



English Syntax Onset

A Discovery-Based Introduction
to Generative Syntax



Yanuaris Yanu Dharmawan
Dini Hidayati

English Syntax Onset

A Discovery-Based Introduction
to Generative Syntax

English Syntax Onset: A Discovery-Based Introduction to Generative Syntax invites students to learn syntax the way working linguists do—by asking questions, testing ideas, and building analyses step by step. Grounded in clear explanations and hands-on problem solving, the book helps readers “think like syntacticians” while enjoying the thrill of cracking real structural puzzles. Drawing constant parallels between English and Indonesian, it turns abstract theory into something concrete, local, and memorable. The journey runs from the basics of constituency and phrase structure through X-bar, movement, binding, and clausal architecture, with carefully scaffolded chapters that blend exposition, worked examples, and guided theory-building tasks. To keep learners moving, the book includes a step-by-step Tree-Drawing Guide, so students can visualize arguments with confidence, plus a Master Glossary & Quick Reference for rapid review at a glance. Whether you’re tackling your first tree or debating the finer points of ellipsis and empty categories, this is a companion text designed to be used, marked up, and returned to—an approachable, discovery-driven introduction that makes the logic of syntax click.



Ganesha
KreasiSemesta

 ganeshakreasisemesta@gmail.com
 www.ganeshakreasisemesta.com
 0852-8000-2192
 Anggota IKAPI
No. 281/JTE/2024

ISBN 978-634-7403-86-5



9

786347

403865

**ENGLISH SYNTAX ONSET:
A DISCOVERY-BASED INTRODUCTION
TO GENERATIVE SYNTAX**

**Yanuarius Yanu Dharmawan
Dini Hidayati**



PENERBIT PT. GANESHA KREASI SEMESTA

**ENGLISH SYNTAX ONSET:
A DISCOVERY-BASED INTRODUCTION
TO GENERATIVE SYNTAX**

Authors : Yanuarius Yanu Dharmawan
Dini Hidayati

Editors : Yanuarius Yanu Dharmawan
Dini Hidayati

Cover Design : Firman Isma'il

Layout : Retno Wilujeng

ISBN : 978-634-7403-86-5

Published by : **GANESHA KREASI SEMESTA,
OKTOBER 2025
MEMBER OF IKAPI CENTRAL JAVA
NO. 281/JTE/2024**

Editorial:

Panongan Street, Kutasari Village Baturraden District

Banyumas Regency Telp. 0852-8000-2192

Email : ganeshakreasisemesta@gmail.com

First Printing : 2025

All right reserved

Copyright protected by law

It is prohibited to reproduce or transfer part or all of the contents of this book in any form and by any means, including photocopying, recording, or other recording techniques without written permission from the publisher.

PREFACE

English Syntax Onset grew out of a deep and ongoing effort to make the study of syntax more approachable, relevant, and engaging for university students of English. Years of teaching generative grammar have shown us that students often connect most with the material when they can explore it actively, through problem-solving, drawing syntactic trees, and comparing structures between English and Indonesian. These experiences revealed a clear need for a resource that not only presents theory but also speaks directly to the context and learning style of our students. Drawing on the theoretical clarity of Carnie (2012), the minimalist perspective of Van Gelderen, and the practical orientation of Chung's *GB Syntax Problem Sets*, this book encourages readers not just to absorb information, but to think like syntacticians, developing the tools and habits of inquiry needed to explore language analytically. At the heart of this project, though, is enjoyment, the satisfaction of working through a tricky sentence, the sudden clarity when a tree structure makes sense, and the surprise of discovering how English and Indonesian differ and align. These were some of the most rewarding moments during both the writing and teaching process, and they form the spirit of this book. More than just a textbook, it is meant to be a companion on your learning journey, helping you navigate both the complexities and the elegance of syntactic theory. Syntax can be challenging, but it is also deeply systematic and, at times, wonderfully intuitive. Whether you're just starting with phrase structure rules or diving into more advanced topics like movement and binding, we hope this book provides not only understanding, but also a genuine sense of curiosity and satisfaction along the way.

**Yanuaris Yanu Dharmawan
Dini Hidayati**

TABLE OF CONTENTS

PREFACE.....	iii
TABLE OF CONTENTS	iv
PART I: FOUNDATIONS OF SYNTAX.....	1
UNIT 1 WHAT IS SYNTAX?.....	2
A. Syntax as a Cognitive Science	2
B. Descriptive vs Prescriptive Grammar	5
C. Syntax and the Scientific Method	7
D. Competence vs Performance	10
E. Universal Grammar and the Innateness Hypothesis	13
F. Summary	15
GLOSSARY	19
UNIT 2 CATEGORIES AND PARTS OF SPEECH.....	20
A. Lexical and Functional Categories	20
B. Open vs Closed Categories	22
C. Tests for Word Classes	25
D. Feature Systems	28
E. Summary	30
GLOSSARY	34
UNIT 3 CONSTITUENCY AND PHRASE STRUCTURE	35
A. What Is a Constituent?	35
B. Tests for Constituency	38
C. Phrase Structure Rules (PSRs)	41
D. Bracket Notation and Tree Diagrams	44
E. Structural Ambiguity	47
F. Summary	50
GLOSSARY	53
PART II: THE ARCHITECTURE OF SENTENCES	54
UNIT 4 X-BAR THEORY AND SENTENCE STRUCTURE	56
A. X-bar Schema and Its Components	56
B. Specifiers, Complements, and Adjuncts	59
C. From XP to CP: Layering the Structure	61
D. Drawing Trees with X-bar Notation	64
E. Summary	67
GLOSSARY	69

UNIT 5	STRUCTURAL RELATIONS IN SYNTAX.....	70
	A. Domination, Precedence, and C-command	70
	B. Syntactic Binding and Scope	73
	C. Representing Structural Relations in Trees	76
	D. Summary	80
	GLOSSARY	82
UNIT 6	CLAUSAL ARCHITECTURE	83
	A. TP, CP, and Functional Layers.....	83
	B. Sentence Types and Complementizers	86
	C. The Clause as a Syntactic Unit	89
	D. Summary	93
	GLOSSARY	94
PART III:	ARGUMENTATION IN SYNTAX.....	95
UNIT 7	AUXILIARIES AND STRUCTURAL EVIDENCE.....	96
	A. Auxiliaries as Heads: Evidence from Ellipsis	96
	B. Distribution and Ordering of Auxiliaries	99
	C. Morphological Restrictions and Verb Forms	102
	D. Summary	105
	GLOSSARY	107
UNIT 8	QUESTIONS AND MOVEMENT	108
	A. Yes-No Questions: Inversion vs. Base Generation	108
	B. WH-Movement: Subject and Object Questions	111
	C. Islands and Constraints on Movement.....	115
	D. Summary	119
	GLOSSARY	120
UNIT 9	TRANSFORMATIONS AND DERIVATIONS	121
	A. Do-Support and Verb Movement.....	121
	B. Head Movement and Affix Hopping.....	124
	C. Derivations and Deep Structure	127
	D. Summary	130
	GLOSSARY	132
PART IV:	DEEP STRUCTURE, CASE, AND BINDING	133
UNIT 10	THETA THEORY AND CASE.....	134
	A. Theta Roles and Argument Structure	134
	B. Thematic Hierarchies	137

C. Case Assignment and the Extended Projection Principle (EPP).....	141
D. Summary	144
GLOSSARY	146
UNIT 11 BINDING THEORY	147
A. Anaphors, Pronouns, and R-expressions	147
B. Principles A, B, and C	151
C. Locality and Binding Domains	154
D. Summary	157
GLOSSARY	159
PART V: SYNTAX IN ACTION.....	160
UNIT 12 ELLIPSIS AND EMPTY CATEGORIES	161
A. VP Ellipsis and Pseudogapping	161
B. Null Subjects and PRO	164
C. Control vs Raising	167
D. Summary	170
GLOSSARY	173
UNIT 13 APPLIED PROBLEM SOLVING IN SYNTAX	174
A. GB Syntax Problems: Assignments and Commentary	174
B. Derivational Analyses and Tree Drawing	177
C. Argumentation and Hypothesis Testing.....	180
D. Summary	183
GLOSSARY	185
UNIT 14 LANGUAGE VARIATION AND UNIVERSAL GRAMMAR.....	186
A. Syntactic Parameters: Word Order, Null Subjects	186
B. Cross-linguistic Syntax: Indonesian Case Only	189
C. Implications for Language Acquisition.....	192
D. Summary	196
GLOSSARY	198
REFERENCES.....	199
APPENDICES.....	205
AUTHOR'S BIOGRAPHY	220



PART I:
FOUNDATIONS OF SYNTAX



UNIT |

1

WHAT IS SYNTAX?

Syntax is the branch of linguistics that studies how words combine to form grammatical sentences. It investigates the set of principles and rules that govern sentence structure in natural languages. While the concept may appear simple at first glance, the discipline opens a window into how the human mind processes and generates language. Every time we utter a sentence, we are unconsciously applying rules we may not even be aware of. Syntax attempts to uncover those hidden mental rules and explain how language users can produce and understand an infinite number of sentences, many of which they have never heard before.

A. Syntax as a Cognitive Science

One of the foundational insights of modern linguistics, especially in the generative tradition, is the recognition that language is a cognitive system, a mental faculty that forms a core part of the human mind. This means that when we study syntax, we are not simply cataloging sentence patterns or correcting grammar mistakes; rather, we are trying to understand a deep and abstract system that is internalized by all speakers of a language. This view positions syntax within the broader interdisciplinary field known as cognitive science, which includes psychology, neuroscience, artificial intelligence, philosophy, and linguistics. The goal of cognitive science is to explore and explain the nature of human intelligence, and language, particularly syntax, plays a central role in this inquiry.

Instead of thinking of language as a long list of memorized phrases and expressions, the generative approach argues that humans possess an internal grammar, sometimes called an “I-language” (Chomsky, 1986), which allows them to construct and interpret entirely new sentences. For instance, you have probably never heard the sentence: “The green elephant danced under the upside-down umbrella on Mars,” and yet, you can easily understand it. This ability to comprehend novel combinations of familiar words is strong evidence that language is not learned through repetition alone, but rather generated through mental rules and principles. These rules are not explicitly taught; they are unconsciously acquired and applied by native speakers from early childhood.

The idea that language is both creative and rule-governed is central to the generative framework. Noam Chomsky, widely considered the founder of modern theoretical linguistics, introduced the concept of Universal Grammar, the idea that all human languages share a set of basic principles due to our shared biological makeup. In his influential works such as *Aspects of the Theory of Syntax* (1965) and *Lectures on Government and Binding* (1993), Chomsky argued that children are not merely passive recipients of linguistic input. Instead, they come equipped with a mental blueprint for language, which guides and constrains how they interpret and produce linguistic data. This blueprint explains why language acquisition in young children is so rapid, consistent, and robust across cultures.

Syntax plays a critical role in this theory because it deals with the structural aspect of language, the way words are arranged into phrases and sentences. Syntax is not concerned with the meanings of individual words (which falls under semantics), but with how those words are combined into larger units according to systematic rules. For example, native speakers know that “The cat chased the dog” and “The dog chased the cat” are both grammatical, but mean very different things, despite containing the same words. The difference in meaning

arises entirely from the syntax, or word order, which tells us who is doing what to whom.

By treating syntax as a branch of cognitive science, linguists can apply scientific methods to investigate it. This involves forming hypotheses about how sentence structures work, testing those hypotheses with examples (both grammatical and ungrammatical), and refining the theories based on new data. For instance, if we hypothesize that English always requires an overt subject, we can test this by examining imperative sentences like “Run!” or “Don’t panic!” These examples appear to lack subjects, which may lead us to revise our hypothesis and conclude that subjects in imperatives are implicitly understood rather than overtly stated. This type of hypothesis testing aligns syntax with the broader scientific method and underlines its cognitive significance.

Moreover, syntax is not only a mental capacity but also appears to be linked to specific regions of the brain. Studies in neurolinguistics have shown that certain areas, such as Broca’s area in the left hemisphere, are heavily involved in syntactic processing. Individuals who suffer damage to this area often experience difficulties constructing or understanding complex sentences, a condition known as agrammatism. This provides further evidence that syntax is a neurologically distinct component of language, with its own dedicated mechanisms in the brain (Friederici, 2002).

The relevance of syntax extends even into artificial intelligence. When developers try to create natural language processing systems or chatbots, one of the hardest challenges is replicating the kind of syntactic understanding that humans perform effortlessly. For machines to understand or generate human-like sentences, they must be equipped with models that reflect human syntactic knowledge. This connection between syntax and computation supports the view that syntax is not only a mental but also a computational system, a formal system of rules that generates well-formed expressions.

To summarize, the study of syntax as a cognitive science allows us to understand language not as a superficial set of rules, but as a deep and dynamic mental capacity. Syntax is part of our biological endowment, shaped by Universal Grammar and realized in the minds and brains of individual speakers. It enables the infinite creativity of human language and connects us to other areas of scientific inquiry, from psychology to neuroscience and artificial intelligence. Understanding syntax, therefore, is essential for understanding not just language, but also the human mind.

B. Descriptive vs Prescriptive Grammar

One of the most important foundational distinctions that students of syntax must grasp early on is the difference between prescriptive grammar and descriptive grammar. These two approaches to language represent very different goals and attitudes, and understanding them helps clarify what syntactic study is really about.

Prescriptive grammar is the traditional type of grammar that many people first encounter in school. It refers to a set of rules about how a language “should” be used, often based on historical usage, literary tradition, or institutional authority. Prescriptive rules are typically found in language handbooks, dictionaries, and style guides. For example, prescriptive grammarians might insist that a sentence should never end with a preposition or that one must always use “whom” instead of “who” in certain formal contexts. These rules are not necessarily based on how people naturally speak, but rather on how they are taught to speak, often with the goal of maintaining a particular standard of correctness or formality.

In contrast, descriptive grammar is the approach that linguists and syntacticians use. Rather than telling people how they should speak, descriptive grammar aims to observe and explain how people actually do speak. It is based on empirical data collected from real-world language use, both spoken and written. Linguists who adopt a descriptive perspective are not

concerned with judging language as right or wrong. Instead, they ask questions like: What are the underlying patterns in the way people construct sentences? What rules do native speakers seem to follow unconsciously? Are there systematic differences in how language is used in different regions, communities, or contexts?

To illustrate the contrast, consider the following example from English:

1. Prescriptive: "To whom did you send the letter?"
2. Descriptive: "Who did you send the letter to?"

According to traditional prescriptive grammar, the first sentence is considered more correct because it avoids ending with a preposition ("to"). However, the second sentence is far more common in natural spoken English. Native speakers frequently use this form, and it is fully grammatical from a syntactic perspective. In fact, attempting to avoid sentence-final prepositions in everyday conversation often results in awkward or unnatural phrasing. The descriptive linguist accepts the second sentence as valid because it reflects actual language use.

This same distinction appears in other languages, including Indonesian. Compare the following two versions of a similar question in Indonesian:

1. Prescriptive: "Kepada siapa kamu mengirim surat itu?"
2. Descriptive: "Kamu kirim surat itu ke siapa?"

In formal settings or in written texts, the first sentence might be preferred because it follows traditional Indonesian grammar taught in schools. However, in casual conversation, native speakers are more likely to use the second version. Just like in English, the prepositional phrase ("ke siapa" or "to whom") naturally moves to the end in informal speech. This demonstrates that, across languages, speakers often prioritize fluency, clarity, and efficiency over formal rules, especially in spontaneous, real-time communication.

It is important to understand that both prescriptive and descriptive grammar have value, but they serve different purposes. Prescriptive grammar is useful for standardized writing, education, and professional communication, where consistency and clarity are especially important. Descriptive grammar, however, is essential for scientific inquiry. It helps linguists build theories about how language functions, how it varies across speakers and contexts, and how it is acquired by children and learners.

From a syntactic point of view, descriptive grammar is the starting point for all analysis. Syntax does not prescribe how you should speak or write. Instead, it seeks to uncover the mental rules and structures that native speakers already know and use intuitively. These rules are often complex and subtle, and they may differ from the prescriptive norms taught in school. For example, while a prescriptive rule may ban double negatives in English (e.g., “I don’t know nothing”), a descriptive analysis will explore why such structures are common in some dialects, what they mean, and how they fit into the broader grammar of those speech communities.

In this book, we take a descriptive approach. We focus on how syntax operates in real language use, not on correcting so-called “errors.” Our goal is to understand the logic behind sentence structures and to appreciate the diversity and richness of human language.

C. Syntax and the Scientific Method

Syntax, like all areas of modern linguistics, is grounded in the principles of the scientific method. It is not simply a collection of language observations or grammatical trivia. Rather, syntax functions as a rigorous and systematic investigation into the structure of human language, based on the formulation and testing of hypotheses. Linguists who study syntax begin by observing actual language use, whether in spoken dialogue, written texts, or even constructed examples, and from these

observations, they develop hypotheses about the underlying rules that govern sentence formation.

The process typically begins with data collection, where linguists examine both naturally occurring utterances and intuitively constructed examples. These examples serve as evidence for possible rules in the grammar of a language. For instance, if speakers consistently use a particular word order, or avoid certain structures, this tells the syntactician something about the constraints or preferences encoded in the mental grammar. The next step is hypothesis formation: the linguist proposes a rule or generalization that explains the observed data. Crucially, this hypothesis must be testable. That is, it must make predictions about what structures should or should not be grammatical.

An essential feature of any scientific theory is falsifiability, the idea that a hypothesis should be disprovable under certain conditions. If a grammatical rule claims that all sentences must contain a visible subject, then a counterexample that contradicts this rule would be important for refining or rejecting the hypothesis. Consider the following hypothesis: "Every sentence in English must have an overt subject." At first glance, this might seem reasonable, especially since English is considered a non-null subject language. However, when we look at imperative constructions like "Run!" or "Don't move!", we encounter a challenge. These sentences are grammatical despite the absence of an explicitly stated subject. This suggests that the hypothesis is too strong or incomplete.

To resolve this, syntacticians might propose that the subject is still present at an abstract level, even if it is not pronounced. In the case of "Run!", the subject is understood to be "you," as in "[You] run!". This leads to a refinement of the original hypothesis: perhaps all English sentences have a subject at the level of syntactic structure, but it may be phonologically null in certain environments. This kind of refinement is characteristic of scientific inquiry, it reflects how theories evolve in response to new evidence.

The same pattern occurs in other languages. For instance, Indonesian also allows imperative sentences without overt subjects. A speaker might say “Lari!” which translates as “Run!” in English. There is no stated subject, but the meaning is clearly directed at the second person (“kamu,” meaning “you”). Just like in English, the subject is understood rather than spoken. This cross-linguistic similarity provides further support for the idea that syntactic rules are both language-specific and, in some respects, universal. It also illustrates how comparing data across languages can help linguists test and refine universal claims about grammar.

In some cases, hypothesis testing can even reveal unexpected aspects of grammar. For example, in English, the position of auxiliary verbs in questions and negations shows that syntactic rules must account for movement, words do not always appear where they are first generated. If we hypothesize that words remain in fixed positions, we are quickly forced to revise that view in light of constructions like “Did she leave?” or “She does not sing.” The scientific method compels linguists to revise their assumptions, test alternative explanations, and build more comprehensive models of syntax.

The use of constructed examples (sometimes called “intuition-based data”) is also a hallmark of syntactic science. Because native speakers can usually judge the grammaticality of a sentence without much effort, their intuitions serve as a rich source of evidence. However, linguists are careful to supplement these judgments with data from corpora, large collections of real-world language use, and from typological studies across many languages. This triangulation of evidence helps ensure that theories are not only internally consistent but also empirically grounded.

Ultimately, what makes syntax a scientific discipline is not just the use of data, but the logic of explanation and prediction. A syntactic theory must not only describe which sentences are grammatical but also explain why they are, how they are derived, and what constraints shape their structure. This

includes explaining ungrammatical sentences, predicting when certain constructions will be ambiguous, and accounting for cross-linguistic variation. In this way, syntax contributes to our understanding of the cognitive mechanisms that underlie language.

The scientific approach to syntax has been a major focus of generative grammar since its inception. Works such as *Syntax: A Generative Introduction* by Andrew Carnie (2012) provide a clear overview of this methodology, emphasizing hypothesis formation and testing as key aspects of syntactic analysis. Likewise, Andrew Radford's *Analyzing English Sentences* (2009) encourages students to think like scientists, identifying patterns, forming generalizations, and evaluating evidence from multiple angles.

This perspective also prepares students for further studies in psycholinguistics, computational linguistics, and language acquisition, where the scientific modeling of syntax intersects with experimental and applied research. By learning to think scientifically about language, students not only develop analytical skills but also deepen their appreciation for the complexity and elegance of human communication.

D. Competence vs Performance

One of the most influential distinctions in modern linguistics was introduced by Noam Chomsky in his landmark work *Aspects of the Theory of Syntax* (1965). This is the distinction between competence and performance, two terms that refer to fundamentally different aspects of linguistic behavior. Understanding this distinction is essential for anyone studying syntax, because it clarifies the scope and focus of syntactic theory, as well as the limits of linguistic data obtained from everyday speech.

Competence refers to a speaker's internalized knowledge of their language. It is the mental grammar that native speakers unconsciously acquire during childhood and that enables them to produce and understand an infinite number of sentences, even

those they have never encountered before. This knowledge includes understanding sentence structure, word order, agreement, movement, and many other syntactic operations. Competence is idealized in the sense that it is abstracted away from external factors. It represents what a speaker knows about the language under optimal conditions, unaffected by distractions or errors.

Performance, on the other hand, is the actual use of language in real-life situations. It includes everything that occurs when someone speaks or writes, listens or reads. Unlike competence, performance is subject to a wide range of influences, such as attention span, memory limitations, fatigue, stress, distractions, and social pressures. These factors can result in speech errors, hesitations, repetitions, false starts, or incomplete sentences. Importantly, such imperfections do not reflect a lack of grammatical knowledge; rather, they illustrate the limits of human cognitive processing in real time.

For example, consider the following complex English sentence: "The cat the dog the girl saw chased was black." Structurally, this sentence is grammatical. Each noun phrase and verb phrase is in the correct position, and the sentence follows the rules of English syntax. However, most people find this sentence very difficult to process and understand, especially when hearing it for the first time. The reason is that it contains multiple nested relative clauses, which place a heavy burden on short-term memory and syntactic parsing. Despite this difficulty, native speakers can recognize that the sentence is well-formed, they may simply prefer to rephrase it for clarity. This demonstrates the distinction between competence (recognizing grammatical structure) and performance (struggling with complex processing).

Chomsky emphasized that linguistic theory, especially in the generative tradition, should focus primarily on competence. The goal is to uncover the underlying rules and structures that make language possible. By abstracting away from the imperfections of performance, linguists can build a model of the

mental grammar that all speakers share. This model can then explain the creativity of language, the fact that speakers can generate and interpret countless sentences, including ones they have never seen before.

