration formation is strictly speaking not part of narrow syntax but part of planning, which is another backstage process that is best attributed to the derivational environment (i.e., the discourse in the broad sense in note 34).

The above answer to the first question also makes the answer to the second question easy to see. Since the derivational environment is not the derivation per se, the discourse by assumption can only access information that is in the numeration or in the discourse but cannot access information that requires derivational procedures—unless such information gets renumerated. A straightforward implication of this is that the ε-operator in Form Sequence can be sensitive to factors like the lexical information in individual lexical items (e.g., their sounds and meanings) but not to any syntax-internal relations, such as c-command, head-complement/specifier, locality, minimality, and many others. This prediction matches our observations of the two applications in this paper.

6. Conclusion

Hilbert's epsilon operator is a valuable formal tool from David Hilbert's work on the foundations of mathematics. In this paper, I have comparatively examined two applications of it in theoretical linguistics, one by the Konstanz School, especially by Klaus von Heusinger, since the 1980s and the other by Noam Chomsky since 2019. The Konstanz School's application is on the semantics of (in)definite NPs and intersentential anaphora, while Chomsky's application is on the syntax of coordination, especially the unbounded unstructured case thereof. This paper has two main goals: *i*) Compare the two applications on a metatheoretical level and thereby examine the status of the ε-operator in linguistic theory. *ii*) Provide an implementation of Chomsky's ε-based new theory (Form Sequence).

I addressed the first goal in Sections 3, 4, and 5 and addressed the second goal in Section 4.3. In particular, since Chomsky's application is very new, with little discussion in the literature, my implementation of it heavily draws on my own understanding of the matter, which is elaborated in Sections 4.1–4.2. I have also built on Zwart's layered derivation theory and de Vries's 3D coordination theory. Specifically, I implemented Form Sequence in a distributed fashion, partly in narrow syntax and partly in the derivational environment or discourse. The formation of the sequence (i.e., the ε-based part) takes place in the discourse, while the construction of the actual CoP (by multidimensional Pair Merge) takes place in syntax. As such, the sequential information is not available in syntax but can be accessed by the interpretative interfaces in the discourse.

I have summarized the main results of my metatheoretical comparison in Section 5. Below is a brief recapitulation. The two linguistic applications of ε have three major similarities. First, the ε-operator is semideterministic and often context-sensitive in both of them. Second, the ε-operator is not an integral part of the linguistic module under study in either application but a general-purpose tool for information processing (i.e., a third-factor tool) available to the human mind. Third, the two applications partially overlap since they both involve a total order component, which can be pinned down by the same usage of ε (that of Chomsky).

There are two main differences between the two applications. First, their uses of the ε-operator take effect at different levels—in semantics in the Konstanz School's application and directly in the discourse and indirectly at the syntax-semantics/phonology interfaces in Chomsky's application. Second, the context-sensitivity of ε is much more regular and also more limited in Konstanz School's application (where the only factor is salience) than in Chomsky's application (where a variety of factors can play a role).

Overall, Chomsky's application of the ε-operator is more general in spirit, as it can be easily extended from Chomsky's immediate area and phenomenon of interest (i.e., clausal syntax, coordination) to other areas and phenomena, as revealed in Section 5 (semantics, definite NP) and my ongoing work (word formation, compounding). All these manifestations of the ε-effect suggests that Chomsky's new theory has a highly fundamental place in the minimalist program, if not in narrow syntax per se.

References

Abrusci, V. M. 2017. Hilbert's τ and ε in proof theoretical foundations of mathematics: An introduction. *The IfCoLog Journal of Logics and their Applications: Special Issue—Hilbert's Epsilon and Tau in Logic, Informatics, and Linguistics* 4, 257–274.

Asser, G. 1957. Theorie der logischen Auswahlfunktionen. *Mathematical Logic Quarterly* 3(1-5), 30–68.

Avigad, J. & R. Zach. 2020. The epsilon calculus. In E. N. Zalta (ed.), *The Stanford encyclopedia of philosophy* (Fall 2020 edn.), Metaphysics Research Lab, Stanford University.