This approach has been widely adopted in formal syntax, especially in frameworks like Government and Binding Theory and the Minimalist Program. These theories aim to describe the core grammatical principles that govern sentence structure in all languages. At the same time, linguists acknowledge that performance data, such as disfluencies or errors, can provide valuable insights, especially in fields like psycholinguistics, which studies how language is processed in real time.

The competence-performance distinction has also influenced research in second language acquisition, where learners' performance often lags behind their competence. For example, a student learning English might know the correct word order for questions (e.g., "What are you doing?") but still produce incorrect forms under pressure (e.g., "What you are doing?"). This kind of error does not mean the learner lacks grammatical knowledge; it may simply reflect processing limitations, anxiety, or lack of fluency.

While Chomsky's original focus was on idealized native speakers, later work has expanded the concept of competence to include sociolinguistic variation and pragmatic knowledge. Researchers such as Hymes (1972) introduced the idea of "communicative competence," which includes the ability to use language appropriately in different social contexts. However, in syntactic theory, the term "competence" usually refers specifically to grammatical competence, the knowledge of syntax, morphology, and related structures.

Understanding the competence-performance distinction allows students to see why syntactic theories often rely on constructed examples and grammaticality judgments. These tools are designed to reveal the structure of mental grammar, not to document every feature of spontaneous speech. Recognizing

this helps students interpret syntactic data more critically and appreciate the theoretical goals of formal syntax.

E. Universal Grammar and the Innateness Hypothesis

One of the most influential and, at times, controversial ideas in modern linguistic theory is the notion of Universal Grammar (UG). The concept of UG was developed and popularized by Noam Chomsky, particularly in works such as *Lectures on Government and Binding* (1993) and *Knowledge of Language* (1986). According to this theory, all human beings are born with a biologically endowed language faculty, a specialized mental system that provides the structural foundation for acquiring any natural language. This innate system, referred to as Universal Grammar, consists of a set of general principles and parameters that are shared across all human languages, despite the observable diversity among them.

The core motivation behind Universal Grammar arises from a set of observations about how children acquire language. Across cultures and environments, children appear to acquire their native language quickly, uniformly, and without formal instruction. By the age of five or six, most children have mastered the basic grammar of their language, including complex syntactic structures that they were never explicitly taught. This rapid and robust acquisition process has led researchers to propose that children do not learn language solely through imitation or trial and error. Instead, they must be equipped with an internal system that guides them in interpreting linguistic input and constructing grammatical rules.

This idea is often supported by what is known as the Poverty of the Stimulus argument (Chomsky, 1980; Pinker, 1994). This argument states that the linguistic input children receive, often called the primary linguistic data, is too limited, too noisy, and too ambiguous to fully account for the complexity of the grammar they eventually acquire. For example, children are rarely exposed to negative evidence (i.e., explicit correction of ungrammatical sentences), and yet they develop grammars that

avoid those errors. Additionally, many syntactic structures are underrepresented in speech directed at children, and yet children demonstrate clear knowledge of them. The logical conclusion, proponents argue, is that some aspects of grammatical knowledge must be pre-specified by our biology.

Universal Grammar is thought to consist of two components: principles and parameters. Principles are grammatical rules that are true of all languages. For example, all languages distinguish between subjects and predicates, and all have mechanisms for forming questions or negatives. Parameters, on the other hand, are points of variation. A parameter is like a switch that can be set differently depending on the language. One well-known parameter is the head-directionality parameter, which determines whether the head of a phrase precedes or follows its complements. In English, which is a head-initial language, the verb typically comes before its object (e.g., “eat apples”), while in Japanese, a head-final language, the verb comes after the object (e.g., “ringo o taberu” – ‘apples eat’).

Even between typologically different languages like English and Indonesian, we find evidence of shared structural properties, supporting the idea of a common underlying grammar. For instance, both English and Indonesian follow a basic Subject-Verb-Object (SVO) word order in declarative sentences:

1. English: “The rabbit nibbled the lettuce.”
2. Indonesian: “Kelinci itu memakan selada.”

Although Indonesian uses affixes extensively for grammatical functions, such as the prefix “meN-” to indicate active voice (“memakan” = ‘eat’), the core syntactic structure remains comparable to English. This suggests that both languages instantiate the same basic word order parameter, even though they differ in morphological complexity and phonological realization. Such structural parallels across unrelated languages strengthen the argument that all languages are manifestations of a shared set of cognitive principles.

The theory of Universal Grammar has far-reaching implications beyond first language acquisition. In the field of second language acquisition (SLA), researchers such as White (2012) and Schwartz & Sprouse (1996) have explored whether adult learners still have access to UG and how it might influence their ability to acquire a second language. Some scholars argue that second language learners draw on their first language's parameter settings, which can either facilitate or hinder acquisition, depending on how similar the two languages are.

Despite its strong empirical grounding, Universal Grammar is not without its critics. Alternative approaches, such as usage-based theories and connectionist models, argue that language acquisition can be explained through domain-general cognitive mechanisms, such as pattern recognition, statistical learning, and social interaction. Nevertheless, UG remains a foundational theory in formal syntax and linguistic theory more broadly. Its central claim, that human beings are born with a predisposition to acquire language, is supported by decades of cross-linguistic, developmental, and neurological evidence.

For students of syntax, the idea of Universal Grammar provides an important framework for understanding why languages share certain features and why learners acquire language so effortlessly during early childhood. It also helps explain how syntactic theory can aim to describe not just English or Indonesian, but the universal properties of all human languages.

F. Summary

This unit has introduced students to the foundational concepts of syntax and its role within the broader study of human language. Syntax is the scientific study of how words are arranged to form meaningful and grammatically correct sentences. Rather than viewing grammar as a static list of rules, syntax treats sentence structure as a dynamic system that is embedded in the human mind. It aims to uncover the rules and principles that allow speakers to produce and understand an

infinite variety of sentences. As such, syntax is not just a tool for editing written language; it is a window into the cognitive processes that underlie linguistic ability.

We have seen that syntax belongs to the interdisciplinary field of cognitive science. This is because syntax deals with mental representations of language, what Chomsky (1986) refers to as “I-language” or internalized language. This mental grammar is what speakers acquire in early childhood, usually without formal instruction, and it is what enables them to interpret and generate new sentences effortlessly. Cognitive scientists, linguists, and psychologists study syntax to understand the mental mechanisms that enable this linguistic creativity, as well as how language interfaces with memory, perception, and meaning. Theoretical linguists like Radford (2012), Carnie (2012), and Adger (2003) have consistently emphasized that syntactic rules are not simply surface conventions but reflect deep, abstract mental structures shared by all speakers of a language.

A central feature of syntactic inquiry is its reliance on the scientific method. Just like researchers in physics or biology, syntacticians collect data, form hypotheses, test predictions, and revise their theories based on new evidence. For example, when linguists encounter data that challenge their assumptions, such as imperative sentences that lack overt subjects, they do not reject the data as anomalies. Instead, they revise their hypotheses to accommodate these observations. The distinction between grammaticality (what speakers intuitively know) and acceptability (what they actually produce under real-world conditions) helps maintain the scientific integrity of syntactic theory. Scholars like Chomsky (1965), Radford (2009), and Haegeman (1994) have all contributed to refining this empirical methodology in syntax.

Another important theme covered in this unit is the distinction between prescriptive and descriptive grammar. While prescriptive grammar dictates how people “should” speak based on social or educational norms, descriptive grammar

focuses on how people actually use language in real contexts. Syntax is a descriptive science, it seeks to analyze language patterns without making judgments about correctness. For instance, while prescriptive grammarians might object to sentences like “Who did you send the letter to?”, syntacticians recognize such constructions as fully grammatical based on real-world usage. This distinction is not merely philosophical, it shapes how we collect data, evaluate grammaticality judgments, and understand linguistic variation. Authors like Aarts (2008), Van Gelderen (2002), and Pullum & Huddleston (2012) have strongly advocated for the importance of descriptive accuracy in grammatical research.

We also explored Chomsky’s competence versus performance distinction, which helps explain the gap between what speakers know and how they use that knowledge in practice. Competence refers to the idealized internal grammar that allows for infinite sentence generation, while performance includes all the real-world limitations, such as memory constraints, hesitation, and slips of the tongue. This distinction allows linguists to abstract away from performance-related errors and focus on the underlying grammar that guides sentence formation. Understanding this distinction is crucial for interpreting linguistic data and evaluating syntactic theories (Chomsky, 1965; O’Grady et al., 2009).

Finally, the concept of Universal Grammar (UG) was introduced as the theoretical framework that explains how children can acquire complex grammatical systems so quickly and with relatively little input. UG consists of a set of innate principles and adjustable parameters that guide the acquisition of any human language. The similarities in word order between English and Indonesian, despite their many surface-level differences, suggest that both languages adhere to a shared set of syntactic constraints. The theory of UG provides a compelling explanation for the uniformity of language acquisition across cultures and has been supported by research in first and second language acquisition, language typology, and neurolinguistics

(Chomsky, 1993; Cook & Newson, 2014; Pinker, 1994; White, 2012).

Altogether, this unit establishes the conceptual foundation for the rest of the book. Understanding syntax means more than learning terminology or memorizing sentence patterns, it means understanding how language operates as a mental system, how it reflects cognitive processes, and how it varies across individuals and communities. The insights from generative syntax, coupled with empirical data from diverse languages, allow us to approach language as both a uniquely human faculty and a scientifically investigable phenomenon.

GLOSSARY

- **Competence** – Idealized mental knowledge of grammar.
- **E-language** – Externalized language (texts, utterances).
- **Generativity** – Finite rules yield infinitely many sentences.
- **Grammar (Descriptive)** – Rules speakers actually use; target of linguistic analysis.
- **Grammar (Prescriptive)** – Norms about how people “should” use language.
- **I-language** – Internal, mental grammar of a speaker.
- **Performance** – Actual language use with errors/limitations.
- **Scientific Method in Syntax** – Hypothesis, data, prediction, testing.
- **Syntax** – Study of how words combine into phrases and sentences.
- **Universal Grammar (UG)** – Innate principles guiding language acquisition.

UNIT 2

CATEGORIES AND PARTS OF SPEECH

Understanding the structure of sentences begins with understanding the fundamental building blocks that make up those sentences, namely, the categories of words we use and how they behave. In linguistic terms, these categories are known as parts of speech or syntactic categories. Every word in a sentence belongs to a specific category that determines how it functions within the sentence and how it interacts with other words. In syntactic analysis, these categories are usually divided into two broad types: lexical categories and functional categories.

A. Lexical and Functional Categories

Lexical categories include the kinds of words that carry rich semantic content. These are typically the words that name things, describe actions, or provide descriptive qualities. The core lexical categories in most languages are nouns (N), verbs (V), adjectives (Adj), and adverbs (Adv). For instance, in English, the word *planet* is a noun, *dance* is a verb, *lovely* is an adjective, and *quickly* is an adverb. These words refer to objects, events, properties, and manners of action, and they form the content-heavy part of the sentence.

In Indonesian, we find corresponding examples for these lexical categories: *planet* (noun), *menari* (verb), *indah* (adjective), and *cepat* (adverb). Just like in English, these words in Indonesian serve to provide the core meaning of a sentence. Although the languages differ in morphology and syntax, they both rely on

these content-bearing words to structure thoughts and communicate ideas. The lexical category of a word determines what kinds of grammatical roles it can play, such as subject, predicate, or modifier.

In contrast to lexical categories, functional categories include words that do not carry much meaning by themselves, but instead serve important grammatical roles. They help build sentence structure by expressing relations between content words. Functional categories include determiners (Det), such as *the* in English or *itu* in Indonesian; auxiliaries (Aux), such as *will* in English or *akan* in Indonesian; and complementizers (C), such as *that* in English or *bahwa* in Indonesian. These words are crucial to syntax because they mark distinctions such as tense, aspect, mood, definiteness, and clause type.

The difference between lexical and functional categories can also be observed through linguistic tests, such as how a word behaves morphologically or what types of phrases it can combine with. For example, lexical categories can usually be modified or inflected: nouns can be pluralized (*planet* → *planets*), verbs can be conjugated (*dance* → *danced*), and adjectives can be compared (*lovely* → *lovelier*). Functional categories, however, do not undergo these changes as readily. We do not pluralize *the*, or conjugate *that*, because their function is structural rather than referential or descriptive.

In syntax, it is essential to distinguish between these two types of categories because they play different roles in phrase structure. Lexical categories typically form the head of phrases, meaning they are the core elements that determine the type of phrase (e.g., noun phrases are centered around a noun). Functional categories typically appear in specifier or higher functional positions, helping to structure the phrase by specifying features such as definiteness, tense, or subordination.

For example, in the English noun phrase *the red book*, the noun *book* is the head of the phrase (a lexical category), while *the* is a determiner (a functional category) that introduces the noun and marks it as definite. Similarly, in the Indonesian noun phrase

buku merah itu, the head noun is *buku* ('book'), the adjective *merah* ('red') modifies it, and the determiner *itu* ('that') functions to mark definiteness, occurring after the noun and modifier, in accordance with Indonesian syntax.

Scholars such as Radford (2012), Van Gelderen (2002), and Carnie (2012) emphasize the central importance of recognizing these category distinctions in syntactic analysis. For instance, Radford explains that lexical categories are often open-class, meaning new words can be added relatively easily (e.g., new nouns like *selfie* or verbs like *google*), while functional categories are typically closed-class, meaning they rarely admit new members. This closed nature reflects the grammatical stability of the functional system in a language.

Cross-linguistically, these distinctions also support the idea of universal structural tendencies. Despite surface-level differences in word order, morphology, and pronunciation, nearly all natural languages appear to differentiate between content words and grammatical words, a point widely supported in typological studies (Baker, 2003; Payne, 2006).

In sum, understanding lexical and functional categories is essential for the study of syntax because these categories are the foundational units from which phrases and sentences are built. Lexical categories give a sentence its semantic content, while functional categories provide the grammatical scaffolding that allows that content to be interpreted within a structured context. Recognizing the different behaviors and functions of these categories enables students to analyze sentence structures more effectively and prepares them to engage with more complex syntactic theories in later units.

B. Open vs Closed Categories

Another important distinction in syntactic categorization is between what linguists call open word classes and closed word classes. These terms refer to how flexible a particular category is in terms of admitting new members. This distinction plays a crucial role in understanding how languages evolve, how

vocabulary expands, and how grammatical structures remain relatively stable over time.

Open classes refer to categories that are constantly expanding. These include the major lexical categories: nouns, verbs, adjectives, and adverbs. These are the content-bearing parts of language, and they are open to innovation, borrowing, and coinage. Because language must be able to accommodate new ideas, objects, technologies, and cultural trends, open classes are flexible and productive. For example, in recent decades, English has added new nouns like *selfie*, *hashtag*, *vlog*, and *glamping*, as well as new verbs like *google*, *ghost*, and *doomscroll*. These words were either created from existing language material or borrowed and adapted from other languages or technological discourse.

Indonesian also demonstrates this tendency. New nouns such as *gedrul* (a playful or invented term that could hypothetically enter slang or popular culture) or borrowed terms like *layangan* (drone) and *gadget* have become common in youth vernacular. Likewise, verbs such as *ngeprint* (from the English *print*) and *ngegas* (from *gas*) illustrate how Indonesian adapts to changing communicative needs by incorporating and localizing foreign words. According to Sneddon et al. (2010), Indonesian is particularly receptive to loanwords and productive derivational processes, especially with affixation that adapts foreign roots into native morphological patterns.

The reason open classes are so adaptable is that their function is to refer to the changing world around us. New technologies, cultural phenomena, and trends demand new ways of describing them, and speakers are naturally inventive in this regard. From a linguistic perspective, the productivity of open classes ensures that the lexicon can continue to meet the expressive needs of the speech community.

In contrast, closed classes are much more resistant to change. These categories include function words such as pronouns (e.g., *he*, *she*, *it*), prepositions (e.g., *in*, *on*, *under*), conjunctions (e.g., *and*, *but*, *because*), determiners (e.g., *the*, *this*,

that), and auxiliary verbs (e.g., *is*, *will*, *must*). Unlike open-class items, closed-class words do not carry rich conceptual content on their own. Instead, their job is to structure grammatical relationships between words and phrases. Because their meanings are abstract and their roles are highly grammaticalized, these words tend to resist innovation. As Huddleston and Pullum (2012) explain, closed-class words are part of the “scaffolding” of grammar, they enable syntactic constructions but are not themselves descriptive of the world.

For example, in English, it is extremely rare to create a new pronoun. Although there have been recent attempts to introduce gender-neutral singular pronouns such as *ze* or *xe* to supplement or replace *he* and *she*, these innovations face considerable resistance and have only gained limited adoption in mainstream usage. Even when such pronouns are introduced for social reasons, they highlight the general rigidity of the pronoun system. Similarly, new prepositions are virtually unheard of in modern English. We do not often hear entirely new spatial or temporal prepositions entering the language, even though the vocabulary of nouns and verbs continues to grow rapidly.

Indonesian exhibits the same pattern. Prepositions such as *di* (in/at), *ke* (to), and *dari* (from) are core members of the closed-class system and have remained stable for generations. Although new compounds or multi-word expressions may arise (e.g., *ke arah* meaning “towards”), the class of single-word prepositions does not readily accept new members. Likewise, Indonesian pronouns such as *saya* (I), *kamu* (you), *dia* (he/she), and *mereka* (they) form a relatively fixed system. Innovations, when they do occur, are often driven by sociolinguistic change rather than syntactic necessity. For example, regional variants or honorifics like *aku*, *engkau*, or *beliau* may vary in usage but do not represent true lexical additions to the class.

From a theoretical perspective, the open/closed class distinction supports the idea that languages are organized around two complementary systems: a lexical system that is dynamic and meaning-rich, and a grammatical system that is

stable and rule-governed. This distinction is supported by cross-linguistic studies as well. Payne (2006) and Baker (2003) both highlight the relative universality of open vs. closed classes in natural languages, arguing that while the specific contents of these classes may vary across languages, the distinction itself is a basic organizing principle of linguistic structure.

Understanding which categories are open and which are closed is essential for syntactic analysis. When constructing phrase structure rules, for example, we must account for the fact that the heads of lexical phrases (e.g., nouns in noun phrases, verbs in verb phrases) are highly variable, while the functional elements that surround them (e.g., determiners, auxiliaries) come from a small, well-defined inventory. This helps us understand why syntax is both flexible and constrained: it allows for expressive diversity through open-class words, but it maintains structural consistency through closed-class elements.

C. Tests for Word Classes

One of the foundational tasks in syntactic analysis is determining what category a word belongs to, that is, whether it is functioning as a noun, verb, adjective, adverb, or something else in a sentence. While native speakers usually "just know" intuitively how words are used, syntacticians must rely on explicit criteria to make such categorizations systematic and testable. This is especially important when analyzing unfamiliar words, novel coinages, or words in languages with different grammatical structures. To do this, linguists use several well-established diagnostic tests. These include distributional tests, morphological behavior, and substitution tests. Each of these diagnostics helps reveal a word's syntactic category by examining how it behaves in relation to other words and within various syntactic contexts.

One of the most powerful and widely used methods is the distributional test. This approach involves observing the environments in which a word appears, what comes before and after it, and comparing this with known patterns. For instance, if

a new or unfamiliar word appears in a position typically occupied by a noun, we can reasonably conclude that the word belongs to the noun category. Consider the invented English sentence, “The zindle was glowing.” Even though *zindle* is not a real English word, it appears in the same syntactic slot as common nouns like *cat*, *child*, or *machine*, it is preceded by the definite article *the* and serves as the subject of the verb *was glowing*. Based on this syntactic distribution, *zindle* behaves like a noun.

A similar test can be applied in Indonesian: “Zindle itu bersinar.” Here, *zindle* is followed by the determiner *itu* (that) and acts as the subject of the verb *bersinar* (to shine). In Indonesian, determiners typically follow nouns, unlike in English where they precede them. Despite this difference, the distributional pattern confirms that *zindle* is functioning as a noun in both cases. As Radford (2012) and Van Gelderen (2002) note, distributional evidence is essential in determining syntactic category across languages because word classes are not determined by meaning alone but by position and function within phrases.

Another important diagnostic is morphological behavior, which focuses on a word’s ability to take on specific grammatical markers or affixes. In English, verbs typically accept tense and aspect markers such as *-ed* for past tense (e.g., *jump* → *jumped*) or *-ing* for progressive aspect (*run* → *running*). If a word can appear in these inflected forms, it is likely a verb. For example, if the word *tarf* can be used as *tarfed* or *tarfing*, we classify it as a verb based on its morphological potential.

Indonesian morphology operates differently, but similar principles apply. In Indonesian, verbs are commonly formed or marked using affixes such as *ber-*, *meN-*, or *di-*. For example, *lari* (run) becomes *berlari* when prefixed, indicating a verb. This process is part of what Sneddon et al. (2010) describe as a highly productive system of derivational and inflectional morphology. A word that accepts such verbal prefixes is likely to belong to the

verb class. Similarly, *tulis* (write) becomes *menulis* (to write), where *meN-* confirms the word's verbal status.

The substitution test provides another useful method for identifying word classes. This test checks whether a word can be replaced by another word that is known to belong to a specific category, without changing the grammatical structure of the sentence. For instance, in the English sentence "The fox ran," we can replace *the fox* with *it*, a known pronoun, to form "It ran." Since *it* is a pronoun that replaces noun phrases, this test confirms that *the fox* is a noun phrase headed by a noun. This method is especially effective in verifying the categorial status of noun phrases, adjectives, and adverbs, as they are often replaceable by other words with the same syntactic role.

In Indonesian, the same substitution logic applies. Take the sentence "Rusa itu lari" ("That deer ran"). The noun phrase *rusa itu* (the deer) can be replaced by the pronoun *ia* (he/she/they), resulting in "Ia lari." This shows that *rusa itu* occupies the subject noun phrase position and confirms its function as a noun phrase. Substitution also works for other categories. For example, an adjective like *besar* (big) can replace another adjective like *tinggi* (tall) in "Gedung itu tinggi" ("The building is tall") → "Gedung itu besar" ("The building is big"), showing that both are adjectives.

While each test provides valuable evidence, it is important to note that no single test is infallible. For instance, some words may exhibit hybrid behaviors, or may belong to more than one category depending on context. Words like *fast* in English can function both as an adjective (*a fast car*) and an adverb (*he ran fast*), demonstrating the importance of contextual analysis. Therefore, syntacticians often apply multiple diagnostics together, combining distributional, morphological, and substitutional evidence to arrive at a more reliable classification.

Linguists such as Baker (2003), Adger (2003), and Carnie (2012) emphasize the value of these diagnostics not only in classifying familiar words, but also in analyzing new lexical items, unknown constructions, and second-language learner

data. These tests are especially powerful in generative syntax, where they help determine the underlying structure of sentences and the rules that generate them.

By applying these empirical methods, students develop the analytical tools necessary to approach syntax not as a matter of memorization, but as a science of discovering patterns, forming hypotheses, and testing them against actual language data.

D. Feature Systems

In addition to belonging to specific syntactic categories such as nouns or verbs, words in natural languages also carry grammatical features, abstract properties that influence how they function and interact with other elements in a sentence. These features are not always visible on the surface of the word, but they are crucial in determining grammatical agreement, compatibility, and structure. In syntactic theory, features are often represented using binary or attribute-value notation, such as [$\pm$ plural], [$\pm$ past], or [person: 3rd, number: plural, tense: past]. These features are particularly important in languages with rich inflectional morphology, such as English, Latin, or Russian, but they are also relevant in languages like Indonesian, which rely more on word order and context than on morphological marking.

One of the most commonly discussed features in syntax is number. In English, number is often marked on both nouns and verbs. Nouns typically display plural morphology using the suffix -s or -es (e.g., *book* vs. *books*, *engineer* vs. *engineers*), and verbs show subject-verb agreement, particularly in the present tense for third-person singular subjects. Consider the following pair:

1. *The engineers work hard.*
2. *The engineer works hard.*

In the first sentence, the subject *engineers* is plural, so the verb appears in its bare form *work*, without the third-person singular -s suffix. In the second sentence, the singular noun

engineer triggers the agreement morphology on the verb, resulting in *works*. This agreement is driven by a feature-matching requirement between the subject noun phrase and the finite verb. In generative syntax, this mechanism is often formalized through a process known as feature checking within the structure of the clause (Chomsky, 1995).

Another critical feature is tense, particularly the distinction between past and non-past forms. In English, tense is marked overtly on verbs: *walked* vs. *walk*, *ate* vs. *eat*. This [$\pm$ past] feature plays a vital role in determining verb morphology and temporal interpretation. Some languages mark tense directly on verbs through inflection, while others may use separate auxiliary verbs or particles.

Person and gender are also important syntactic features in many languages. English marks person (1st, 2nd, 3rd) on pronouns (*I*, *you*, *he/she/they*), and this interacts with verb agreement. In other languages, such as Spanish or Arabic, person and gender features also influence verb morphology, noun agreement, and even adjective forms.

When we compare this to Indonesian, a different picture emerges. Indonesian does not overtly mark number or tense on verbs. For example, the sentence *Para insinyur bekerja* ("The engineers work") and *Insinyur itu bekerja* ("The engineer works") both use the same verb form *bekerja*, regardless of whether the subject is singular or plural. The word *para* is a plural marker used with animate human nouns, but the verb form does not change. Likewise, *itu* is a demonstrative used to indicate definiteness, but again, it does not influence verb form. This shows that Indonesian syntax does not rely on subject-verb agreement morphology as English does.

Despite the lack of morphological agreement, Indonesian still relies on syntactic and contextual cues to represent grammatical relationships. As Sneddon et al. (2010) explain, the Indonesian clause structure is relatively analytic, meaning it depends on the fixed order of words and particles to indicate grammatical functions rather than inflection. Thus, while

Indonesian verbs do not carry overt [$\pm$ plural] or [$\pm$ past] features, these distinctions are often inferred from other elements in the sentence or from pragmatic context.

This contrast illustrates how languages vary in the realization of syntactic features. In feature-based syntactic theories such as Lexical-Functional Grammar (LFG) and Minimalist Syntax, features are assumed to be universal, but their surface expression is language-specific. That is, all languages operate with some version of number, person, tense, and aspect features, but not all languages express these features overtly on the word form. In Indonesian, the verb form is more invariant, whereas in English and other Indo-European languages, the verb often carries multiple grammatical features.

Feature systems are also important in syntactic theory for modeling agreement mechanisms, movement operations, and syntactic licensing. For example, in Chomsky's Minimalist Program (1995), features such as [Case], [Person], and [Number] must be checked and valued through operations like Merge and Agree. The failure to check features can lead to ungrammaticality in languages with rich agreement systems.

Understanding feature systems is critical for students of syntax because it reveals how grammatical relations are encoded across different languages and explains why some structures are acceptable in one language but not in another. It also demonstrates the richness of the syntactic component of the grammar, even when surface morphology is minimal, the underlying syntactic representation may still involve complex feature structures.

E. Summary

In this unit, we examined the foundational concept of syntactic categories, which serve as the basic building blocks of sentence structure in all natural languages. Every word in a sentence belongs to a particular category, such as noun, verb, adjective, or determiner, and these categories determine the syntactic roles that words can play and the grammatical

environments they can appear in. Understanding these categories is essential for parsing sentences, constructing syntactic trees, and developing accurate grammatical models.

We began by distinguishing between lexical categories and functional categories, a fundamental division in syntactic theory. Lexical categories, such as nouns, verbs, adjectives, and adverbs, are open-class items that carry the semantic content of a sentence. They denote entities, actions, qualities, and manner. Functional categories, such as determiners, complementizers, auxiliaries, and pronouns, are closed-class items that carry grammatical meaning and perform structural roles within phrases and clauses. Scholars like Carnie (2012), Radford (2012), and Van Gelderen (2002) emphasize that while lexical items are rich in meaning and often subject to innovation, functional elements are relatively fixed and help scaffold syntactic structure.

The unit also explored the distinction between open and closed word classes. Open categories allow for the introduction of new items, as seen in the frequent emergence of new nouns and verbs in both English and Indonesian. Examples include newly coined English terms like *selfie* and *vlog*, or Indonesian adaptations such as *ngeprint* (to print) and *layangan* (drone). Closed categories, in contrast, such as pronouns, prepositions, and auxiliary verbs, rarely admit new members and tend to remain stable over time. Linguists like Baker (2003) and Payne (2006) have documented this universal tendency, arguing that it reflects a basic division between expressive and grammatical components of language.

To identify word classes more rigorously, syntacticians apply several diagnostic tests, including distributional analysis, morphological behavior, and substitution. Distributional tests examine where a word appears in relation to others in a sentence. For example, invented words like *zindle* in “The zindle was glowing” clearly function as nouns because they occupy the same structural position as familiar nouns. Morphological tests explore which affixes a word can take. In English, a word that

accepts tense suffixes like *-ed* or *-ing* is typically a verb. In Indonesian, the ability to accept verbal prefixes such as *meN-* or *ber-* serves a similar diagnostic purpose. Substitution tests further confirm word classes by seeing whether a word can be replaced by another item of the same class without altering the grammaticality of the sentence. This triad of diagnostics provides a systematic and replicable approach to syntactic categorization (Adger, 2003; Radford, 2012; Sneddon et al., 2010).

Beyond these surface-level diagnostics, the unit introduced the concept of feature systems in syntax. Words are not only categorized by their part of speech but also by the grammatical features they carry. These include number, person, tense, aspect, gender, and animacy, among others. In English, such features are often marked overtly. For instance, subject-verb agreement in sentences like “The engineer works” versus “The engineers work” reflects feature matching for number. In contrast, Indonesian does not mark number or tense on verbs overtly, as shown in examples like *Insinyur itu bekerja* (“The engineer works”) and *Para insinyur bekerja* (“The engineers work”), where the verb form *bekerja* remains invariant. While both languages share similar core categories, their morphological strategies for realizing grammatical features differ significantly. These differences underscore the need for abstract syntactic representations that can capture universal principles while accommodating language-specific variation (Baker, 2008; Chomsky, 1995).

The unit also highlighted that features are not merely descriptive tags but play an active role in syntactic operations such as agreement, movement, and licensing. In generative frameworks like the Minimalist Program, features are checked and valued through syntactic processes such as Merge and Agree (Chomsky, 1995). This explains why some sentences are ungrammatical if a feature mismatch occurs, even if the word order appears acceptable.

In sum, Unit 2 has laid the groundwork for understanding how words function within syntactic structures by introducing the concepts of categories, diagnostics, and features. These tools will be crucial for analyzing more complex syntactic phenomena in later units. By comparing English and Indonesian, we have also seen that while all languages categorize words and encode grammatical relationships, they do so in diverse ways, offering valuable insights into the universal and variable aspects of human language.

GLOSSARY

- **Distribution** – Where a word can appear; clues to category.
- **Features** – Person, number, tense, etc. used in agreement/licensing.
- **Functional Categories** – Det, Aux, C, T, P, etc. (grammatical scaffolding).
- **Indonesian Determination** – Demonstratives (itu/ini) function like determiners.
- **Lexical Categories** – N, V, Adj, Adv (content words).
- **Morphology** – Inflection/derivation signals category.
- **Open vs. Closed Class** – Open (new words enter) vs. closed (stable inventory).
- **Substitution Test** – Replace with a pro-form to test category.

UNIT 3

CONSTITUENCY AND PHRASE STRUCTURE

In syntax, one of the most fundamental insights is that words in a sentence are not simply strung together in a linear sequence. Rather, they are grouped into hierarchical units known as constituents. These groupings reflect the way that speakers mentally organize language and the rules that determine grammatical structure. Understanding constituency is essential for syntactic analysis because it reveals how sentences are constructed, how meaning is interpreted, and how syntactic rules apply at various levels of sentence structure.

A. What Is a Constituent?

A constituent is a group of words that functions as a single unit within a larger syntactic structure. Constituents can act as subjects, objects, predicates, modifiers, or entire clauses, depending on their role in the sentence. They may consist of one word or multiple words, but what unites them is their behavior as a single grammatical unit. In generative syntax, identifying constituents is a key step in building phrase structure trees, which visually represent how sentences are organized.

Consider the English sentence:

The girl in the red coat smiled.

Here, the phrase *The girl in the red coat* functions as a noun phrase (NP). It serves as the subject of the sentence and contains multiple elements, a determiner (*the*), a head noun (*girl*), a prepositional phrase (*in the red coat*), that together act as a single

constituent. Syntactically, all of these elements group under a single NP node in phrase structure, and this unit interacts with the rest of the sentence as a block.

A similar structure exists in Indonesian:

Gadis dengan mantel merah itu tersenyum.

In this example, *Gadis dengan mantel merah itu* is also a noun phrase. It consists of the noun *gadis* (girl), followed by the prepositional phrase *dengan mantel merah* (with a red coat), and the demonstrative *itu* (that/that one). Although Indonesian differs in word order and the placement of modifiers and determiners, the same constituent structure is present. The entire phrase serves as the subject and behaves as a single grammatical unit, just like its English counterpart.

The concept of constituency is supported by several kinds of evidence and diagnostic tests. One such test is the substitution test, if a group of words can be replaced by a pronoun (for NPs) or a pro-form (for other types of phrases), then that group is likely a constituent. For instance, in the English sentence *The girl in the red coat smiled*, we can replace the entire noun phrase with *she*, yielding *She smiled*. This confirms that *The girl in the red coat* is a constituent.

Another test is movement, which examines whether a group of words can be moved together to another position in the sentence. For example:

1. Original: *The girl in the red coat smiled.*
2. Fronted: *Smiled, the girl in the red coat did.* (somewhat marked, but possible in emphatic constructions)

The fact that this phrase can be moved as a unit suggests it functions as a constituent. In Indonesian, while such fronting is less common due to the language's relatively fixed word order, clefting or topicalization can be used for emphasis, e.g.

3. *Gadis dengan mantel merah itu, dia tersenyum.* (roughly, "That girl in the red coat, she smiled.").

A third common test is coordination. If two similar units can be joined by a conjunction such as *and* or *atau* (in Indonesian), and the result is grammatical, this suggests each joined unit is a constituent of the same type. For example:

The girl in the red coat and the boy with the backpack smiled.

Both *The girl in the red coat* and *the boy with the backpack* are noun phrases, and their coordination confirms their parallel status as constituents.

Linguists such as Radford (2012), Carnie (2012), and Van Gelderen (2002) stress that understanding constituency is a gateway to more complex syntactic theory. It provides the basis for constructing X-bar theory, phrase structure rules, and understanding hierarchical structure, all of which are central to generative grammar. Constituents form not just the surface structure of a sentence but also its deep structure, influencing how transformations like movement and deletion operate within the syntax.

Moreover, recognizing constituents helps explain ambiguities in sentence interpretation. Consider the sentence:

She saw the man with the telescope.

This could mean either that she used a telescope to see the man (with the telescope modifying *saw*), or that she saw a man who had a telescope (with the telescope modifying *the man*). Each interpretation corresponds to a different constituent structure. Understanding constituency allows syntacticians to diagram both structures and explain how different meanings arise from different groupings.

From a cross-linguistic perspective, the presence of constituency is considered a universal of natural languages, although the specific forms and markers of constituents may vary. Indonesian, for instance, often uses post-nominal modifiers and lacks case marking, but it still builds syntactic structures around clear constituent groupings. This supports the hypothesis in typological linguistics that all human languages exhibit some form of hierarchical organization in sentence construction (Dryer, 2007; Payne, 2006).

In sum, a constituent is not merely a convenient label but a fundamental syntactic concept that reflects how language users structure thought, encode relationships, and generate well-formed sentences. The identification and analysis of constituents are essential for understanding how syntax works both within a single language and across the world's languages.

B. Tests for Constituency

In order to identify whether a particular string of words functions as a constituent within a sentence, syntacticians employ a series of diagnostic tests. These tests do not rely on intuition alone but provide systematic and replicable procedures for uncovering hierarchical structure. Since constituents operate as unified syntactic units, they exhibit predictable behaviors in relation to movement, replacement, and coordination. These behaviors are instrumental in building phrase structure trees and in constructing formal syntactic theories. Among the most reliable diagnostics are the substitution test, the movement or fronting test, and the coordination test. These tests, which are applicable across languages, are foundational in works such as Radford (2009), Carnie (2012), and Adger (2003), and are frequently used in linguistic fieldwork, language acquisition research, and formal grammar modeling.

The substitution test is one of the most straightforward diagnostics. If a sequence of words can be replaced by a single pro-form (such as a pronoun or another type of function word) without altering the grammaticality or core meaning of the sentence, that sequence is likely a constituent. In English, for example, consider the sentence *The girl in the red coat smiled*. The entire phrase *The girl in the red coat* can be replaced by the pronoun *she*, resulting in *She smiled*. The grammaticality and interpretability of the resulting sentence confirm that the replaced group of words forms a noun phrase (NP), functioning as a subject constituent. Similarly, in Indonesian, *Gadis itu tersenyum* ("The girl smiled") may be reformulated as *Dia tersenyum* ("She smiled"), with *dia* substituting for the full noun

phrase. This parallel substitution further supports the notion that Indonesian, like English, exhibits constituent structure even though it lacks overt inflectional morphology in many cases (Payne, 2006; Sneddon et al., 2010).

The movement test, sometimes referred to as fronting or topicalization, investigates whether a sequence of words can be relocated to a different syntactic position, often to the front of a sentence, while preserving grammaticality and coherence. Constituents can typically undergo such displacement as a single unit. In English, the sentence *I saw the girl in the red coat* can be transformed into *The girl in the red coat, I saw*, where the noun phrase has been fronted for topical emphasis. Although somewhat marked in everyday speech, this construction is syntactically acceptable and reinforces the phrase's constituent status. In Indonesian, which generally adheres to subject-verb-object (SVO) order, a similar strategy is used for emphasis or topic-focus constructions. The sentence *Saya melihat gadis dengan mantel merah itu* ("I saw the girl in the red coat") can be reordered as *Gadis dengan mantel merah itu, saya lihat*, which likewise emphasizes the noun phrase and confirms its integrity as a constituent (Musgrave, 2001; Sneddon et al., 2010). While the degree of flexibility varies by language, the ability to move a string of words as a block is strong evidence for their constituency.

The coordination test examines whether a sequence of words can be conjoined with another sequence of the same syntactic type using a coordinating conjunction such as *and* in English or *dan* in Indonesian. Only constituents of the same category can typically be coordinated. For instance, in English, the sentence *The girl in the red coat and the boy with the blue hat smiled* successfully joins two noun phrases. The symmetry in structure and function supports the conclusion that both coordinated elements are constituents. In Indonesian, a similar structure is found in *Gadis dengan mantel merah dan anak laki-laki dengan topi biru itu tersenyum*, where two complex noun phrases are conjoined and function together as the subject of the verb

tersenyum ("smiled"). As Radford (2009) and Adger (2003) note, coordination is a particularly robust test because it is highly sensitive to syntactic type, if coordination is grammatical, it provides strong confirmation that both elements share constituent status.

It is important to recognize, however, that no single constituency test is flawless. Each test may yield ambiguous or inconclusive results in certain contexts, especially when applied to complex or elliptical constructions. Therefore, syntacticians are advised to apply multiple diagnostics in combination and interpret the outcomes holistically. The convergence of positive results across different tests significantly increases confidence in constituent identification. As Carnie (2012) points out, syntactic theory relies on such cumulative evidence to build and justify tree structures, to propose transformations, and to identify functional projections in clause architecture.

Furthermore, the application of constituency tests across languages, such as English and Indonesian, reinforces the cross-linguistic validity of syntactic principles. While word order, morphology, and movement constraints differ between languages, the underlying structural behavior of constituents remains consistent. This supports the broader claim in generative grammar that human languages are built from similar cognitive and structural principles, despite surface-level variation (Baker, 2002; Chomsky, 1995).

In conclusion, the substitution, movement, and coordination tests offer essential empirical tools for identifying syntactic constituents. These tests reveal that linguistic structures are not flat but hierarchically organized, with words grouped into units that function together. Recognizing and analyzing these constituents allows linguists and students to construct more accurate representations of sentence structure, develop syntactic theories, and understand the interplay between form and meaning in natural language.

C. Phrase Structure Rules (PSRs)

Phrase Structure Rules (PSRs), also known as rewrite rules, are formal syntactic devices used to represent the hierarchical structure of phrases and sentences in a language. These rules express how smaller units (words and phrases) are systematically combined to form larger syntactic constituents. In traditional generative grammar, PSRs provide a way of specifying the permissible combinations of categories in a language, contributing to the construction of syntactic trees that visually represent the grammatical architecture of sentences.

At the most basic level, a sentence (S) can be represented using a PSR such as:

$$S \rightarrow NP VP$$

This rule states that a sentence consists of a noun phrase (NP) followed by a verb phrase (VP). This formulation captures the essential subject-predicate structure of sentences in English and many other languages. Within each of these major constituents, further rules specify their internal composition. For instance:

1. $NP \rightarrow \text{Det } N \text{ (PP)}$
2. $VP \rightarrow V \text{ (NP) (PP)}$

These rules indicate that a noun phrase may consist of a determiner followed by a noun, optionally accompanied by a prepositional phrase (PP). A verb phrase may consist of a verb, optionally followed by a noun phrase (as a direct object) and/or a prepositional phrase. The parentheses indicate optional elements, a notational convention widely used in linguistic texts such as Radford (2012), Carnie (2012), and Adger (2003). Such representations are particularly useful in the development of phrase structure grammars, which form the basis for X-bar theory and later models like Bare Phrase Structure within the Minimalist Program (Chomsky, 1995).

These rules not only help identify possible sentence structures but also rule out ungrammatical combinations. For example, the ungrammatical English sentence "*The ran book*" violates PSRs because a determiner and a verb cannot co-occur

as a noun phrase. PSRs constrain syntactic formation in ways that reflect native speaker intuitions about grammaticality.

When applied to Indonesian, these rules remain largely valid, albeit with important language-specific adjustments. Indonesian, like English, generally adheres to a Subject-Verb-Object (SVO) word order (Sneddon et al., 2010). Thus, the high-level PSR:

$S \rightarrow NP VP$

Applies equally well to both languages. Consider the sentence:

Siswa itu membaca buku.

("The student read a book.")

Here, *siswa itu* ("the student") functions as the subject NP, and *membaca buku* ("read a book") forms the VP. The VP consists of the verb *membaca* (to read) and the object NP *buku* (book). Although syntactically parallel to English, a notable typological distinction is that Indonesian often lacks overt determiners, relying instead on demonstratives or contextual inference for definiteness. The word *itu* (that) can serve as a post-nominal determiner-like element, but its presence is not obligatory in most noun phrases. For instance, *siswa membaca buku* (a grammatically valid sentence) lacks any overt determiner but remains syntactically complete. As such, while the $NP \rightarrow Det N$ rule is standard in English, a more accurate PSR for Indonesian would accommodate the optionality or absence of determiners:

1. $NP \rightarrow (Det) N (PP)$
2. $VP \rightarrow V (NP) (PP)$

This modified rule reflects the fact that noun phrases in Indonesian may consist of bare nouns, with or without modifiers. Adelaar and Himmelmann (2004) emphasize that many Austronesian languages, including Indonesian, rely heavily on word order, discourse context, and pragmatic markers to convey definiteness, animacy, or number, rather than relying on a closed class of determiners.

Another important syntactic feature of Indonesian is its treatment of verbs. Unlike English, which marks tense morphologically (*work* vs. *worked*), Indonesian verbs typically do not inflect for tense. Instead, aspect and temporal interpretation are inferred from adverbs or context. This morphological simplicity does not imply syntactic impoverishment; rather, it highlights the importance of phrase structure and word order in clause construction.

Moreover, Indonesian verbal phrases can contain pre-verbal auxiliaries (e.g., *akan* for future tense), as in *siswa itu akan membaca buku* ("The student will read a book"). This supports an extension of the VP rule to accommodate auxiliary structures:
$$VP \rightarrow (Aux) V (NP) (PP)$$

Phrase structure rules thus offer a formal mechanism for capturing both the universal properties of syntax and the specific grammatical characteristics of individual languages. These rules are instrumental in pedagogical grammar, computational parsing, and theoretical syntax alike. They not only specify linear order but also establish hierarchical relationships between constituents, forming the foundation for further analyses involving binding, movement, and clause embedding (Haegeman, 1994; Lasnik et al., 2005).

In sum, Phrase Structure Rules are a critical component of syntactic theory. They provide a concise and formal means of describing how words combine to form constituents and how those constituents form larger structures. While languages like English and Indonesian share core structural patterns, they also exhibit unique variations, particularly in the presence or absence of overt determiners, morphological marking, and auxiliary constructions. These differences underscore the need for flexible and language-specific adaptations of PSRs while maintaining a core framework of universal syntactic principles.

D. Bracket Notation and Tree Diagrams

A fundamental principle in syntactic theory is that sentence structure is not merely linear but hierarchical. Words combine into phrases, which in turn combine into larger phrases and ultimately into full clauses and sentences. This recursive and nested architecture is best captured through constituent structure representations, of which two widely used forms are bracket notation and syntactic tree diagrams. These tools make visible the internal organization of sentences, revealing how smaller units group together and interact within the larger structure.

Bracket notation, sometimes called box notation, is a linear method of representing syntactic hierarchy. It uses square brackets to indicate constituents and category labels to identify the function of each grouping. Consider the following English sentence:

The cat chased the mouse.