Bernays, P. 1991/1958. *Axiomatic set theory*. New York: Dover Publications.

Blümel, A. 2020. A uniform NP-approach to (poly-)definiteness and nominal concord. Talk at CamCoS 9 New (online), Sep 8–11.

Bourbaki, N. 1970/1954. *Éléments de mathématique 1: Theorie des ensembles*. Paris: Hermann.

Chatzikyriakidis, S., F. Pasquali & C. Retoré. 2017. From logical and linguistic generics to Hilbert's tau and epsilon quantifiers. *The IfCoLog Journal of Logics and their Applications: Special Issue—Hilbert's Epsilon and Tau in Logic, Informatics, and Linguistics* 4, 231–256.

Chomsky, N. 1995. *The minimalist program*. Cambridge MA: MIT Press.

Chomsky, N. 2000. Minimalist inquiries: The framework. In R. Martin, D. Michaels & H. Uriagereka (eds.), *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 89–156. Cambridge MA: MIT Press.

Chomsky, N. 2004. Beyond explanatory adequacy. In A. Belletti (ed.), *Structures and beyond*. The Cartography of Syntactic Structures, vol. 3, 104–131. Oxford: Oxford University Press.

Chomsky, N. 2005. Three factors in language design. *Linguistic Inquiry* 36(1), 1–22.

Chomsky, N. 2013. Problems of projection. *Lingua* 130, 33–49.

Chomsky, N. 2015. Problems of projection: Extensions. In E. Di Domenico, C. Hamann & S. Matteini (eds.), *Structures, strategies and beyond: Studies in honour of Adriana Belletti*, 1–16. Amsterdam: John Benjamins.

Chomsky, N. 2019. Lecture 4, UCLA lecture series, May 2. https://linguistics.ucla.edu/noam-chomsky/.

Chomsky, N. 2020. Lecture at the 161st Meeting of the Linguistic Society of Japan, Nov 22. https://youtu.be/X4F9NSVVVuw.

Curry, H. B. 1963. *Foundations of mathematical logic*. New York: McGraw-Hill.

David, H. 1923. Die logischen Grundlagen der Mathematik. *Mathematische Annalen* 88, 151–165.

Egli, U. 1991. (in)definite Nominalphrase und Typentheorie. Arbeitspapier 27. Fachgruppe Sprachwissenschaft Universität Konstanz.

Egli, U. & K. von Heusinger. 1995. The epsilon operator and E-type pronouns. In U. Egli, P. E. Pause, C. Schwarze, A. von Stechow & G. Wienold (eds.), *Lexical knowledge in the organization of language*, 121–141. Amsterdam: John Benjamins.

Evans, G. 1977. Pronouns, quantifiers, and relative clauses I. *Canadian Journal of Philosophy* 7(3), 467–536.

Gödel, K. 1931. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* 38(1), 173–198.

Heim, I. 1982. The semantics of definite and indefinite noun phrases. University of Massachusetts, Amherst dissertation.

von Heusinger, K. 1997a. Definite descriptions and choice functions. In S. Akama (ed.), *Logic, language, and computation*, 61–91. Berlin: Springer.

von Heusinger, K. 1997b. *Salienz und Referenz: Der Epsilonoperator in der Semantik der Nominalphrase und anaphorischer Pronomen*. Berlin: Akademie Verlag GmbH.

von Heusinger, K. 2000. The reference of indefinites. In K. von Heusinger & U. Egli (eds.), *Reference and anaphoric relations*, 247–265. Dordrecht: Springer.

von Heusinger, K. 2002. Specificity and definiteness in sentence and discourse structure. *Journal of Semantics* 19(3), 245–274.

von Heusinger, K. 2004. Choice functions and the anaphoric semantics of definite NPs. *Research on Language and Computation* 2(3), 309–329.

von Heusinger, K. 2013. The salience theory of definiteness. In A. Capone, F. Lo Piparo & M. Carapezza (eds.), *Perspectives on linguistic pragmatics*, 349–374. Cham: Springer.