In bracket notation, this is rendered as:

[S [NP The cat] [VP chased [NP the mouse]]]

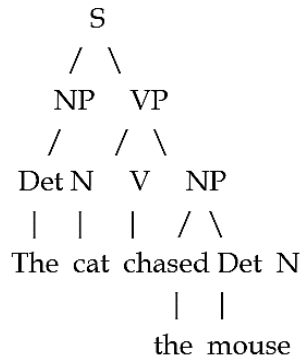
This notation tells us that the sentence (S) is composed of two primary constituents: a noun phrase (NP) *The cat*, and a verb phrase (VP) *chased the mouse*. Within the VP, we find a verb *chased* and another NP *the mouse*. The recursive nature of the structure, phrases containing other phrases, is captured neatly through nested brackets. As Radford (2012) and Carnie (2012) emphasize, bracket notation offers an efficient way to trace hierarchical relationships without the spatial complexity of full tree diagrams, especially in early-stage syntactic analysis or when typing text.

The equivalent structure in Indonesian can be represented similarly:

1. *Kucing itu mengejar tikus itu.*
2. [S [NP Kucing itu] [VP mengejar [NP tikus itu]]]

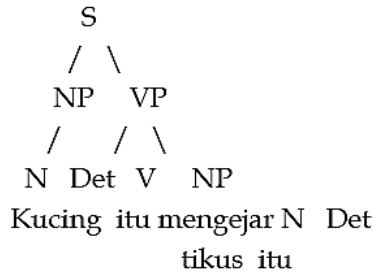
Here, *Kucing itu* ("the cat") functions as the subject NP, and *mengejar tikus itu* ("chased the mouse") functions as the VP. Like in English, the VP consists of a verb followed by an object NP. While Indonesian often places demonstratives like *itu* ("that") post-nominally, the overall constituent hierarchy remains parallel. This comparison supports cross-linguistic syntactic universals: both languages form sentences through subject-verb-object (SVO) structuring, even if morphological or lexical specifics differ (Payne, 2006; Sneddon et al., 2010).

While bracket notation is efficient and compact, tree diagrams (also known as phrase structure trees) offer a more intuitive and visual way of representing syntactic structure. In a tree diagram, each node represents a syntactic category, and branching lines show how constituents are combined. For the sentence *The cat chased the mouse*, the tree would appear as follows:



This structure clearly shows the internal composition of the noun phrases and verb phrase. The determiner and noun combine to form NP; the verb and object NP form the VP; and these together construct the sentence. Each branch corresponds to a rule from phrase structure grammar (e.g., $NP \rightarrow Det\ N$, $VP \rightarrow V\ NP$, $S \rightarrow NP\ VP$), reinforcing the direct relationship between PSRs and tree representations (Adger, 2003; Haegeman, 1994).

The same sentence in Indonesian can be represented analogously:



Here, the NP *Kucing itu* is decomposed into a noun and a post-nominal determiner. Similarly, *tikus itu* forms the object NP. This structure illustrates a key cross-linguistic difference: while English determiners are typically pre-nominal, Indonesian determiners such as *itu* follow the noun. Despite these surface differences, both languages share hierarchical structure and phrase composition.

Tree diagrams are not merely illustrative; they play a critical role in theoretical syntax. They allow linguists to identify constituency, c-command relations, scope, and movement domains, which are essential for understanding binding theory, question formation, and syntactic dependencies. Trees also allow researchers to model grammatical derivations within frameworks such as X-bar theory and the Minimalist Program, both of which rely heavily on binary branching and hierarchical phrase structures (Chomsky, 1995; Lasnik et al., 2005).

Moreover, tree representations help expose ambiguity in sentence structure. Consider the sentence *I saw the man with the telescope*. This can be parsed in two different ways:

1. [VP saw [NP the man [PP with the telescope]]] (the man has the telescope)
2. [VP [VP saw the man] [PP with the telescope]] (I used the telescope)

Each reading corresponds to a different tree structure, and understanding how these are formed helps explain how syntactic parsing relates to meaning.

In linguistic pedagogy, tree diagrams also serve as an effective visual tool for students to internalize phrase structure rules and explore sentence composition. As Carnie (2012) notes, drawing trees allows learners to practice abstract reasoning, hypothesis testing, and rule application, core skills in both formal linguistics and cognitive science.

In summary, bracket notation and tree diagrams are indispensable tools in syntactic theory. Brackets offer a compact way to trace constituent boundaries, while trees provide a visual representation of hierarchical structure and phrase composition. Both methods reveal the nested architecture of natural language, enabling precise analysis of sentence structure. Through comparative examples in English and Indonesian, it becomes evident that the syntactic systems of natural languages, though varied in surface realization, share deep structural similarities rooted in universal grammar.

E. Structural Ambiguity

A central insight in syntactic theory is that sentence meaning is determined not only by the meanings of individual words but also by how those words are structurally combined. When a sentence permits more than one possible syntactic structure, it is said to exhibit structural ambiguity. This phenomenon occurs when a single surface string corresponds to two (or more) distinct hierarchical representations, each of which yields a different interpretation. Structural ambiguity highlights the importance of constituency and phrase structure in semantic interpretation and is an essential topic in both theoretical syntax and psycholinguistics.

Consider the following English sentence:

She watched the man with binoculars.

This sentence is structurally ambiguous because the prepositional phrase *with binoculars* can attach to two different parts of the sentence, yielding two possible interpretations. In the first reading, the phrase modifies the verb phrase (VP), implying that *she* used binoculars as an instrument:

[VP watched [NP the man]] [PP with binoculars] → She used binoculars.

In the second reading, the same prepositional phrase modifies the noun phrase (NP), suggesting that the *man* is the one associated with the binoculars:

[VP watched [NP the man [PP with binoculars]]] → The man had the binoculars.

Although the linear word order remains unchanged, the hierarchical structure, and hence the meaning, differs significantly. These two interpretations correspond to distinct syntactic trees, with *with binoculars* attached at different levels. As Radford (2012), Carnie (2012), and O'Grady et al. (2009) point out, such ambiguity cannot be resolved by lexical meaning alone; it is the structural configuration that determines the scope and attachment of modifiers.

A comparable case exists in Indonesian:

Dia melihat pria dengan teropong.

Here, the phrase *dengan teropong* ("with binoculars") introduces similar ambiguity. As in English, this prepositional phrase may either modify the verb *melihat* ("see"), indicating the means by which the seeing was performed (i.e., she used binoculars), or it may modify the noun *pria* ("man"), specifying which man was seen (i.e., the man who had binoculars). In ordinary discourse, additional contextual or prosodic cues may resolve the ambiguity, but in isolation, the sentence remains structurally ambiguous.

The two interpretations can be made explicit as follows:

1. Verb-modifying reading (instrumental):

Dia melihat [pria] [dengan teropong] → She used binoculars to see the man.

2. Noun-modifying reading (descriptive):

Dia melihat [pria dengan teropong] → She saw a man *who* had binoculars.

This ambiguity arises due to the attachment site of the prepositional phrase. When the prepositional phrase attaches high, modifying the VP, it serves as an adjunct describing the action. When it attaches low, within the NP, it restricts the reference of the noun. This distinction between high and low attachment is a core concept in syntactic theory and has been extensively studied in cross-linguistic parsing research (Clifton & Frazier, 1989; Frazier, 1987). Structural ambiguity of this kind provides compelling evidence that sentence interpretation must be mediated by hierarchical syntax rather than surface sequence alone.

Such ambiguity is also relevant in syntactic theory because it illustrates the limits of phrase structure rules when divorced from semantics or prosody. PSRs can generate both structures, but only further syntactic analysis, e.g., using trees or binding theory, or external cues such as intonation or context can disambiguate them. For instance, a pause or prosodic boundary in speech might signal which constituent the prepositional phrase is modifying. In writing, punctuation or background information may provide similar cues.

Moreover, structural ambiguity has implications beyond theoretical linguistics. In fields such as computational linguistics and natural language processing (NLP), ambiguity presents a major challenge for sentence parsing and machine translation. A system that fails to distinguish between alternative structures may produce incorrect or nonsensical outputs. Similarly, in second language acquisition, learners often struggle to detect or resolve ambiguities without explicit instruction, particularly when their L1 lacks analogous structures (White, 2012).

From a cognitive perspective, the way speakers and listeners resolve structural ambiguity sheds light on real-time language processing. Garden-path sentences, such as *The horse raced past the barn fell*, exploit structural ambiguity to illustrate

how readers make initial syntactic decisions and then revise them. Research by Frazier and others has demonstrated that native speakers typically favor low attachment or right-branching interpretations due to processing efficiency and memory constraints (Clifton & Frazier, 1989; Frazier, 1987).

In conclusion, structural ambiguity arises when a sentence allows multiple syntactic parses, each yielding a different interpretation. This phenomenon underscores the importance of constituent structure and hierarchical syntax in sentence processing and meaning construction. Both English and Indonesian exhibit structural ambiguity in similar syntactic environments, reinforcing the universality of the issue and highlighting the relevance of syntactic theory for understanding meaning, parsing, and ambiguity resolution across languages.

F. Summary

This unit has introduced one of the most foundational concepts in syntactic theory: constituency. Constituents are syntactic units, groups of words that function as a single element within the hierarchical structure of a sentence. They form the internal architecture of natural language, organizing linear strings of words into recursive and meaningful configurations. Constituents serve various grammatical functions, such as subjects, predicates, objects, or modifiers, and understanding their structure is a prerequisite for analyzing sentence formation in any natural language.

To systematically identify constituents, syntacticians employ several empirical diagnostic tests, the most prominent of which are substitution, movement (or fronting), and coordination. The substitution test allows a constituent to be replaced by a pro-form (such as a pronoun or pro-verb), while movement tests involve fronting the suspected constituent to another position in the sentence. Coordination further confirms constituency by combining two similar units with a coordinating conjunction, demonstrating their parallel syntactic status. These tests, extensively discussed in works by Radford (2012), Carnie

(2012), and Adger (2003), not only confirm constituent boundaries but also reveal deeper hierarchical structure that is not always apparent in surface word order.

Building on these diagnostic strategies, the unit explored how phrase structure rules (PSRs) formally define the composition of syntactic categories and constituents. Rules such as $S \rightarrow NP VP$ or $VP \rightarrow V (NP) (PP)$ specify the permissible configurations of syntactic units in a given language. In English, for example, PSRs explain how determiners, nouns, and modifiers combine to form noun phrases, and how these phrases are nested within larger structures. Indonesian exhibits similar constituent behavior, with surface-level differences such as post-nominal determiners and a more flexible use of modifiers. Despite these differences, both languages can be modeled using equivalent PSRs, albeit with necessary adjustments that account for language-specific morphological and syntactic patterns (Adelaar & Himmelmann, 2004; Sneddon et al., 2010).

To visualize syntactic structure, linguists use bracket notation and phrase structure trees, both of which make explicit the hierarchical and recursive nature of language. Bracket notation provides a compact, linear representation of constituent grouping, while tree diagrams represent the same structures visually, using branching nodes to indicate the dominance and dependency relations among constituents. These tools are indispensable for illustrating abstract grammatical rules and for modeling how languages generate infinite sentence structures from a finite set of rules. Trees are also essential in theoretical models such as X-bar theory and the Minimalist Program, where hierarchical relations underpin transformational operations and feature checking (Chomsky, 1995; Lasnik et al., 2005).

A particularly illuminating topic addressed in this unit is structural ambiguity, which arises when a sentence allows for more than one syntactic parse, each yielding a different interpretation. For instance, the English sentence *She watched the man with binoculars* is ambiguous: it can mean that she used binoculars, or that the man possessed them. This ambiguity

stems from the potential for the prepositional phrase *with binoculars* to attach at two different structural levels, either as a modifier of the verb phrase or as a modifier of the noun phrase. Indonesian exhibits parallel patterns, as in *Dia melihat pria dengan teropong*, which is similarly ambiguous without contextual clarification. Such examples underscore the critical role of syntactic structure in determining meaning, and they demonstrate how ambiguity cannot be resolved through lexical semantics alone but must be addressed at the level of constituent grouping and phrase attachment (Clifton & Frazier, 1989; Frazier, 1987).

In a cross-linguistic context, the analysis of English and Indonesian reveals that while the two languages differ in morphosyntactic realization, such as the presence or absence of overt determiners, or variations in modifier placement, the underlying principles of constituency, phrase structure, and hierarchical grouping remain remarkably consistent. This supports the notion, central to generative grammar, that human languages share a universal structural foundation, even when surface-level features vary (Chomsky, 1993; O'Grady et al., 2009).

In conclusion, unit 3 provides the essential conceptual and technical framework for syntactic analysis. Through the study of constituents, phrase structure rules, tree diagrams, and ambiguity, students gain the tools to investigate the deep organization of sentence structure. These insights not only inform linguistic theory but also have applications in areas such as language acquisition, computational linguistics, and translation studies. Mastery of constituency lays the groundwork for more advanced topics in syntax, including binding theory, movement phenomena, and the architecture of the clause.

GLOSSARY

- **Constituent** – Words forming a single syntactic unit.
- **Coordination Test** – Conjoin parallel units (NP and NP).
- **Head** – Category projecting the phrase (e.g., N in NP).
- **Movement Test** – Move the sequence as a block (topic/fronting).
- **Phrase Structure Rules (PSRs)** – Formal rewrite rules (e.g., $S \rightarrow NP VP$).
- **Structural Ambiguity** – Multiple parses yield multiple meanings.
- **Substitution Test (Constituency)** – Replace sequence with a pro-form (e.g., pronoun).
- **Tree Diagrams** – Hierarchical representation of structure.



**PART II:
THE ARCHITECTURE
OF SENTENCES**



Having established the foundational principles of syntax in Part I, we now shift our focus to the internal architecture of sentence structure. Part II will introduce the architecture of sentence structure through X-bar theory and functional projections, providing a deeper understanding of how phrases are constructed, expanded, and hierarchically organized in natural language. This section builds upon the notion of constituency by exploring how heads, complements, specifiers, and adjuncts systematically combine to form complex syntactic units. Through the formalization offered by X-bar theory and its successors, we aim to uncover the universal blueprint that governs the composition of noun phrases, verb phrases, and entire clauses across languages. Special attention will be paid to how these structures manifest in both English and Indonesian, illustrating both cross-linguistic parallels and typological distinctions. These concepts serve as essential groundwork for more advanced topics in sentence derivation, argument structure, and transformational syntax in later units.

UNIT 4

X-BAR THEORY AND SENTENCE STRUCTURE

A. X-bar Schema and Its Components

The structural analysis of sentence components in generative grammar has evolved from simple phrase structure rules (PSRs) to more sophisticated frameworks that aim to account for universal principles of syntactic organization. Among the most influential of these frameworks is the X-bar theory, developed within the Government and Binding framework of Chomsky (1981) and further formalized by Jackendoff (1977,1990). X-bar theory offers a universal template for phrase structure that can be applied consistently across lexical categories, nouns, verbs, adjectives, and prepositions, thereby promoting cross-categorial symmetry and theoretical elegance.

At the core of X-bar theory is the idea that all phrases, regardless of category, are hierarchically organized into three structural layers: a head (X), an intermediate projection (X' or X-bar), and a maximal projection (XP). These layers correspond to the structural functions of the elements that compose a phrase: the head, the complement, the specifier, and adjuncts. This tripartite structure allows syntacticians to represent both obligatory and optional constituents in a principled way, providing greater explanatory adequacy than flat PSRs.

The canonical X-bar schema is represented through the following rules:

1. $XP \rightarrow (\text{Specifier}) X'$

2. $X' \rightarrow X$ (Complement)
3. $X' \rightarrow X'$ (Adjunct)

These rules indicate that the specifier (e.g., determiners, quantifiers) appears at the edge of the phrase, the complement appears closest to the head, and adjuncts are recursively adjoined to intermediate projections. Crucially, X-bar theory posits that all phrases conform to this architecture, whether they are noun phrases (NP), verb phrases (VP), adjective phrases (AP), or prepositional phrases (PP).

Consider the following English noun phrase analyzed using X-bar theory:

[NP [Det those] [N' [N ideas] [PP about marsupials]]]

In this structure, *those* functions as the specifier (Det), *ideas* is the head noun (N), and *about marsupials* is a prepositional phrase functioning as a complement or adjunct depending on interpretation. If the prepositional phrase is selected by the noun (e.g., *the idea about something*), it is considered a complement. If it is freely attached and not lexically selected, it is treated as an adjunct.

A comparable structure in Indonesian may appear as follows:

[NP [Det ide-ide] [N' [N itu] [PP tentang marsupial]]]

Here, *ide-ide* is the pluralized noun (with reduplication indicating plurality), *itu* functions as a post-nominal demonstrative determiner, and *tentang marsupial* is a prepositional phrase that modifies the noun. While Indonesian typically places demonstratives and modifiers in a post-nominal position, unlike the pre-nominal position common in English, the hierarchical structure remains consistent. This demonstrates how X-bar theory accommodates typological variation by abstracting away from surface word order and focusing on structural relations (Adelaar & Himmelmann, 2004; Sneddon et al., 2010).

One of the key advantages of the X-bar framework is that it captures the recursivity of syntactic structures. Adjuncts, for example, can be added repeatedly without limit, forming ever larger and more complex phrases. For instance:

[NP [Det those] [N' [N ideas] [PP about marsupials] [PP from Australia]]]

Each additional prepositional phrase is adjoined to the intermediate projection (N'), preserving the recursive nature of syntax. This property is crucial for accounting for the generative capacity of human language, the ability to produce and interpret an unbounded number of novel sentences from a finite set of elements (Chomsky, 1965, 1995).

Furthermore, X-bar theory provides a uniform syntactic template that simplifies the formulation of transformational rules. Because all phrases share the same underlying structure, rules such as movement, agreement, and feature checking can be generalized across categories. For example, the movement of a *wh*-phrase in interrogatives, or the agreement between subject and verb, can be explained more systematically when both noun and verb phrases follow the same projectional template (Haegeman, 1994; Radford, 2012).

From a cross-linguistic perspective, the application of X-bar theory to Indonesian reveals both shared syntactic principles and language-specific realizations. While English typically marks specifiers with determiners and modifies nouns through pre-nominal adjectives or post-nominal prepositional phrases, Indonesian relies on juxtaposition, word order, and particles such as *yang* to indicate modification. Nevertheless, the hierarchical structure predicted by X-bar theory still applies. The maximal projection (NP) in Indonesian still contains a head (N), optional complement or adjunct (PP), and optional specifier or demonstrative (Det), even if the surface order deviates from the English pattern (Musgrave, 2001; van Valin, 2012).

In conclusion, X-bar theory enriches our understanding of syntactic structure by introducing a generalized and hierarchical template that applies across all phrasal categories. It replaces ad

hoc phrase structure rules with a principled architecture that accommodates both obligatory and optional constituents. Through the components of head, complement, adjunct, and specifier, X-bar theory captures the recursive and universal nature of syntactic structure. Its applicability to both English and Indonesian, despite their typological differences, illustrates its theoretical robustness and cross-linguistic validity.

B. Specifiers, Complements, and Adjuncts

In X-bar theory, syntactic constituents are not simply ordered in a flat sequence; rather, they are arranged within a hierarchical template that reflects the internal structure of phrases. Three core components of this architecture are the specifier, the complement, and the adjunct. Each of these elements plays a distinct syntactic role and occupies a specific position relative to the head of a phrase. The clear differentiation among these components enables linguists to systematically account for syntactic variation, semantic scope, and structural recursion across languages.

The specifier occupies a position at the periphery of a phrase, specifically the left edge in head-initial languages like English. Specifiers often function to narrow or contextualize the reference of the phrase. In noun phrases (NPs), the specifier position is typically filled by determiners such as *the*, *this*, or *those*; in verb phrases (VPs), it may host subjects (depending on the analysis); and in tense phrases (TPs), it is commonly occupied by the grammatical subject of the clause. Specifiers are not selected by the head but are licensed by the phrase as a whole. Their function is often discourse-sensitive, helping to anchor the phrase in the larger syntactic or pragmatic context (Haegeman, 1994; Radford, 2012).

The complement is a phrase that is semantically and syntactically required by the head. It is selected by the head and typically appears adjacent to it. Complements are essential for the saturation of a predicate or nominal projection, meaning that without the complement, the phrase may be incomplete or

uninterpretable. For instance, in the noun phrase *the fear of snakes*, the prepositional phrase *of snakes* is a complement of the noun *fear*—it completes the meaning by specifying what the fear is directed toward. Similarly, in *give a book to the teacher*, the noun phrase *a book* and the prepositional phrase *to the teacher* are complements required by the ditransitive verb *give*. The notion of theta roles and selectional restrictions is central to identifying complements, as discussed in Chomsky (1993) and Carnie (2012).

The adjunct differs from complements in that it is optional and recursively iterable. Adjuncts are not selected by the head but instead serve to provide additional, often descriptive or circumstantial, information. They are typically adverbial or adjectival in nature and may express time, manner, place, or reason. For example, in *the creature from the cave with glowing eyes*, *from the cave* and *with glowing eyes* may both appear post-nominally, but their syntactic status differs. If *with glowing eyes* is required to complete the meaning (e.g., if *creature* in context must be associated with such a PP), it may be analyzed as a complement; otherwise, it functions as an adjunct. As adjuncts can be inserted or removed without affecting grammaticality, they are structurally adjoined to X-bar levels, typically X' projections (Haegeman, 1994; R. Jackendoff, 1977).

Consider the following English example, analyzed within this framework:

The creature [with glowing eyes] [from the cave] attacked.

In this noun phrase, *creature* is the head (N), *with glowing eyes* serves as a complement if it is semantically selected by the noun, and *from the cave* functions as an adjunct, providing locative context but not required by the noun. The determiner *the* serves as the specifier of the NP.

Now examine the Indonesian equivalent:

Makhluk [bermata menyala] [dari gua] menyerang.

In this sentence, *makhluk* is the head noun, *bermata menyala* ("with glowing eyes") functions as a modifier that may be interpreted as a complement depending on semantic selection, and *dari gua* ("from the cave") is clearly an adjunct expressing

location. Notably, Indonesian exhibits a high degree of head-initiality, but its post-nominal modifiers (including relative clauses, PPs, and demonstratives) offer rich data for analyzing adjunct-complement distinctions (Musgrave, 2001; Sneddon et al., 2010).

Furthermore, Indonesian utilizes morphological derivation (such as *ber-* and *meN-* prefixes) in place of many prepositions and auxiliary constructions found in English. This results in surface differences in how complements and adjuncts are realized, but the underlying syntactic principles remain consistent. For example, *bermata menyala* consists of the prefix *ber-* ("having") and the noun *mata* ("eyes"), forming a complex predicate that functions similarly to *with glowing eyes* in English.

From a theoretical perspective, the clear separation of specifiers, complements, and adjuncts allows for more precise rule formulation, improves the predictability of syntactic structures, and underpins much of the transformational machinery used in generative grammar. In Minimalist syntax, for example, complements are merged first (closer to the head), followed by adjuncts (external merge), and finally specifiers (often the result of internal merge or movement) (Adger, 2003; Chomsky, 1995).

In sum, understanding the distinction among specifiers, complements, and adjuncts is fundamental to analyzing phrase structure. Each plays a structurally and functionally unique role within the architecture of phrases. While English and Indonesian may differ in morphological expression and surface word order, both languages conform to the same deeper syntactic principles articulated through X-bar theory. Recognizing these roles enhances our ability to parse, represent, and interpret complex sentences in both descriptive and theoretical syntax.

C. From XP to CP: Layering the Structure

As syntactic phrases combine to form larger units, they eventually culminate in clauses, which function as the maximal projections within sentence structure. In the framework of X-bar

theory and later syntactic models such as the Government and Binding Theory (Chomsky, 1993) and the Minimalist Program (Chomsky, 1995), clauses are not flat or loosely assembled structures; rather, they are highly layered, with a hierarchical organization that encodes grammatical features such as tense, aspect, modality, and subordination. These layers are formalized through a series of functional projections, the most prominent of which are the Tense Phrase (TP) and the Complementizer Phrase (CP).

At the heart of clausal architecture is the idea that the clause is headed not simply by a verb, but by a tense-bearing element, typically realized as an auxiliary verb, modal, or inflectional morpheme. This head projects a Tense Phrase (TP), which hosts the subject in its specifier position and governs the verb phrase (VP) as its complement. This structure can be represented through the following schema:

$$\text{TP} \rightarrow \text{NP T}'$$

$$\text{T}' \rightarrow \text{T VP}$$

Here, the specifier of TP (usually a noun phrase) corresponds to the grammatical subject of the sentence. The head T encodes temporal and modal information (e.g., *will*, *has*, *did*), and its complement is a VP that carries the core predication of the sentence.

Above the TP sits the Complementizer Phrase (CP), which forms the outermost clausal projection. The CP serves several syntactic and semantic functions: it introduces embedded clauses, marks interrogatives or declaratives, and provides a position for moved constituents in transformations such as *wh*-movement or topicalization. The basic structure of the CP is given as:

$$\text{CP} \rightarrow \text{C TP}$$

In this configuration, C (the complementizer) may be overt, such as *that*, *if*, or *whether* in English, or covert, depending on the clause type. The specifier of CP is the canonical landing site for fronted elements in interrogatives and relative clauses (Adger, 2003; Radford, 2012).

Consider the English sentence:

That Mira saw the painting amazed me.

In this example, *that* is the complementizer (C), introducing an embedded declarative clause. The embedded clause (*that Mira saw the painting*) functions as the subject of the matrix predicate *amazed me*. The internal structure of the embedded clause reveals multiple layers: *Mira* is the specifier of TP (the subject), *saw the painting* is the VP (the predicate), and *that* occupies the head of CP.

A structurally parallel construction is found in Indonesian:

Bahwa Mira melihat lukisan itu mengejutkan saya.

Here, *bahwa* serves as the complementizer (C), introducing the embedded clause *Mira melihat lukisan itu*. The noun phrase *Mira* functions as the subject (specifier of TP), and *melihat lukisan itu* is the verb phrase complement. The embedded clause as a whole (*bahwa Mira melihat lukisan itu*) serves as the subject of the matrix predicate *mengejutkan saya* ("surprised me").

Although the surface order in Indonesian often appears more rigid and lacks certain inflectional morphology present in English, the underlying hierarchical structure remains compatible with the CP-TP-VP schema. Indonesian does not exhibit subject-auxiliary inversion or inflection for tense in the same way as English, but it compensates for this through temporal adverbs, modal particles, and auxiliary-like verbs (e.g., *akan* for future, *sudah* for perfective), which may be analyzed as occupying T. The complementizer *bahwa* functions analogously to *that* in English and marks the beginning of embedded declarative clauses (Musgrave, 2001; Sneddon et al., 2010).

This layering of functional projections is essential for capturing syntactic phenomena such as movement, agreement, control, and scope. For instance, wh-movement in questions typically involves displacement from the specifier of VP to the specifier of CP. Similarly, embedded clauses that serve as arguments of verbs (e.g., *believe that*, *think that*) require CPs to function as complements to verb heads, further illustrating the clausal projection beyond the VP.

The CP-TP-VP architecture also supports cross-linguistic generalizations. In both English and Indonesian, the structure of embedded clauses, complementizer positions, and clause-typing markers (e.g., declarative vs. interrogative) can be consistently modeled through this layered structure. While surface variation exists, such as the optionality of certain functional elements or the absence of overt subject agreement, the deep syntactic principles articulated by X-bar theory remain robust across typologically diverse languages (Chomsky, 1986; van Valin, 2012).

In summary, clauses in natural language are built through a series of nested projections, with CP occupying the highest functional layer. This architecture allows for a modular and recursive system of sentence construction, whereby heads like C and T project their own phrases and organize complements and specifiers in consistent structural configurations. Understanding the roles of CP and TP within the syntactic hierarchy enables students and linguists to explain the position of subjects, the nature of subordination, and the syntactic properties of embedding across languages like English and Indonesian.

D. Drawing Trees with X-bar Notation

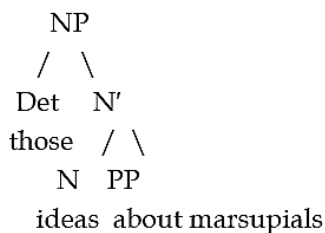
While syntactic theory is grounded in abstract principles, the visualization of structure is essential for both analysis and pedagogy. Tree diagrams offer a concrete representation of hierarchical relationships between constituents, and they are especially powerful when constructed using X-bar notation. This notational system, introduced by Jackendoff (1977) and widely adopted in generative syntax, provides a uniform structural template that represents how heads, complements, adjuncts, and specifiers combine to form phrases. By drawing syntactic trees with X-bar notation, students can better grasp the recursive and layered nature of phrase structure, moving beyond linear word order to the underlying architecture of syntactic units.

At the core of the X-bar model is the idea that each lexical category (N, V, A, P) projects a three-tiered hierarchy: a minimal projection (X), an intermediate projection (X'), and a maximal projection (XP). Each projection encodes a specific layer of syntactic composition:

1. The head (X) is the lexical nucleus of the phrase.
2. The complement is merged directly with the head to form an intermediate projection (X').
3. An adjunct, if present, is adjoined to the intermediate projection, forming a new X'.
4. The specifier is merged with the highest X' to form the maximal projection (XP).

To illustrate this structure, consider the English noun phrase:
those ideas about marsupials

In this phrase, *ideas* is the head noun (N), *those* is a determiner functioning as the specifier (Det), and *about marsupials* is a prepositional phrase (PP) that functions as a complement or adjunct, depending on whether the noun *ideas* semantically selects the PP. The structural tree under X-bar theory is as follows:



This tree demonstrates that the noun *ideas* projects a structure wherein the PP *about marsupials* is closely tied to the head. The determiner *those*, which provides specificity and definiteness, occupies the specifier position of the NP.

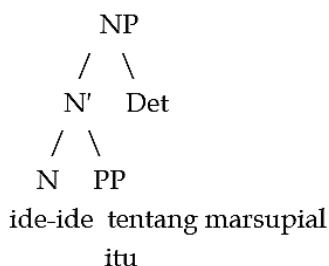
A comparable structure exists in Indonesian:

ide-ide itu tentang marsupial

Here, *ide-ide* is the reduplicated noun (*ide* meaning "idea"), which indicates plurality through morphological reduplication.

The word *itu* (“that” or “those”) serves as a demonstrative determiner and occurs in a post-nominal position, a characteristic feature of Indonesian NP structure (Adelaar & Himmelmann, 2004; Sneddon et al., 2010). The PP *tentang marsupial* functions similarly to its English counterpart, offering additional specification regarding the content or topic of the ideas.

This structure may be represented as:

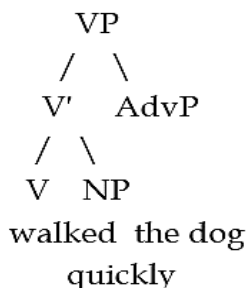


In this representation, the NP differs from the English one in terms of surface order, but the underlying syntactic roles remain consistent. Indonesian allows post-nominal determiners such as *itu*, which are analyzed as specifiers despite their linear position. Linguists like Van Gelderen (2002) and Musgrave (2001) have argued that such positional differences are surface phenomena and should not obscure the universal syntactic roles played by heads, complements, and specifiers.

The tree structures above reflect the binary branching constraint of X-bar theory, which ensures that every non-terminal node branches into two constituents. This constraint aligns with both linguistic economy and computational parsability, supporting models of human sentence processing and machine parsing (Chomsky, 1995; Radford, 2012). Moreover, drawing such trees provides learners with tools for diagnosing constituent structure, testing phrase boundaries, and understanding how syntactic movement operates in more complex constructions.

Tree diagrams in X-bar notation are not limited to noun phrases; they apply equally to other phrasal categories, such as

VP, AP, PP, and CP. For instance, a verb phrase might be represented as:



This recursive architecture makes X-bar theory particularly powerful and versatile for modeling linguistic phenomena such as coordination, ambiguity, and ellipsis. It also lays the groundwork for transformations, such as wh-movement and passivization, which rely on clear structural representations.

In sum, drawing trees using X-bar notation provides a rigorous yet accessible method for visualizing phrase structure. These trees encode the hierarchical, constituent-based organization of natural language and reveal how superficially linear strings are built from recursive and ordered layers. The comparison of English and Indonesian confirms that, despite typological variation, both languages conform to the same underlying syntactic principles, demonstrating the cross-linguistic validity of the X-bar framework.

E. Summary

This unit introduced X-bar theory as a unified model of phrase structure that refines traditional phrase structure rules by specifying a three-level hierarchy: the head (X), an intermediate projection (X'), and the maximal projection (XP). All phrasal categories, whether noun phrases, verb phrases, adjective phrases, or prepositional phrases, are analyzed using this schema. The unit explained how complements, specifiers, and adjuncts attach to these layers in consistent and predictable ways. Using both English and Indonesian examples, it

demonstrated how phrases can vary in surface word order while adhering to a shared hierarchical structure. The unit also explored how X-bar theory supports the construction of syntactic trees that reflect constituency and recursive structure. Tree-drawing exercises were used to illustrate how syntactic relations like head-complement and specifier-head are represented visually and structurally. This framework forms the basis for more advanced theories of sentence architecture, including the derivation of full clause structures in the next units.

GLOSSARY

- **Adjunct** – Optional modifier attached recursively.
- **Bare Phrase Structure** – Minimalist variant without bar levels.
- **Complement vs. Adjunct Tests** – Obligatoriness, selection, iterability.
- **Functional Projections** – CP, TP, vP, etc. layered above VP.
- **Head/Specifier/Complement** – Core roles in phrase structure.
- **Headedness** – Head-initial vs. head-final ordering.
- **Projection** – Head $\rightarrow$ X' $\rightarrow$ XP (levels of structure).
- **X' (X-bar) Schema** – Universal phrase template: XP $\rightarrow$ Spec X'; X' $\rightarrow$ X (Comp) (Adjunct).

UNIT 5

STRUCTURAL RELATIONS IN SYNTAX

A. Domination, Precedence, and C-command

In formal syntax, especially within generative grammar, understanding structural relations among constituents is essential for analyzing sentence structure, explaining syntactic movement, and accounting for interpretive phenomena such as binding, scope, and ellipsis. Among the most central structural relations are domination, precedence, and c-command. These relations are not defined in terms of linear word order alone, but in terms of hierarchical organization as represented in syntactic tree structures.

1. Domination

A node A dominates node B if and only if A is higher up in the tree and there is a continuous downward path from A to B. Domination is a transitive relation: if A dominates B, and B dominates C, then A also dominates C. This relation is crucial for understanding constituent structure and for defining other relations, such as c-command. For example, in a sentence like *The fish near the robot floated*, the NP *the fish near the robot* dominates all of its internal components, including *near the robot* and *robot* itself. Similarly, the TP dominates the VP and all its constituents.

In Indonesian, a parallel structure such as:

Ikan di dekat robot itu mengapung.

Shows the same kind of hierarchical structure. The subject NP *ikan di dekat robot itu* dominates the PP *di dekat robot itu*, just as in English, reinforcing the cross-linguistic applicability of domination as a syntactic relation (Adger, 2003; Radford, 2012; Sneddon et al., 2010).

2. Precedence

A node A precedes node B if A and B are sisters (i.e., they share the same mother node), and A appears to the left of B in linear order. More broadly, precedence can be extended to all cases where A comes before B in surface structure, provided A does not dominate B nor is dominated by B. It is important to distinguish precedence from linear order alone; precedence is defined within tree structures and is used in formulating syntactic constraints.

For instance, in *The fish near the robot floated*, within the VP, the verb *floated* follows the subject NP and thus does not precede it structurally. However, in linear order, *floated* comes after *robot*, but because *floated* and *robot* are in separate constituents, *robot* does not precede *floated* in the syntactic sense relevant to rule application or movement constraints.

In Indonesian:

Ikan di dekat robot itu mengapung.

The linear order is parallel, and syntactic precedence operates similarly. Within the PP *di dekat robot itu*, the noun *robot* is preceded by the preposition *di*, which is its sister node in the PP projection, satisfying the structural definition of precedence in a localized domain (Adger, 2003; van Valin, 2012).

3. C-command

Among all structural relations in syntax, c-command is perhaps the most powerful and widely used. Originally introduced by Reinhart (1976) and formalized further in Chomsky (1993), c-command plays a key role in determining binding domains, licensing ellipsis, constraining movement, and governing scope interpretations.

Formally, node A c-commands node B if:

- a. A does not dominate B, and
- b. The first branching node that dominates A also dominates B.

In essence, c-command defines a symmetric, lateral relationship within a tree, operating between nodes that are not necessarily in a dominance relation but share a common ancestor. For example, in the structure:

The fish near the robot floated.

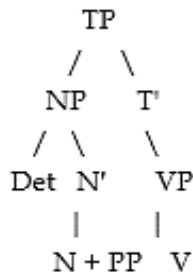
The subject NP *the fish near the robot* c-commands the VP *floated*. Within that NP, the noun *fish* c-commands the prepositional phrase *near the robot* only if they are sisters under a shared N' node. More importantly, the entire subject NP c-commands out of its domain into the TP, which is relevant for binding theory (discussed in Unit 11).

Similarly, in Indonesian:

Ikan di dekat robot itu mengapung.

The NP *ikan di dekat robot itu* c-commands the VP *mengapung*. This relation becomes crucial when evaluating how pronouns, reflexives, or quantifiers interact within a sentence. For example, if a reflexive pronoun must be c-commanded by its antecedent to be interpreted correctly (Principle A of binding theory), c-command provides the structural condition for that interpretation.

The following abstract tree structure demonstrates c-command visually in both English and Indonesian (simplified for illustration):



In this configuration:

- a. The NP (subject) c-commands the VP.
- b. The determiner and the noun do not c-command each other, as they are not in the required branching relationship.
- c. Within the PP, the preposition c-commands its complement noun, but the reverse is not true.

The theoretical utility of c-command is vast. It underlies the definition of syntactic binding domains, affects scope interpretation of quantifiers (e.g., *everyone didn't leave*), and is used in formulating constraints on movement such as island constraints. Cross-linguistic research (Carnie, 2012; O'Grady et al., 2009) has shown that c-command operates universally, even in languages with differing word orders and morphological structures, suggesting that hierarchical structure, not linear sequence, is primary in human syntactic cognition.

In conclusion, the structural relations of domination, precedence, and c-command form the backbone of syntactic theory. They allow linguists to represent and explain how constituents are organized in the mental grammar of speakers, and they serve as the foundation for more advanced theoretical constructs such as binding theory, movement, and scope. Understanding these relations enables students to move beyond surface order and appreciate the abstract architecture of syntax that underlies linguistic variation.

B. Syntactic Binding and Scope

The notion of syntactic binding is central to the theory of anaphora and plays a critical role in understanding how reference is established within sentence structure. Binding involves a structural relationship between two expressions, typically a pronoun or anaphor and its antecedent, such that the interpretation of one element (the bound element) depends on the other (the binder). This relationship is structurally

constrained and defined not merely by linear proximity but by hierarchical relations, especially c-command and coindexation.

1. Formal Definition of Binding

A binding relation is established under the following conditions (adapted from Chomsky, 1993; Reinhart, 2016):

An element A binds element B if:

- a. A c-commands B, and
- b. A and B are coindexed (i.e., they refer to the same entity).

The concept of c-command, introduced in Section A, ensures that binding occurs within structurally defined domains rather than arbitrary distances. Binding is foundational to the three principles of Binding Theory, which will be treated in more depth in Unit 11. For now, the focus is on how syntactic structure constrains the reference of reflexives, pronouns, and quantified expressions.

2. English Example: Reflexive Binding

Consider the English sentence:

Lara painted herself.

In this example, the reflexive pronoun *herself* is interpreted as referring to *Lara*. Structurally, the NP *Lara* c-commands the reflexive *herself* within the verb phrase and they are coindexed, satisfying the conditions for syntactic binding. The reflexive is grammatically licensed because it appears within the correct binding domain, typically the smallest clause or VP that contains it and a c-commanding antecedent. If we substitute a non-reflexive pronoun (*Lara painted her*), the sentence would suggest that Lara painted someone else, not herself, unless co-reference is forced by context or disambiguated by prosody.

3. Indonesian Equivalent: Reflexive Binding

A parallel structure can be found in Indonesian:

Lara melukis dirinya.

Here, *dirinya* is a reflexive pronoun formed from *diri* (“self”) and the enclitic *-nya*, which adds definiteness or third-person reference. Like English *herself*, *dirinya* is interpreted as referring back to the subject *Lara*, and the binding relation is licensed by the same structural conditions. Indonesian reflexives are subject to similar constraints of locality and c-command: the antecedent must appear within the appropriate syntactic domain and must c-command the reflexive (Musgrave, 2001; Sneddon et al., 2010).

The tree structure of both sentences would show that the subject NP c-commands the VP and hence all of its internal constituents, including the reflexive object. This shared configuration across typologically distinct languages like English and Indonesian suggests that binding is governed by universal structural principles, even when morphosyntactic expressions differ.

4. Scope and Binding Interactions

Binding theory is also closely related to the notion of scope, especially when pronouns and quantifiers interact. Scope refers to the syntactic and semantic domain over which a quantifier or operator has influence. For example, in the English sentence:

Every student thinks she will succeed.

This sentence is ambiguous. The pronoun *she* may refer to the same person in every quantifier case (bound reading), or it may refer to a specific individual (referential reading). Only under the bound variable reading does *every student* bind the pronoun *she*, meaning that for each student, she believes that she herself will succeed. This reading requires that *every student* c-command and be coindexed with *she*, illustrating how syntactic structure affects interpretive possibilities (Heim, 1998; Reinhart, 2016).

In Indonesian, a similar scope ambiguity can be constructed: *Setiap siswa berpikir bahwa dia akan berhasil.*

Here, *dia* ("he/she") can either refer to each student individually (bound variable interpretation) or to a single individual outside the quantifier's scope (referential interpretation). The ambiguity hinges on whether *setiap siswa* c-commands and binds *dia*. The syntactic configuration of CP and TP layers, as discussed in Unit 4, determines whether such a binding relation is licensed.

5. Reflexive vs Non-reflexive Pronouns

It is important to distinguish reflexive binding from pronominal co-reference. In:

Lara said that she would win.

She can refer to Lara (co-referential reading) or to another person. However, in:

Lara said that herself would win.

His sentence is ungrammatical, because *herself* is a reflexive and must be bound locally within the same clause. The failure here results from violating Principle A of binding theory (to be discussed in Unit 11), which requires that reflexives be bound in their local domain. Indonesian shows a similar pattern:

Lara berkata bahwa dirinya akan menang.

This sentence is grammatical and acceptable, as *dirinya* is used in a non-local binding context. Indonesian reflexives like *dirinya* appear more flexible than English reflexives in terms of locality, raising interesting questions about the parametric variation in binding domains cross-linguistically (Cole et al., 2006; Safir, 2004).

C. Representing Structural Relations in Trees

The accurate representation of structural relations in tree diagrams is not only a pedagogical tool but also a fundamental aspect of syntactic theory. Trees make visible the hierarchical configuration of constituents, thereby revealing the grammatical dependencies and interpretive constraints that cannot be captured through linear word order alone. Among these

dependencies, c-command plays a central role in determining the interpretation of pronouns and quantifiers, especially in cases of scope ambiguity.

Scope ambiguity arises when the hierarchical relationship between a quantifier and a pronoun allows for multiple readings. This phenomenon is particularly well-illustrated in sentences where a quantificational noun phrase appears to bind a possessive pronoun within another constituent. The ambiguity emerges from the fact that c-command is a structural, not linear, relation, two noun phrases may appear in a particular order, but only when one c-commands the other can binding occur.

1. English Example: Quantifier-Pronoun Ambiguity

Consider the English sentence:

Every guard watched his reflection.

This sentence is ambiguous between two interpretations:

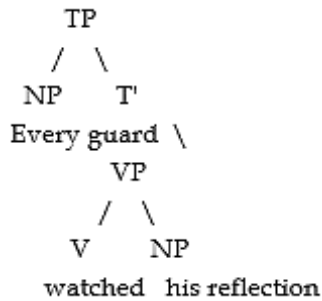
- a. Bound variable reading: Each guard watched his own reflection.
- b. Coreferential reading: All guards watched the same reflection, associated with a specific male individual.

The first reading arises when *every guard* c-commands and binds the possessive pronoun *his*. In tree terms, this occurs when the quantificational NP *every guard* is in the subject position of TP, and *his reflection* is part of the object NP inside the VP. Since the subject c-commands the object, and coindexation is possible, binding is licensed.

The second reading emerges when *his* is interpreted as a referential pronoun, linked to an individual in the discourse, not structurally bound by *every guard*. This difference in interpretation is a result of how scope and c-command interact in the syntactic structure (Heim, 1998; Radford, 2012; Reinhart, 2016).

2. Visualizing in X-bar Trees

In an X-bar tree representation, we can visualize this structure as follows:



In this structure:

- The NP *every guard* c-commands the entire VP.
- The NP *his reflection* is embedded within the VP and thus is within the scope of *every guard*.
- C-command from *every guard* to *his* allows the bound variable reading.

This visual representation demonstrates how the structural configuration enables or blocks interpretive possibilities, making the binding conditions transparent.

3. Indonesian Parallel: Scope and Ambiguity

A syntactically and semantically parallel example exists in Indonesian:

Setiap penjaga melihat bayangannya.