Hilbert, D. 1926. Über das Unendliche. *Mathematische Annalen* 95(1), 161–190.

Hilbert, D. & P. Bernays. 1939. *Grundlagen der Mathematik*, vol. 2. Berlin: Springer.

Jech, T. J. 1973. *The axiom of choice*. Amsterdam: North-Holland Publishing Company.

Johnson, K. 2003. Towards an etiology of adjunct islands. *Nordlyd* 31(1), 187–215.

Kamp, H. 1981. A theory of truth and semantic interpretation. In J. Groenendijk, T. Janssen & M. Stokhof (eds.), *Formal methods in the study of language*, 277–322. Mathematical Center: Amsterdam.

Kratzer, A. 1998. Scope or pseudoscope? Are there wide-scope indefinites? In S. Rothstein (ed.), *Events and grammar*, 163–196. Dordrecht: Springer.

Leisenring, A. C. 1969. *Mathematical logic and Hilbert's ε symbol*. London: Macdonald & Co. Ltd.

Lewis, D. 1979. Scorekeeping in a language game. In R. Bäuerle, U. Egli & A. von Stechow (eds.), *Semantics from different points of view*, 172–187. Berlin: Springer.

Muskens, R. 1989. A relational formulation of the theory of types. *Linguistics and Philosophy* 325–346.

Ott, D. 2021. Unstructured sequences: A non-syntactic approach. In T. S. Kim & S. Bae (eds.), *Proceedings of the 23rd Seoul International Conference on Generative Grammar*, 104–121.

Pylkkänen, L. 2008. *Introducing arguments*. Cambridge MA: MIT Press.

Retoré, C. 2014. Typed Hilbert epsilon operators and the semantics of determiner phrases. In G. Morrill, R. Muskens, R. Osswald & F. Richter (eds.), *Proceedings for the 19th international conference on formal grammar*, 15–33. Heidelberg: Springer.

Schröder, B. 2016. *Ordered sets: An introduction with connections from combinatorics to topology* (2nd edn.). Cham: Birkhäuser.

Siddiqi, D. 2009. *Syntax within the word: Economy, allomorphy, and argument selection in distributed morphology*. John Benjamins.

Sigurðsson, H. Á. 2020. Universality and variation in language: The fundamental issues. *Evolutionary Linguistic Theory* 2(1), 5–29.

Slater, H. 1986. E-type pronouns and ε-terms. *Canadian Journal of Philosophy* 16(1), 27–38.

Slater, H. 2017. $(\exists y)(y = \varepsilon x f x)$. *The IfCoLog Journal of Logics and their Applications: Special Issue—Hilbert's Epsilon and Tau in Logic, Informatics, and Linguistics* 4, 276–286.

Song, C. 2017. Prosodically-driven morphosyntactic change? Revisiting the history of Chinese disyllabic words. In C. Song & J. Baker (eds.), *Cambridge occasional papers in linguistics*, vol. 10, 152–196. Department of Theoretical and Applied Linguistics, University of Cambridge.

de Vries, M. 2004. Parataxis as a different type of asymmetric Merge. In *Proceedings of the Interfaces Conference hosted by the International Conference on Advances in the Internet, Processing, Systems, and Interdisciplinary research (CD-ROM)*, .

de Vries, M. 2005. Coordination and syntactic hierarchy. *Studia Linguistica* 59, 83–105.

Whitehead, A. N. & B. Russell. 1910. *Principia mathematica*, vol. 1. Cambridge: Cambridge University Press.

Winter, Y. 1997. Choice functions and the scopal semantics of indefinites. *Linguistics and Philosophy* 399–467.

Zwart, J.-W. 2007. Layered derivations. Advanced Core Training in Linguistics lecture, Mar 9, London.

Zwart, J.-W. 2009. Prospects for top-down derivation. *Catalan Journal of Linguistics* 8, 161–187.

Zwart, J.-W. 2011. Recursion in language: A layered-derivation approach. *Biolinguistics* 5(1-2), 043–056.