Here, *setiap penjaga* is a quantificational subject NP ("every guard"), and *bayangannya* is a possessive NP ("his reflection"). Like its English counterpart, this sentence is ambiguous between:

- A distributive interpretation: each guard saw his own reflection (bound variable).
- A referential interpretation: all guards saw someone else's reflection.

The structural explanation is the same. The subject NP *setiap penjaga* c-commands the VP containing the object *bayangannya*. Assuming coindexation between *setiap penjaga* and the possessive suffix *-nya*, the bound variable interpretation becomes available. If, however, *-nya* refers to someone else in the discourse, the referential interpretation results.

This supports the idea that c-command governs the scope of quantifiers across languages, even when the morphological realization of possession differs. In Indonesian, possession is marked by the enclitic *-nya* rather than a free-standing pronoun, but the structural relations remain intact (Landau, 2000; Musgrave, 2001; Sneddon et al., 2010).

4. Implications for Syntax and Semantics

Scope ambiguity highlights the importance of syntactic trees as interpretive mechanisms. The same string of words can yield different meanings based on underlying hierarchical relations. Furthermore, this phenomenon underscores that syntactic form constrains semantic interpretation, particularly through structural tools like c-command and binding domains. Linguists use these representations not only to explain ambiguity but also to model how native speakers assign meaning to complex structures in real-time processing.

Cross-linguistic studies (e.g., O'Grady et al., 2009; Reinhart, 2016) show that despite variations in morphology and surface word order, languages consistently rely on hierarchical constituency to regulate scope and reference. This universality lends empirical support to the idea that structural principles such as c-command are part of Universal Grammar, a notion central to the generative tradition.

D. Summary

This unit introduced key structural relations that underpin syntactic theory, including domination, precedence, and c-command. These relations are central to understanding how syntactic elements interact within tree structures and are necessary for explaining phenomena such as binding, scope, and movement. Domination refers to hierarchical containment in a syntactic tree, while precedence captures the linear ordering of elements. C-command, the most critical of these relations, was defined as a condition where one node structurally governs another within a branching structure. The unit demonstrated how c-command determines the availability of syntactic operations and influences the interpretation of pronouns and anaphors.

Building on these structural foundations, the unit explored binding theory, particularly how anaphors like *himself* must be bound within a specific domain, and how pronouns must remain free. English examples were used to illustrate how incorrect structural configurations lead to ungrammaticality, while Indonesian data showed that binding constraints also operate across languages, albeit with surface differences in form.

The unit also examined scope relations, such as quantifier scope ambiguity, and how these depend not on word order alone but on the hierarchical configuration of constituents. For instance, ambiguity in sentences like *Every guard watched his reflection* arises from multiple possible c-command relations, leading to different interpretations.

By emphasizing structural relations rather than surface patterns, the unit underscored the importance of hierarchical syntax in explaining core linguistic behavior. Tree diagrams were employed to illustrate how these abstract concepts are visually represented, allowing linguists to model derivational steps and test theoretical claims. Overall, the unit provided a foundation for understanding how structural configurations

govern grammaticality and interpretation across diverse syntactic environments.

GLOSSARY

- **Binding** – Relations among anaphors, pronouns, R-expressions.
- **C-command** – Structural relation crucial for binding/scope.
- **Domination/Immediate Domination** – Hierarchical inclusion in trees.
- **Governing Category** – Domain relevant in GB-era constraints.
- **Locality** – Structural “closeness” required for dependencies.
- **Precedence** – Linear ordering of sisters.
- **Scope** – Semantic range of operators (negation, quantifiers).
- **Sisterhood** – Nodes sharing the same mother.

UNIT 6

CLAUSAL ARCHITECTURE

A. TP, CP, and Functional Layers

In generative syntax, a clause is not merely a linear string of words but a hierarchically layered structure that contains multiple levels of grammatical information. These layers are projected through functional categories, each of which serves a distinct syntactic and interpretive function. Among the most prominent of these categories are the Tense Phrase (TP) and the Complementizer Phrase (CP). Together, they represent the backbone of clause structure, providing the scaffolding upon which the rest of the syntactic derivation is built (Adger, 2003; Chomsky, 1986; Radford, 2012).

The TP and CP are part of what is often referred to as the clausal spine, a series of projections that sit above the Verb Phrase (VP) and host essential grammatical information such as tense, aspect, modality, agreement, and clause type. These projections are typically not filled by lexical material (like nouns or verbs), but by functional heads that encode grammatical features and license syntactic operations such as movement, agreement, and embedding.

1. The Tense Phrase (TP)

The TP is the projection of the Tense head (T), which is responsible for anchoring the event described in the clause to a point in time. In English, T can be realized overtly through inflected auxiliaries (e.g., *will*, *has*, *did*) or through bound tense morphology on the main verb (e.g., *walked*, *eats*). The

specifier of TP (Spec-TP) typically hosts the subject, while the complement of T is the Verb Phrase (VP) that encodes the predicate and its arguments. The internal structure of TP can be schematized as follows:

- a. $TP \rightarrow NP\ T'$
- b. $T' \rightarrow T\ VP$

Consider the English sentence:

She will leave.

Here, *she* is the subject in Spec-TP, *will* is the T head expressing future tense, and *leave* is the lexical verb in the VP. The TP structure encodes the propositional content of the sentence along with its temporal anchoring.

In Indonesian, the TP is realized differently but serves the same syntactic function. For example:

Dia akan pergi.

In this sentence, *dia* ("she/he") is the subject, and *akan* is a modal particle expressing futurity, equivalent in function to English *will*. *Pergi* ("go/leave") is the main verb. Although Indonesian lacks tense inflection on the verb itself, the presence of temporal and aspectual markers such as *sudah* (perfective), *sedang* (progressive), and *akan* (future) indicates the presence of a TP-like structure (Sneddon et al., 2010). These markers can be analyzed as occupying the T position or functioning as auxiliaries merged with T.

2. The Complementizer Phrase (CP)

Sitting at the top of the clause is the CP, which is headed by a complementizer (C). The CP introduces embedded clauses, marks clause type (e.g., declarative, interrogative, relative), and serves as the target for certain syntactic movements, such as *wh*-movement and topicalization. The internal structure of the CP can be represented as:

- a. $CP \rightarrow (Spec)\ C'$
- b. $C' \rightarrow C\ TP$

The head C may be overt, as in English *that*, *if*, or *whether*, or covert (null), especially in main clause declaratives. The specifier of CP (Spec-CP) is often the landing site for fronted elements in questions and relative clauses.

Take, for example, the English sentence:

I believe [that she will leave].

The embedded clause *that she will leave* is a CP in which *that* occupies the C position, introducing the embedded TP *she will leave*. This structure allows the embedded clause to function as the complement of the verb *believe*. Without the CP layer, it would be impossible to distinguish between matrix and embedded propositions or to account for different clause types.

A structurally equivalent sentence in Indonesian would be:

Saya percaya [bahwa dia akan pergi].

Here, *bahwa* is the Indonesian complementizer, analogous to English *that*. It introduces the embedded clause *dia akan pergi*, which is a full TP. Despite the fact that Indonesian does not require agreement or tense morphology on the verb, the structure mirrors English in terms of hierarchical layering. The presence of *bahwa* confirms that the embedded clause is a CP and functions as the syntactic object of *percaya* ("believe").

Notably, in both languages, CPs can host not only declaratives but also interrogatives and relatives. In English:

- a. *I asked [if she would come].*
- b. *The book [that she read] was excellent.*

In Indonesian:

- a. *Saya bertanya [apakah dia akan datang].*
- b. *Buku [yang dia baca] itu bagus.*

Here, *apakah* functions as a question-introducing complementizer, and *yang* serves as a relative marker, both of which occupy the C position. These examples further

illustrate that CP is a robust and flexible functional category across languages (Adger, 2003; Cole et al., 2006).

3. Functional Layering and Syntactic Modularity

The organization of clauses into functional projections like TP and CP reflects the modular architecture of syntax, where distinct types of grammatical information are handled in different structural zones. The VP handles argument structure and predicate formation, the TP manages temporal and modal features, and the CP governs clause typing and embedding. This modularity allows for a highly compositional syntax, where complex sentences can be built incrementally and transparently.

Furthermore, the layering of CP and TP is essential for the application of syntactic operations. For example, subject-auxiliary inversion in English questions involves movement of the auxiliary from T to C:

Will she leave?

In this construction, the auxiliary *will* raises from T to C, leaving the subject in Spec-TP. This movement is only possible because the clause is structured with separate TP and CP layers. Indonesian does not display such overt T-to-C movement, but this absence does not negate the underlying structure; it reflects parametric variation in how languages realize movement and clause types (Adger, 2003; Chomsky, 1995; Radford, 2012).

B. Sentence Types and Complementizers

In generative syntax, the Complementizer Phrase (CP) serves as the syntactic gateway for embedding clauses and distinguishing among various sentence types. The CP layer is headed by a complementizer (C) that encodes important information about the force of the clause, that is, whether it is a declarative, interrogative, imperative, or exclamative. These clause types are not merely distinctions in mood or illocutionary force; they are syntactically encoded through different functional

heads and projections, particularly within the CP domain (Adger, 2003; Radford, 2012; Rizzi, 1997).

The complementizer introduces the CP and governs the embedded TP. In English, overt complementizers such as *that*, *if*, *whether*, and *for* signal different clause types and embedding contexts. In Indonesian, parallel complementizers such as *bahwa*, *apakah*, and *kalau* (among others) fulfill equivalent grammatical functions. Despite typological differences, both languages use CP structures to introduce subordinate clauses and to mark the distinction between statement and question, making CP a robust cross-linguistic projection.

1. Declarative CPs

Declarative clauses represent statements or propositions. In embedded contexts, English uses the complementizer *that* to introduce a declarative clause. Consider the example:

He said [that she left].

Here, *that* occupies the C position and introduces the embedded TP *she left*. The verb *said* selects for a CP complement, and the structure can be formally represented as:

- a. $CP \rightarrow C\ TP$
- b. $C = \text{that}$
- c. $TP = \text{she left}$

In this configuration, the CP functions as the direct object of the verb *said*, illustrating that not only noun phrases but also entire clausal projections can serve as arguments of verbs.

In Indonesian, a similar structure is observed:

Dia bilang [bahwa dia pergi].

Here, *bahwa* serves as the complementizer introducing the embedded declarative. The clause *bahwa dia pergi* ("that she/he left") functions as the object of the verb *bilang* ("say"). As in English, the embedded CP provides the propositional content of the matrix predicate. Although *bahwa* is often

considered more formal or literary than simply embedding a clause without an overt complementizer, its syntactic role is identical to that of English *that* (Musgrave, 2001; Sneddon et al., 2010).

2. Interrogative CPs

Interrogative clauses pose questions and are structurally distinct from declaratives in that they may involve movement (in the case of *wh*-questions) or specific interrogative complementizers. In embedded interrogatives, English typically uses *if* or *whether*.
I wonder [if she left].

In this example, *if* is a complementizer that introduces an embedded yes-no question. Unlike main clause interrogatives, embedded questions do not exhibit subject-auxiliary inversion in English. The complementizer *if* functions as the C head of the CP and indicates that the embedded clause is interrogative in force. The verb *wonder* subcategorizes for an interrogative CP.

Indonesian marks embedded interrogatives using *apakah*, a particle that introduces yes-no questions:
Saya bertanya-tanya [apakah dia pergi].

Here, *apakah* occupies the C position and indicates that the embedded clause is interrogative. The embedded clause *apakah dia pergi* (“whether she/he left”) is the syntactic object of the matrix verb *bertanya-tanya* (“wonder” or “ponder”). As in English, no subject-predicate inversion occurs, and the clause retains declarative word order, underscoring that CP, not surface order, determines clause type (Adger, 2003; Cole et al., 2006).

3. Clause Typing and Force Projection

More recent developments in syntactic theory, particularly within the Cartographic Framework (Rizzi, 1997; Rizzi & Cinque, 2016), propose that CP is not a monolithic projection but rather a complex structure composed of multiple layers, such as ForceP, TopicP, and FinitenessP.

Within this view, the Force head at the top of the CP encodes the clause type (declarative, interrogative, etc.), while other projections manage discourse-related elements and syntactic licensing.

Although these finer-grained projections are not always overt in Indonesian or English, their effects are visible in sentence types and movement possibilities. For instance, only interrogative CPs allow *wh*-elements to move into Spec-CP, and only CPs marked for subordination can serve as complements to verbs.

4. Implications for Syntax

Understanding the relationship between sentence types and complementizers is crucial for multiple areas of syntactic inquiry:

- a. Selectional properties: Verbs differ in the types of CPs they can take (*wonder* selects interrogative CPs; *believes* selects declarative CPs).
- b. Clause-typing: Complementizers serve as overt markers of syntactic force, guiding both interpretation and structure.
- c. Syntactic movement: The CP layer provides targets and landing sites for elements undergoing displacement (e.g., in *wh*-questions or topicalization).

Cross-linguistically, the presence of CP projections, whether overtly realized or covert, indicates that clause typing is a universal syntactic phenomenon, despite differences in morphology and word order.

C. The Clause as a Syntactic Unit

In generative grammar, the clause is treated not merely as a string of words or a vehicle for conveying complete thoughts, but as a syntactic unit, a phrasal projection that conforms to structural principles and occupies a specific position within larger syntactic derivations. A clause is typically defined as a projection of TP (Tense Phrase) and, when embedded, as a projection of CP (Complementizer Phrase). As such, it

functions as a syntactic constituent that can serve various grammatical roles, including subject, object, predicate complement, or adjunct. This treatment allows syntacticians to analyze not only simple sentences but also complex and recursive structures, as observed in subordination, coordination, and relative clauses.

1. Clauses as Arguments

One of the clearest indicators that clauses function as syntactic units is their ability to occupy argument positions. In the following English example:

It is strange [that Mira found the tunnel].

The bracketed clause *that Mira found the tunnel* is a CP headed by the complementizer *that*. It is the logical subject of the sentence, even though the surface subject position is filled by the expletive *it*. This kind of extraposition construction is common in English and reflects the interaction between syntax and discourse: the full clausal subject is heavy and is therefore displaced rightward for processing ease, while *it* functions as a placeholder in Spec-TP.

The embedded clause carries propositional content and serves as the complement of the adjective *strange*. Such constructions exemplify that clauses can function similarly to noun phrases (NPs), i.e., as complements or subjects of predicates. Syntactically, this CP is selected by the matrix predicate and licensed within the overall TP structure of the main clause.

In Indonesian, an equivalent construction exists:

Aneh [bahwa Mira menemukan terowongan itu].

Here, *bahwa Mira menemukan terowongan itu* ("that Mira found the tunnel") functions as the logical subject of the sentence, introduced by the complementizer *bahwa*. Indonesian typically does not employ an expletive subject like English *it*, and instead allows for the CP to surface in clause-initial position. Nonetheless, both structures convey the same semantic and syntactic configuration, reinforcing

the parallelism between English and Indonesian clause architecture (Musgrave, 2001; Sneddon et al., 2010).

This similarity supports the claim, advanced by Chomsky (1993, 1995), that all human languages are built from shared abstract structures, which are parametrically varied in surface realization but universally grounded in Universal Grammar (UG). The clause as a unit is one such invariant structure, projected by functional heads like T (Tense) and C (Complementizer) and composed hierarchically through Merge operations.

2. Clausal Embedding and Recursion

Clauses are not only syntactic units but also recursively embedded structures, capable of being nested within each other indefinitely. Consider the following sentence:

It surprised me [that she claimed [that he said [that the tunnel was hidden]]].

In this construction, multiple layers of CPs are embedded, each serving as the complement of the verb in the next higher clause. This phenomenon, referred to as recursive embedding, illustrates that clauses are self-similar syntactic objects, they can be embedded within other clauses of the same type without limit. Theoretically, recursion is taken to be a defining feature of human language (Fitch et al., 2005; Hauser et al., 2002), and the clause is the minimal domain in which this recursive property operates.

In Indonesian, similar recursive embedding is possible:

Saya terkejut [bahwa dia mengklaim [bahwa dia mengatakan [bahwa terowongan itu tersembunyi]]].

Each *bahwa*-headed clause occupies a CP position and serves as the complement to the verb in the higher clause. The ability to embed CPs within CPs confirms that clauses in Indonesian, like in English, are syntactically and semantically compositional units, each one contributing to the meaning and structure of the larger expression. Despite differences in linear structure, agreement, and prosody, the hierarchical

properties of clause formation remain consistent (Cole et al., 2006; Musgrave, 2001).

3. Clauses as Constituents: Diagnostic Tests

To confirm the constituency status of clauses, syntacticians apply several diagnostics:

- a. Substitution: The clause can be replaced by a pronoun or pro-form such as *it* in English.

That Mira found the tunnel surprised me. → *It surprised me.*

- b. Movement: The clause can undergo topicalization or extraposition.

[That Mira found the tunnel], I find surprising.

- c. Coordination: The clause can be coordinated with another clause.

He said [that she left] and [that he would follow].

Indonesian demonstrates the same behaviors:

- a. Substitution is possible through ellipsis or contextual inference.
- b. Topicalization is common through fronting constructions.
- c. Coordination is productive, often with *dan* ("and") joining CPs.

These diagnostics further support the unit status of clauses across languages.

4. The Clause in Theoretical Perspective

From a theoretical standpoint, the clause plays a crucial role in defining the boundaries of syntactic domains, such as binding domains (e.g., for reflexives), islands for movement (e.g., wh-islands), and scope domains (e.g., for quantifiers). Clauses are cyclic domains in the derivation, meaning that operations such as movement and agreement apply locally within clauses before proceeding to higher structures. This cyclicity is central to modern approaches in Minimalist Syntax, which posit that derivations proceed phase by phase, with CP and vP often treated as phases, domains of syntactic computation that are passed on to the interfaces (Adger, 2003; Chomsky, 2001).

D. Summary

This unit examined the internal structure of clauses and their role as fundamental syntactic, semantic, and computational units across languages. In both English and Indonesian, clauses can function as arguments, undergo embedding, and participate in structural operations such as movement, coordination, and substitution. These patterns provide strong support for the theory of Universal Grammar, which posits that all human languages share a common blueprint of clausal projections such as TP (Tense Phrase) and CP (Complementizer Phrase). While the morphological and lexical realizations of these projections vary across languages, their structural and functional roles remain consistent.

A key focus of the unit was the clause's layered architecture, particularly the role of the TP and CP domains. TP encodes tense and agreement features and serves as the syntactic locus for subjects, while CP introduces embedded clauses and determines clause type. The CP layer is headed by complementizers, which vary across languages in form but not in function. For example, declarative complementizers such as *that* in English and *bahwa* in Indonesian introduce embedded statements, while interrogative complementizers like *if* and *apakah* mark subordinate questions. These elements contribute to the clause's syntactic identity and mediate the interface between syntax and semantics.

Understanding the clause as a hierarchically organized unit with nested projections is essential for explaining a wide range of syntactic phenomena, including clause embedding, subject-auxiliary inversion, topicalization, and long-distance dependencies. Although English and Indonesian differ in their use of morphological agreement and word order flexibility, they both exhibit clause-typing mechanisms that rely on similar structural principles. Thus, the analysis of clausal architecture not only illuminates cross-linguistic variation but also reinforces the universality of the syntactic categories and operations that define sentence structure in natural language.

GLOSSARY

- **Clause as Unit** – Syntactic domain for movement/agreement.
- **Clause Types** – Declarative, interrogative, imperative, embedded.
- **Complementizers** – that, if, whether; Indonesian: bahwa, jika/kalau.
- **CP (Complementizer Phrase)** – Clause type and force (that, if).
- **Subject Position (Spec-TP)** – Canonical subject landing site.
- **TP (Tense Phrase)** – Tense/agreement position for the clause.
- **Voice** – Active/passive alternations in structure.
- **vP** – Light verb phrase hosting external argument (agent).



PART III:
ARGUMENTATION IN SYNTAX



UNIT 7

AUXILIARIES AND STRUCTURAL EVIDENCE

A. Auxiliaries as Heads: Evidence from Ellipsis

One of the key insights from generative syntax is that auxiliaries do not merely precede main verbs in surface word order, but function as structural heads of their own syntactic projections. Specifically, auxiliaries are generally analyzed as occupying the head of functional categories such as Tense Phrase (TP), Aspect Phrase (AspP), or Modality Phrase (ModP), depending on the features they express. These heads govern the verbal phrase (VP) and play a crucial role in tense, aspect, modality, and subject-verb agreement. Their structural status is not just theoretical but is empirically observable through a range of syntactic phenomena, including ellipsis.

One particularly robust diagnostic for the head status of auxiliaries is VP ellipsis. In ellipsis constructions, part of a clause, typically the VP, is omitted under identity with a previously mentioned VP. The fact that auxiliaries remain overt in such constructions, while the lexical verb and its complements are elided, provides compelling evidence that auxiliaries are positioned outside the VP, and thus function as heads within the syntactic hierarchy (Adger, 2003; Carnie, 2012; Radford, 2012).

Consider the following English example:

Kai has eaten the noodles, and Nia has too.

In the first clause, the full VP *eaten the noodles* appears. In the second clause, however, the VP is elided (omitted) but understood. The auxiliary *has* remains overt in both clauses. The

only way this sentence remains grammatical and interpretable is if the auxiliary *has* occupies a higher syntactic position, specifically, the head of Tense Phrase (T), while the VP it governs is structurally subordinate and therefore deletable. The structural representation is roughly as follows:

1. $TP \rightarrow NP\ T'$
2. $T' \rightarrow T\ VP$
3. $T = \text{has}$
4. $VP = \text{eaten the noodles (overt in first clause; elided in second)}$

This ellipsis pattern strongly suggests that the auxiliary *has* is not part of the VP but rather a functional head that governs it. Its ability to remain in the clause while the VP is deleted supports its higher position in the phrase structure.

In English, auxiliary verbs such as *have*, *be*, *will*, *can*, and *must* demonstrate this pattern consistently. The remaining clause after ellipsis contains only the auxiliary, which indicates that the auxiliary serves as the anchor for tense and agreement features, while the elided material corresponds to the lexical predicate and its internal arguments.

A similar structure can be observed in Indonesian, even though the language does not mark tense through verb inflection. Instead, Indonesian employs aspectual markers such as *sudah* (perfective), *sedang* (progressive), and *akan* (future), which function semantically and structurally as auxiliaries. These elements are analyzed as aspectual heads that license the VP and encode aspectual interpretation (Musgrave, 2001; Sneddon et al., 2010).

Take the following Indonesian example:

Kai sudah makan mie itu, dan Nia juga sudah.

Here, *sudah* functions as a perfective aspect marker, conveying that the action of eating has been completed. In the second clause, the VP *makan mie itu* ("eat the noodles") is omitted, but the sentence remains grammatical and interpretable. The elided structure is fully recoverable from the antecedent clause due to parallelism, a key condition in ellipsis licensing (Merchant, 2001). The fact that *sudah* remains overt

while the VP is omitted provides evidence that it occupies a structural position higher than the VP, functioning analogously to auxiliaries in English.

From a syntactic perspective, this indicates that Indonesian, like English, organizes its clausal architecture into functional layers, with aspectual heads such as *sudah* occupying the head position of an Aspect Phrase (AspP) or even Tense Phrase (TP) if reinterpreted in minimalist terms (Adger, 2003; van Gelderen, 2002). The structure is approximately:

1. TP or AspP $\rightarrow$ NP T'/Asp'
2. T'/Asp' $\rightarrow$ T/Asp VP
3. T/Asp = *sudah*
4. VP = *makan mie itu* (elided in second clause)

Thus, ellipsis provides empirical support for the hierarchical configuration of auxiliary elements in both languages. Whether expressed through inflection (as in English) or through lexical aspect markers (as in Indonesian), auxiliaries consistently occupy positions that govern the verbal domain, reinforcing their syntactic role as functional heads rather than modifiers or lexical elements.

In minimalist syntax, this analysis fits within the framework of feature checking and movement operations. Auxiliaries are introduced in functional projections that host tense, aspect, or modal features, which must be checked by moving the appropriate constituents (often the subject or the verb) to these positions. Ellipsis then applies at PF (Phonological Form), deleting the lower VP while preserving the structure above it.

The similarity in ellipsis patterns between English and Indonesian supports a key hypothesis in generative linguistics: universal principles of syntactic organization are shared across languages, even when surface realizations differ. Auxiliaries, regardless of their morphosyntactic manifestation, behave as clausal heads that dominate and license the verbal predicate.

B. Distribution and Ordering of Auxiliaries

Auxiliaries play a crucial role in expressing grammatical features such as modality, aspect, tense, and voice, and they do so by occupying distinct positions within a clause's hierarchical structure. In English and many other languages, these auxiliary elements appear in a fixed and semantically predictable order, reflecting the syntax-semantics interface and the layered architecture of the clause. The distribution of auxiliaries is not arbitrary; it adheres to strict syntactic rules that are aligned with universal patterns of functional projection (Adger, 2003; Cinque, 1999; Radford, 2012).

In English, auxiliaries typically occur in the following hierarchical order:

Modal > Perfect > Progressive > Passive > Main Verb

This order corresponds to a sequence of functional heads within the clause, each contributing distinct grammatical information. Consider the following example:

She might have been being observed.

This sentence includes a full auxiliary stack:

1. *Might* – expresses modality (epistemic or deontic)
2. *Have* – marks perfect aspect
3. *Been being* – represents progressive and passive voice
4. *Observed* – is the main verb in past participle form

Each auxiliary selects the next lower projection as its complement, forming a tightly nested structure. The rigidity of this ordering is confirmed by the ungrammaticality of reordering them:

**She has might been observed.* (ungrammatical due to incorrect ordering)

The strict sequence follows the universal hierarchy of functional projections, often discussed in the literature as a manifestation of the cartographic approach (Cinque, 1999; Rizzi & Cinque, 2016). In this view, each auxiliary corresponds to a unique functional head within the clause, such as ModP,

PerfP, ProgP, and PassP, each hosting its corresponding auxiliary and projecting in a fixed order.

1. Indonesian Comparison: Auxiliary Distribution and Flexibility

While Indonesian does not inflect verbs for tense or aspect morphologically, it uses a series of aspectual and modal particles that serve similar syntactic and semantic functions. These include:

- a. *Mungkin* – modal (equivalent to "might", "may")
- b. *Sudah* – perfective aspect (completed action)
- c. *Sedang* – progressive aspect (ongoing action)
- d. *Diamati* – passive participle (passive construction of *mengamati*)

In the following example, these auxiliary-like elements appear in a sequence that reflects the English ordering:

Dia mungkin sudah sedang diamati.

This translates as:

She might already be being observed.

Here, *mungkin* is a modal adverb, *sudah* indicates completion (perfective), *sedang* marks progressive action, and *diamati* is the passive form of *mengamati* ("to observe"). Although Indonesian does not require this ordering as rigidly as English does, native speaker intuitions still prefer sequences that mirror the semantic naturalness and scope relations of these elements (Musgrave, 2001; Sneddon et al., 2010).

Unlike English, Indonesian may permit some reordering or even omission depending on discourse context, prosody, and formality. However, when all auxiliaries are used together, the preferred linear order generally reflects the same hierarchical logic as English: modals scope over perfect, which scopes over progressive, which precedes the passive or lexical verb. For instance, reversing the order to *sudah mungkin sedang diamati* results in semantic awkwardness or ambiguity.

This correspondence supports the cross-linguistic hypothesis that auxiliary ordering is governed by a universal base hierarchy, even in languages where the functional projections are not as morphologically transparent or syntactically rigid. In minimalist terms, these elements correspond to functional heads merged in a specific order, with the differences in word order arising from parameterized movement operations or phonological realization (Adger, 2003; Chomsky, 1995).

2. Syntactic Representation and Scope Interpretation

From a structural perspective, each auxiliary corresponds to a head of its own projection, and these projections are nested hierarchically, not simply linearly. For English:

ModP → PerfP → ProgP → VoiceP → VP

Each head selects the projection below it as its complement. For example, the modal *might* selects a TP (or PerfP), *have* selects ProgP, and so on. In Indonesian, even though auxiliaries appear as free morphemes, they still function in a manner consistent with this structure, with *mungkin* taking widest scope, and *diamati* as the lowest verbal projection.

These structures are semantically compositional: the modal scopes over the perfect aspect, which scopes over the progressive, etc. The structural positions therefore determine not just form, but interpretation. This close interaction between syntax and semantics underscores why auxiliary ordering is syntactically significant and not merely stylistic or idiomatic.

3. Cross-Linguistic Universality and Language Typology

Studies in typological syntax (Cinque, 1999; Dryer, 2007) have shown that most languages respect similar auxiliary hierarchies, even if the auxiliaries are expressed differently, through inflection, particles, or lexical verbs. The sequence:

Modality > Tense/Perfect > Aspect > Voice > Lexical Verb

Has been found in a wide range of genetically unrelated languages. This suggests that the clausal spine is universal, and that auxiliaries are manifestations of deeper syntactic architecture.

Therefore, both English and Indonesian, despite their differences in morphology, adhere to underlying principles that govern the distribution and scope of auxiliary elements. Understanding these hierarchies helps illuminate how grammatical meaning is encoded in structure, and how syntax organizes complex predicate formation in a principled way.

C. Morphological Restrictions and Verb Forms

One of the defining characteristics of auxiliary verbs in English is that they impose strict morphological requirements on the main verbs that follow them. These restrictions are not arbitrary but derive from the syntactic and morphological structure of the clause, where auxiliaries and verbal heads project distinct categories that interact through selectional properties. In formal syntax, these properties are typically analyzed under the framework of feature checking, whereby the auxiliary licenses or selects a specific verbal complement form based on tense, aspect, or voice features (Adger, 2003; Radford, 2012).

In English, auxiliaries require particular verb forms according to their syntactic role in the auxiliary chain:

1. Modals (e.g., can, must, will) select for the bare infinitive form of the verb:
 - a. Correct: *She can dance.*
 - b. Incorrect: *She can dances.*
2. Perfect aspect auxiliaries (have/has/had) select the past participle form:
 - a. Correct: *They have eaten.*
 - b. Incorrect: *They have eat.*

3. Progressive aspect auxiliaries (be, is, was, etc.) require the present participle (-ing) form:
 - a. Correct: *She is walking.*
 - b. Incorrect: *She is walk.*
4. Passive constructions also use the auxiliary *be*, but it must be followed by the past participle:
 - a. Correct: *The cake was baked.*
 - b. Incorrect: *The cake was baking.*

These form restrictions are obligatory and unviolated in standard English syntax. They demonstrate that auxiliaries not only convey grammatical meaning but also exert morphosyntactic control over the verbal phrase, effectively selecting the form of the verb that follows them.

This selectional behavior reveals an important aspect of functional head-complement relationships in generative grammar. Each auxiliary is treated as the head of a projection, such as ModP, PerfP, ProgP, or VoiceP, and it selects a complement of a particular morphological category, such as a bare verb (V), present participle (V-ing), or past participle (V-ed). Violations of these morphological restrictions result in ungrammatical sentences because the required features have not been matched or licensed properly in the syntactic derivation.

1. Indonesian Comparison: Morphological Simplicity, Semantic Precision

Indonesian, by contrast, is a non-inflectional language in which verbs do not change form to mark tense, number, or gender. However, this does not mean that verb forms are unconstrained. Rather, Indonesian relies on aspectual particles and derivational affixes to shape verb meaning and interpretational nuances. These particles and affixes function analogously to auxiliaries in English and often influence the morphological shape or acceptability of the main verb.

Consider the progressive marker *sedang* in the following examples:

- a. *Dia sedang jalan*
- b. *Dia sedang berjalan*

Both expressions are understandable and acceptable in colloquial Indonesian, but they differ in formality, grammaticality, and morphosyntactic precision. The first sentence, *dia sedang jalan*, uses the bare noun/verb stem *jalan* ("walk"), while the second, *dia sedang berjalan*, uses the verbal derivative *berjalan*, which incorporates the *ber-* prefix. The latter is more standard in formal usage and is structurally better aligned with the progressive marker *sedang*, which selects for an active intransitive verb.

This preference indicates that although Indonesian lacks inflectional morphology, its derivational morphology and aspectual system exert comparable restrictions on verb usage. In formal or written Indonesian, aspectual markers such as *sedang* (progressive), *sudah* (perfective), and *akan* (future) typically co-occur with derived verb forms, especially those prefixed with *ber-* (intransitive) or *me-* (transitive), which mark voice, transitivity, and Aktionsart (Musgrave, 2001; Sneddon et al., 2010).

For example:

- a. *Dia sudah makan.* → "He/she has eaten." (simple aspectual pairing)
- b. *Dia sudah memakan kue itu.* → "He/she has eaten the cake." (with object, transitive prefix *me-* required)

The compatibility between aspect markers and verb forms in Indonesian reveals a more flexible, but still rule-governed, system of morphological selection. Unlike English, where auxiliaries demand specific verb forms, Indonesian allows the base verb to appear in certain contexts, particularly in informal speech, but prefixed verbs are often required for clarity and grammatical well-formedness in more formal registers.

Thus, while the surface morphology of Indonesian is less overtly constrained than that of English, the underlying system still reflects selectional properties and syntactic regularities, particularly in the interaction between aspectual elements and derived verb forms.

2. Theoretical Implications

The phenomenon of auxiliary–verb selection in English and the interaction of aspectual markers with verb derivation in Indonesian both underscore the importance of morphosyntactic dependency in syntactic theory. In both languages, functional heads impose conditions on the form of their verbal complements, although the morphological realization and degree of obligatoriness vary typologically.

In minimalist syntax, this dependency can be modeled through feature checking mechanisms, where functional heads carry uninterpretable features that must be matched with interpretable features of the verb. Failure to satisfy these conditions leads to crash at the PF interface (Adger, 2003; Chomsky, 1995). While English reflects this morphologically through overt verb forms, Indonesian expresses it through derivational prefixes and semantic-pragmatic constraints, suggesting that the same underlying principles apply across typologically distinct languages.

D. Summary

This unit focused on the structural role of auxiliary verbs and their significance in syntactic theory. It began by distinguishing auxiliaries from main verbs and emphasizing their function as heads of functional projections, particularly in the Tense (T) domain. Evidence from ellipsis constructions was used to argue that auxiliaries license ellipsis and signal syntactic constituency, supporting the claim that they occupy structurally privileged positions. The unit also introduced the fixed hierarchical ordering of auxiliaries in English—Modal > Perfect > Progressive > Passive—and showed that only certain sequences are grammatical, revealing the internal structure of the verb phrase. This hierarchy was compared with Indonesian, where auxiliary-like aspect markers such as *sudah*, *sedang*, and *akan* also follow an interpretable order, though with more surface flexibility.

Another major focus of the unit was the interaction between auxiliaries and verb morphology, particularly the requirement that English auxiliaries select specific verb forms such as bare infinitives or participles. The phenomenon of do-support was briefly revisited to reinforce how English resolves morphosyntactic constraints when no auxiliary is present. In contrast, Indonesian verbs, though lacking inflectional morphology, exhibit structural sensitivity to aspect markers and derivational prefixes like *ber-* and *me-*.

The unit concluded by demonstrating how the distribution and behavior of auxiliaries serve as critical diagnostics for sentence structure, constituency, ellipsis, and the distinction between functional and lexical categories. These properties provide evidence for underlying syntactic organization and reinforce the view that even surface-level verbal expressions are shaped by deep hierarchical structure.

GLOSSARY

- **Aux Ordering** – Modal > Perfect have > Progressive be > Passive be > V.
- **Auxiliary (Aux/T)** – Do, have, be, modals; encode tense/aspect/mood.
- **Distribution** – Where auxiliaries occur relative to subjects/verbs.
- **Do-support** – Insertion of do for negation/questions in English.
- **Ellipsis Evidence** – VP-ellipsis shows Aux as a head position.
- **Indonesian Aux-like Items** – akan (future), sudah (perfect-like), sedang (progressive-like).
- **Morphological Support** – Affix hopping vs. head movement debates.
- **Negation (NegP)** – Not/negators and their position.

UNIT 8

QUESTIONS AND MOVEMENT

A. Yes-No Questions: Inversion vs. Base Generation

Yes-no questions, interrogative sentences that can be answered with "yes" or "no", exhibit distinct syntactic behaviors across languages. In English, such questions are formed through a syntactic operation known as subject-auxiliary inversion, a hallmark of interrogative clause formation in many Germanic languages. In contrast, Indonesian forms yes-no questions through base generation strategies, typically relying on interrogative particles rather than word order changes. These typological differences underscore how languages syntactically encode interrogative force and reveal the range of strategies governed by universal principles but realized through language-specific rules (Adger, 2003; Cheng, 1991; Radford, 2012).

1. English: Inversion as a Movement Operation

In English, yes-no questions are derived from declarative clauses by moving the auxiliary verb to the left of the subject. This operation involves head movement of the Tense (T) head (which hosts the auxiliary) to the Complementizer (C) position in the CP layer. This inversion results in a surface structure where the auxiliary appears before the subject, signaling the interrogative status of the sentence. Consider the following pair:

- a. Declarative: *Ewan will travel to Mars.*
- b. Interrogative: *Will Ewan travel to Mars?*

The auxiliary *will* originates in the T head and undergoes T-to-C movement, leaving the subject *Ewan* in Spec-TP. The movement of the auxiliary to C is required to satisfy the interrogative feature [+Q] on C, which in English is morphologically realized through inversion. This operation explains why only the first auxiliary undergoes movement in multi-auxiliary constructions:

- a. Declarative: *Ewan might have been traveling.*
- b. Interrogative: *Might Ewan have been traveling?*

Here, only *might* (the modal in T) moves to C, further reinforcing the idea that inversion targets the highest auxiliary. The absence of an auxiliary, as in *Ewan travels to Mars*, necessitates do-support in questions:
Does Ewan travel to Mars?

This insertion of *do* reflects the syntactic requirement that something must occupy the C position to mark interrogativity (Chomsky, 1962; Pollock, 1989)). The English system is thus strongly characterized by movement-based interrogative formation.

2. Indonesian: Base Generation with Interrogative Particles

In contrast, Indonesian does not employ subject-auxiliary inversion. Instead, yes-no questions are formed through the use of an interrogative particle, most commonly *apakah*, which is inserted at the beginning of the clause. This strategy is known as base generation, as no movement is involved; rather, the interrogative element is generated directly in the C head to mark the clause as interrogative. Consider the following Indonesian examples:

- a. Declarative: *Ewan akan pergi ke Mars.*
- b. Interrogative: *Apakah Ewan akan pergi ke Mars?*

The sentence structure remains identical to the declarative, with the only addition being the particle *apakah*. The absence of subject-auxiliary inversion demonstrates that Indonesian employs a fixed SVO word order across clause

types and relies on morphological or lexical marking to encode clause type (Musgrave, 2001; Sneddon et al., 2010).

The syntactic structure of the interrogative clause in Indonesian can be represented as:

- a. CP → C TP
- b. C = *apakah*
- c. TP = Ewan akan pergi ke Mars

In this structure, *apakah* is inserted directly into C without requiring any movement of auxiliaries or verbs. The underlying TP structure remains unchanged from the declarative. This mechanism is not only syntactically economical but also reveals that languages vary in how they satisfy the interrogative [+Q] feature, either through movement or base-generated particles.

Indonesian also allows for yes-no questions without *apakah*, relying instead on intonation alone, particularly in informal speech:

Ewan akan pergi ke Mars?

In this case, interrogativity is expressed prosodically rather than morphosyntactically. This dual strategy (with or without *apakah*) reflects the flexibility of Indonesian syntax and its typological preference for analytic structures over movement-based derivations.

3. Cross-Linguistic Implications

The contrast between English and Indonesian question formation strategies illustrates a broader typological distinction between movement-based and particle-based interrogative systems (Cheng, 1991). In movement-based systems, the interrogative clause is marked by the displacement of elements, usually to the CP domain. In particle-based systems, interrogativity is encoded by inserting lexical items like *apakah* that bear [+Q] features. Despite these surface differences, both systems conform to the broader generative model of clause structure, with CP serving as the domain in which clause type is determined.

From a theoretical standpoint, both English and Indonesian instantiate the presence of a C head bearing interrogative features, but diverge in how that head is spelled out. In English, it requires local movement of a verbal head to satisfy its feature-checking needs; in Indonesian, it is lexically realized by an interrogative particle inserted directly into C. These differences reflect parameter settings within Universal Grammar, where languages select between feature-checking through movement or through merge of a lexical item (Chomsky, 1995; Rizzi, 1997).

B. WH-Movement: Subject and Object Questions

Interrogative clauses introduced by WH-words, such as *who*, *what*, *where*, *when*, *why*, and *how*, are structurally more complex than yes-no questions. In most languages, WH-questions serve as a diagnostic tool for syntactic movement, constituency, and clause structure. In English, they typically involve WH-movement, where the WH-phrase moves to the specifier position of the Complementizer Phrase (Spec-CP) to check a [+WH] feature. In Indonesian, WH-questions are also formed via fronting but exhibit some important differences, especially in the use of the relativizer *yang* and a generally less rigid movement behavior, which makes Indonesian an interesting case for cross-linguistic comparison (Cheng, 1991; Musgrave, 2001; Radford, 2012).

1. Subject WH-Questions: Base Position and Surface Alignment

In English, WH-questions targeting the subject do not require syntactic movement of the verb or auxiliary. The WH-word originates in the subject position and remains there, although it is still interpreted as occupying Spec-CP for feature-checking purposes. The structure below illustrates this:

Who discovered the cave?

In this sentence, *who* is the subject of the verb *discovered*. Because it already appears at the front of the sentence and occupies the subject position, there is no need for auxiliary insertion or subject-auxiliary inversion. The derivation of the sentence proceeds directly from the base-generated subject WH-phrase in Spec-CP with no movement of verbal elements required.

From a syntactic perspective, the derivation of the WH-subject question is relatively simple:

CP

→ Spec-CP: *who*

→ C'

→ C: [Ø]

→ TP: *t discovered the cave*

Here, *t* marks the trace left by the subject in its base-generated position. Since there is no auxiliary, there is no T-to-C movement.

In Indonesian, subject WH-questions exhibit a similar surface structure, but they include the particle *yang*, which functions as a relative clause marker. The particle serves to introduce the predicate domain and maintain clause structure while keeping the WH-word in a high position. Consider the following example:

Siapa yang menemukan gua itu?

"Who discovered the cave?"

In this construction, *siapa* ("who") appears at the beginning of the clause, and *yang* introduces the predicate. The structure may be analyzed as:

CP

→ Spec-CP: *siapa*

→ C'

→ C: *yang*

→ TP: *t menemukan gua itu*

Like in English, the WH-word is in subject position, but unlike English, the particle *yang* is obligatory and functions to separate the subject from the predicate. While *yang* is not a complementizer in the traditional sense, it serves a similar clause-linking function in WH-questions and relative clauses (Sneddon et al., 2010).

2. Object WH-Questions: Movement and Syntactic Complexity

Object WH-questions in English involve A'-movement of the WH-phrase from the object position to Spec-CP. This movement is obligatory and is often accompanied by do-support if no auxiliary is present:

- a. Declarative: *She discovered the cave.*
- b. WH-Question: *What did she discover?*

In the WH-question, *what* originates as the direct object of *discover*, but moves to Spec-CP. To fulfill the need for a filled C head in interrogative clauses, English uses *did* (the past tense form of *do*) to support T-to-C movement when the main verb cannot raise. The movement process can be represented structurally as:

CP
 → Spec-CP: *what*
 → C'
 → C: *did*
 → TP
 → NP: *she*
 → T'
 → T: *t* → VP
 → V'
 → V: *discover*
 → NP: *t*

Here, *t* represents the trace of the moved WH-phrase *what*, and the auxiliary *did* appears in C due to the requirement for tense support.

In Indonesian, object WH-questions also involve fronting, but the strategy differs in that movement is optional in some registers, and the relativizer *yang* plays a key role in demarcating clause boundaries. Consider:

Apa yang dia temukan?

"What did she discover?"

Here, *apa* ("what") appears in clause-initial position, while *yang* introduces the predicate clause *dia temukan*. As in subject questions, *yang* functions to preserve clause structure and prevent ambiguity, especially since Indonesian does not have overt case markers or inflection to distinguish subjects from objects. This syntactic device maintains clarity in WH-extractions.

In informal Indonesian, an alternative construction might retain the base word order, relying on intonation and context:

Dia temukan apa?

This form is more typical in spoken discourse and does not involve movement. However, in formal registers, fronting with *yang* is preferred for clarity and grammatical acceptability.

Structurally, the formal WH-object question in Indonesian can be modeled similarly to English, with *apa* moving to Spec-CP and *yang* occupying the C head:

CP

→ Spec-CP: *apa*

→ C'

→ C: *yang*

→ TP: *dia temukan t*

This structure highlights the role of *yang* in introducing the relative clause-like structure and marking the boundary between the WH-phrase and the rest of the clause.

3. Theoretical and Cross-Linguistic Observations

The contrast between English and Indonesian WH-question formation reveals key parametric variations. English relies on syntactic movement to Spec-CP and often requires auxiliary insertion for interrogative marking. Indonesian, by contrast, employs a fixed clausal particle (*yang*) and may use WH-in-situ strategies in informal settings, suggesting greater syntactic flexibility.

Nevertheless, both languages exhibit WH-fronting for subject and object questions in formal registers, indicating shared principles at the level of Universal Grammar. In both cases, the WH-phrase occupies a prominent position at the left edge of the clause, and the clause is interpreted as interrogative through the presence of [+WH] features.

WH-movement is thus a fundamental operation in syntactic theory, linked to broader mechanisms such as feature-checking, intermediate landing sites (e.g., Spec-vP or CP), and island constraints in more complex sentences (Chomsky, 1995; Rizzi, 1997). Indonesian's use of *yang* reflects a typological preference for overt clause structure markers, while English satisfies the same structural need via movement and auxiliary behavior.

C. Islands and Constraints on Movement

One of the most well-studied phenomena in formal syntax is the existence of island constraints, syntactic environments that restrict or block the movement of constituents, particularly Wh-elements. These constraints are not idiosyncratic; rather, they reflect universal limitations on syntactic operations such as A'-movement, which includes Wh-movement, topicalization, and relativization. The term *island* was introduced by Ross (1967) in his foundational dissertation, and it continues to serve as a crucial diagnostic for exploring the nature of grammatical dependencies in human languages (Chomsky, 1973; Huang, 1982; Rizzi, 1990).

1. Island Constraints in English: The Complex NP Constraint

In English, various configurations have been identified as islands. These include:

- a. Complex noun phrases (e.g., *the rumor that she bought a car*)
- b. Relative clauses
- c. Coordinate structures
- d. Adjunct clauses

A classic example of an island violation involves a Complex NP Island, where a WH-phrase is extracted from within a noun phrase containing a complement clause:

What did you hear the rumor that she bought ___?

(Intended meaning: You heard a rumor that she bought *what*?)

This sentence is ungrammatical because the WH-element *what* originates inside the complement clause (*that she bought ___*), which itself is embedded inside the noun phrase (*the rumor that...*). According to the Complex NP Constraint, movement out of such structures is prohibited.

Syntactically, the problem lies in attempting to extract material from within a DP, which is considered an opaque domain for movement:

[CP What_i [did you hear [DP the rumor [CP that she bought t_i]]]] → × (ungrammatical)

The trace of *what* lies too deep inside the DP boundary, making it inaccessible to A'-movement operations that target the matrix clause's CP.

This ungrammaticality is not due to semantics or pragmatics, but to syntactic structure, a reflection of how constituent boundaries restrict dependency formation.

2. Indonesian Island Constraints: Parallel Effects

Indonesian exhibits similar island effects, particularly with complex noun phrases, suggesting that such constraints are not language-specific but rather reflect deeper properties of syntactic computation. Consider the following parallel in Indonesian:

**Apa kamu mendengar rumor bahwa dia membeli __?*
"What did you hear the rumor that she bought __?"

This sentence is degraded or ungrammatical under the intended interpretation, just like its English counterpart. The WH-element *apa* ("what") cannot be interpreted as extracted from the embedded clause (*bahwa dia membeli __*) because the clause is contained within a complex noun phrase (*rumor bahwa...*).

In standard Indonesian syntax, WH-extraction from within a noun phrase, particularly when that phrase contains a clausal complement, results in severe processing difficulty or outright ungrammaticality. The failure of this sentence is consistent with the Complex NP Island Constraint, supporting the claim that island effects are cross-linguistically robust (Musgrave, 2001; Sneddon et al., 2010).

Moreover, native speakers often resort to paraphrasing or restructuring to avoid such island violations. For example:
Acceptable: *Kamu mendengar rumor bahwa dia membeli apa?*
(You heard the rumor that she bought what?)
→ WH-in-situ strategy avoids extraction.

This alternative sentence avoids movement by leaving *apa* in situ, allowing interpretation via semantic reconstruction or discourse inference rather than overt syntactic movement. Such strategies are commonly employed in Indonesian and other WH-in-situ languages (Cheng, 1991).

3. Theoretical Perspectives on Islands

From a theoretical standpoint, island constraints are often interpreted as arising from interface conditions between syntax and other components of the grammar, such as semantics and phonology. In minimalist theory, islands are viewed not as filters imposed post-derivationally, but as the result of violations of locality conditions such as Phase Impenetrability Condition (PIC) and Relativized Minimality (Chomsky, 2001; Rizzi, 1990).

In particular, DPs are often considered phases in the sense of Chomsky (2001), and movement out of a phase without passing through its edge (i.e., Spec-DP) violates the PIC. Since embedded clauses within DPs are not accessible for feature-checking at higher levels, extraction is blocked. This analysis elegantly captures why *what* in the earlier example cannot move across the boundary of *the rumor that...*

Additionally, Rizzi's (1990) work on Relativized Minimality suggests that certain intervening elements (e.g., intervening subjects, CPs, or specifiers) can block movement if they share features with the moving element. These locality constraints offer a unified explanation for a broad array of island phenomena, reinforcing the insight that island effects reflect computational limits of syntactic derivation rather than language-specific rules.

4. Cross-Linguistic Relevance and Typological Implications

The similarity in island sensitivity between English and Indonesian challenges early claims that island effects are more severe in languages with overt WH-movement (e.g., English) than in WH-in-situ languages (e.g., Chinese or Malay). While it is true that some WH-in-situ languages tolerate broader dependencies, Indonesian demonstrates that base-generated WH-elements are still constrained by syntactic structure, especially in formal registers.

This finding aligns with research by Huang (1982) and Cheng (1991), which argues that covert movement or operator binding in WH-in-situ languages is still subject to locality constraints, even when not phonologically overt. Therefore, whether WH-phrases are fronted or not, their interpretability is shaped by the same abstract constraints governing constituent boundaries.

Indonesian's degraded acceptability in complex NP extraction contexts supports the universality of island constraints, suggesting that syntactic movement is not entirely free across languages, even when movement is less visible.

D. Summary

This unit examined the syntactic processes involved in question formation, with a focus on yes-no questions and WH-questions in English and Indonesian. It introduced subject-auxiliary inversion as a diagnostic feature of English yes-no questions and contrasted this with Indonesian, where interrogative particles like *apakah* serve a similar function without requiring movement. The unit then turned to WH-movement, explaining how WH-elements move to clause-initial position in English, leaving a trace and creating a long-distance dependency. Indonesian, in contrast, uses WH-elements followed by relative markers such as *yang*, often preserving base-generated order while relying on semantic interpretation. Island constraints and movement limitations were introduced through examples that show when extraction is blocked, as in complex noun phrase islands and adjunct islands. The unit emphasized that while surface strategies differ between languages, the abstract principles governing movement and question formation are shared and constrained by universal syntactic laws.

GLOSSARY

- **Echo vs Information Questions** — Distinct structures/intonations.
- **Island Constraints** — Domains that block movement (e.g., complex NP).
- **Pied-Piping** — Moving a larger phrase with the wh-word.
- **Subjacency** — Locality condition limiting movement distance.
- **Superiority** — Ordering effects among multiple wh-elements.
- **Wh-in-situ** — Wh stays in place (common in Indonesian).
- **Wh-Movement** — Fronting of wh-phrases to Spec-CP.
- **Yes-No Questions** — T-to-C movement/inversion in English.

UNIT 9 | TRANSFORMATIONS AND DERIVATIONS

A. Do-Support and Verb Movement

One of the notable syntactic phenomena in English is the insertion of the auxiliary “do”, known as do-support, which occurs when grammatical operations such as negation, question formation, or emphatic assertion require tense to be expressed independently of the main verb. This syntactic strategy is particularly important in contexts where verb movement is blocked or unavailable. Do-support is thus a reflex of deeper morphosyntactic constraints in English, highlighting the interaction between morphology, tense marking, and clause structure (Chomsky, 1962; Pollock, 1989; Radford, 2012).

1. English: Morphological Tense and Do-Support

In English declarative affirmative sentences, tense and agreement are typically attached directly to the main verb:

Jamal sings.

Here, *sings* reflects present tense and third person singular agreement.

However, in contexts where tense needs to surface without the verb bearing morphological inflection, such as in negation, yes-no questions, or emphatic constructions, the auxiliary *do* is inserted to host the tense and agreement features. For example:

- a. *Jamal does not sing.*
- b. *Does Jamal sing?*
- c. *Jamal does sing!* (emphatic)

In these examples, the main verb *sing* remains in its bare infinitive form, and the auxiliary *do* carries the tense and agreement information. The necessity of do-support in English arises from its syntactic configuration, in which only auxiliaries can undergo T-to-C movement (in interrogatives) or host negation (e.g., *not* must follow an auxiliary). Since main verbs in English cannot move to T or C, an auxiliary must be inserted when needed. This is not merely a stylistic choice but a syntactic requirement to satisfy formal feature-checking mechanisms (Pollock, 1989; Radford, 2012).

This contrasts with languages like French, where main verbs can raise to T, rendering do-support unnecessary. Thus, do-support in English compensates for the lack of verb movement, reflecting broader typological and parametric variation.

2. Indonesian: No Do-Support, No Verb Movement

In stark contrast to English, Indonesian does not employ do-support. Verbs in Indonesian are invariant with respect to tense and agreement, and they do not inflect morphologically for subject features. Temporal and polarity distinctions are instead expressed through independent lexical elements, such as adverbs and negators. For instance:

- a. Affirmative: *Jamal bernyanyi*.
“Jamal sings / is singing.”
- b. Negative: *Jamal tidak bernyanyi*.
“Jamal does not sing.”

Here, the negator *tidak* is placed immediately before the verb phrase without requiring any auxiliary support. The main verb *bernyanyi* remains in its base form, and no additional syntactic machinery is required to express negation. Unlike English, Indonesian has no overt tense morphology, and therefore no requirement for a verb or auxiliary to host tense features.

Syntactically, Indonesian negation follows a NegP structure, where *tidak* serves as the head of NegP and takes VP as its complement:

[TP Jamal [NegP tidak [VP bernyanyi]]]

This structure satisfies the requirements of negation without necessitating auxiliary insertion, because the verb remains base-generated within VP and there is no competition for tense-marking positions.

Furthermore, yes-no questions in Indonesian do not trigger auxiliary fronting or do-support. They are typically formed either with intonation or the insertion of interrogative particles (e.g., *apakah*), with the main verb structure remaining unaltered:

a. *Apakah Jamal bernyanyi?*

“Does Jamal sing?”

b. *Jamal bernyanyi?* (with rising intonation)

In neither case is there any morphological or syntactic evidence for do-support. This reinforces the typological distinction between inflectional languages like English and analytic languages like Indonesian, in which auxiliaries are not syntactically required for tense or polarity.

3. Cross-Linguistic Typology and Theoretical Insights

The presence of do-support in English and its absence in Indonesian reveals a broader typological divide based on syntactic head movement and verbal morphology. According to Pollock (1989), English underwent a historical loss of verb movement to T, which created the need for periphrastic constructions involving *do*. In contrast, languages that allow main verb movement, or that lack overt tense morphology, do not require auxiliary support in such contexts.

From a minimalist perspective (Chomsky, 1995), English *do* is inserted to satisfy feature-checking at PF (the interface with phonological form), not as a result of syntactic derivation per se. That is, *do* is a last-resort operation to spell out tense and agreement when the verb cannot raise to T.

Indonesian, by contrast, avoids this problem entirely. Since tense is not morphologically encoded, and verbs do not move, there is no need for auxiliary insertion. The use of

negators and aspectual particles (e.g., *sudah*, *sedang*, *akan*) reflects a system in which functional features are expressed lexically, rather than through verb inflection or auxiliary support (Musgrave, 2001; Sneddon et al., 2010).

This distinction has pedagogical significance as well: English learners of Indonesian often overgenerate auxiliaries, while Indonesian learners of English must learn that English syntax requires auxiliaries in contexts of negation, questioning, and emphasis, even when the verb remains semantically unchanged.

B. Head Movement and Affix Hopping

In generative syntax, head movement is a central operation that accounts for the derivational relationship between a verb and higher functional heads such as Tense (T) or Inflection (I). This operation plays a critical role in explaining the surface word order of verbs, auxiliaries, and affixes across languages. It is tightly linked with the morphological realization of inflectional features such as tense, aspect, agreement, and mood (Chomsky, 1962; Pollock, 1989; Radford, 2012).

In English, head movement is particularly salient in auxiliary constructions, while main verbs generally remain in their base position within the Verb Phrase (VP). This asymmetry between auxiliaries and lexical verbs underlies many morphosyntactic processes, including affix hopping, a post-syntactic operation whereby tense and agreement morphemes are phonologically attached to the main verb without actual syntactic raising.

1. English: Head Movement and Auxiliary Inversion

Consider the sentence:

She might sing.

Here, *might* is a modal auxiliary that is base-generated in the Tense head (T). There is no need for movement because modals like *might*, *can*, *will*, and *should* are inherently tensed and appear directly in T. The lexical verb *sing* remains within VP and surfaces in its base form, reflecting the fact that main

verbs in English do not raise to T. This structural configuration is supported by empirical evidence from adverb placement and negation (Pollock, 1989), which shows that only auxiliaries appear above certain adverbs (e.g., *often*, *never*), while main verbs remain below them.

Syntactic representation:

[TP She [T' might [VP sing]]]

This structure illustrates the lack of head movement for main verbs in English, with auxiliaries occupying T and governing the verb phrase.

2. Affix Hopping: Post-Syntactic Morphological Attachment

English uses a mechanism known as affix hopping to handle inflectional morphology when the main verb does not raise. This operation, which applies in the phonological component, allows bound morphemes in T (such as third person singular -s) to “hop” onto the verb located in VP.

For instance:

She walks.

The verb *walk* is base-generated in VP, and the T head carries the [+present, +3SG] feature. The -s suffix in T phonologically lowers onto the verb to yield the surface form *walks*.

This process is non-syntactic and occurs after syntactic structure has been built, thereby complying with the strict locality conditions of minimalist syntax (Chomsky, 1995). The derivation avoids actual verb movement but achieves the same surface morphology.

A derivation involving affix hopping looks like this:

[TP She [T' [T -s] [VP walk]]]

→ PF: *walk* + -s = *walks*

Affix hopping has been analyzed as a morphological repair strategy to satisfy the requirement that affixes must be phonologically supported (Halle & Marantz, 1993). It demonstrates the interface dependency between syntax and phonology.

3. Indonesian: No Affix Hopping or Verb Raising

In contrast to English, Indonesian does not exhibit either head movement of the verb to T nor any process resembling affix hopping. Verbs in Indonesian do not inflect for tense or subject agreement. Instead, verbal morphology is handled through derivational affixes, such as *ber-*, *me-*, and *di-*, which indicate voice, aspect, or valency, but are not syntactically tied to functional projections like T or I.

Consider:

Dia berjalan.\

“She walks / is walking.”

Here, *berjalan* is a derived intransitive verb formed by the prefix *ber-* and the root *jalan* (“walk”). The derivation of this verb occurs in the lexical component, not in the syntactic or morphological structure associated with Tense. There is no equivalent of tense-bearing morphology (*-s*, *-ed*) that needs to be attached via hopping or movement. The clause can be represented as:

[TP Dia [VP berjalan]]

There is no morphological dependency between T and V, and thus, no requirement for either movement or morphological attachment post-syntax. The functional category T is syntactically present, but it is realized through temporal adverbs (e.g., *kemarin*, *sekarang*, *nanti*) or aspectual particles (*sudah*, *sedang*, *akan*), not through verbal morphology.

This contrast makes Indonesian an example of an analytic language, where grammatical meaning is expressed through independent words, rather than bound morphemes or syntactic movement. The absence of affix hopping supports the typological classification of Indonesian as a language with non-concatenative, non-fusional morphology (Musgrave, 2001; Sneddon et al., 2010).

4. Theoretical Implications

The English system of head movement and affix hopping reflects a split IP structure, where tense and agreement are distributed across functional heads, as proposed in Pollock (1989) and further developed in minimalist theory. Affix hopping resolves the phonological deficiency of bound affixes stranded in T when the verb does not move. This post-syntactic repair ensures that morphological requirements are met at PF (Phonological Form), demonstrating the modularity of the grammar.

Indonesian, lacking inflectional morphology, provides a counterpoint that enriches our understanding of syntactic variation. It shows that languages can satisfy tense and aspectual interpretation through alternative strategies that do not involve head movement or morphological attachment. These differences suggest that Universal Grammar allows parameterization of movement and morphology, depending on how languages realize functional categories (Adger, 2003; Chomsky, 1995).

Understanding this variation helps students appreciate the cross-linguistic diversity of surface forms while recognizing the underlying structural similarities that govern sentence architecture across languages.

C. Derivations and Deep Structure

A foundational assumption of transformational-generative grammar, particularly in early generative models such as Government and Binding Theory (Chomsky, 1986, 1993), is that syntactic structures are derived from underlying representations, known as Deep Structure (D-Structure), through a series of rule-governed transformations, ultimately yielding a Surface Structure (S-Structure) that corresponds to the actual utterance. These derivations capture both the semantic interpretation and phonological form of a sentence. Although minimalist frameworks (Chomsky, 1995) have reduced the number of representational levels, the conceptual distinction

between abstract syntactic generation and observable surface realization remains crucial for understanding the derivational architecture of grammar.

1. Derivations in English: Tense and Auxiliary Alternations

In English, transformational derivations account for the movement and insertion operations that shape a clause's surface word order. Consider the following contrast:

a. *He did eat the sandwich.*

(Emphatic or diagnostic use; tense carried by auxiliary *did*)

b. *He ate the sandwich.*

(Canonical declarative sentence)

The first sentence represents a morphosyntactic structure closer to the underlying deep configuration, where the verb *eat* appears in its base form, and tense is carried by *do*, a strategy used when movement or affixation is blocked. The second sentence reflects the surface realization after verb–tense fusion, where the main verb *eat* carries the past tense inflection *-ed* (irregularly realized as *ate*). The derivation involves tense feature valuation, and the morphological realization is determined at PF (Phonological Form) through processes such as affix hopping (Pollock, 1989; Radford, 2012).

In minimalist terms, the derivation proceeds via Merge, Move, and Agree operations, where features from T (tense) are valued either through head movement (e.g., with auxiliaries) or morphological processes (e.g., regular or irregular past tense marking). The distinction between the emphatic *did eat* and the canonical *ate* highlights the interplay between syntax, morphology, and semantics in clause derivation.

Furthermore, surface variation such as VP ellipsis (e.g., *He did eat the sandwich, and she did too*) or negative polarity licensing (e.g., *He didn't eat the sandwich*) confirms the syntactic reality of deeper structures that may not always be phonologically realized.

2. Derivations in Indonesian: Aspectual Simplification and Informality

In Indonesian, derivational transformations are not governed by inflectional tense morphology, but by the insertion or omission of aspectual markers, which reflect the temporal interpretation of the clause. Consider the pair:

a. *Dia sudah makan sandwich itu.*

“He has already eaten the sandwich.” (Aspect marked)

b. *Dia makan sandwich itu.*

“He ate the sandwich.” (Unmarked; neutral or generic past)

The first sentence includes the aspectual marker *sudah*, which indicates completion or anteriority (Musgrave, 2001; Sneddon et al., 2010). This can be interpreted as the deep or fully specified structure, containing a functional projection (e.g., AspP or PerfP) that encodes temporal information explicitly. The second sentence, by contrast, omits the aspect marker, yielding a more surface-like realization that is common in colloquial or context-dependent speech.

The derivational reduction from *sudah makan* to simply *makan* reflects a form of surface simplification that is pragmatically licensed. In discourse, the temporal interpretation of *makan* (e.g., whether it refers to past or habitual action) is often recovered via contextual inference, eliminating the need for explicit aspect marking. The omission of *sudah* is not ungrammatical but represents a syntactic economy that reflects Indonesian's analytic nature and reliance on discourse pragmatics for interpretation.

From a structural perspective, the two forms may share a similar core VP but differ in the presence of an Aspect Phrase (AspP) in the functional layer:

a. Deep: [TP Dia [AspP sudah [VP makan sandwich itu]]]

b. Surface: [TP Dia [VP makan sandwich itu]]

In this way, the derivational difference corresponds not to morphological inflection but to the optional projection of aspectual heads, supporting the view that derivational

variation can be captured through variable syntactic structure rather than obligatory transformations.

3. Cross-Linguistic Comparison and Theoretical Significance

The derivational contrasts between English and Indonesian illuminate broader typological patterns in the mapping from syntactic structure to surface realization. In English, transformations are tightly constrained by morphosyntactic features, such as tense and agreement, necessitating operations like verb raising, auxiliary insertion, or affix hopping. In Indonesian, by contrast, derivational differences are reflected more in the lexical choice of aspectual particles and constituent omission, with relatively limited morphosyntactic dependencies.

Both languages, however, exhibit the principle of economy, wherein optional material (e.g., *did*, *sudah*) is often omitted in informal registers unless it is required for emphasis, contrast, or disambiguation. This reinforces the idea that derivational structure underlies surface variation, even in languages without inflectional morphology.

Moreover, the ability to reconstruct deep structure from surface forms, whether through auxiliary insertion in English or aspectual omission in Indonesian, supports the generative hypothesis that human syntax operates through abstract, hierarchical representations, which may be partially or fully realized in actual utterances (Adger, 2003; Chomsky, 1993).

D. Summary

This unit introduced the concept of syntactic transformations as operations that map deep structures onto surface structures. Key transformational processes such as do-support, verb movement, and affix hopping were discussed in detail. English requires do-support when tense or negation cannot be hosted by a lexical verb, highlighting the difference between auxiliary and main verb behavior. Head movement was examined as a way to explain how affixes like *-s* in *walks* appear

in surface position despite being generated elsewhere. The unit also discussed derivations as a step-by-step sequence of syntactic operations governed by principles such as feature checking, locality, and economy. Raising, passive formation, and question movement were analyzed as derivational phenomena. Indonesian examples were used to show how similar interpretations are achieved without the same morphological processes, relying instead on word order, aspect markers, and discourse strategies. Overall, the unit emphasized that although languages differ in how transformations are realized, their syntactic derivations are guided by common principles.

GLOSSARY

- **Affix Hopping** – Historical analysis of tense affixes lowering.
- **Chain/Trace (Copy)** – Record of movement in earlier/modern terms.
- **Deep vs. Surface Structure (GB)** – Earlier two-level view of syntax.
- **Derivation** – Step-by-step building of a sentence.
- **Do-Support (Again)** – Transformation to host tense/negation.
- **Feature Checking/Agree** – Formal licensing in Minimalism.
- **Head Movement** – Verb-to-T/C in some languages.
- **Movement (Move/Attract)** – Displacement to satisfy features.



PART IV:
DEEP STRUCTURE, CASE,
AND BINDING



UNIT 10

THETA THEORY AND CASE

A. Theta Roles and Argument Structure

In the study of syntax and semantics, verbs are not arbitrary lexical items; they are semantically rich predicates that impose specific requirements on the number and types of arguments they co-occur with. These requirements are formalized in what is known as Theta Theory, a subcomponent of X-bar and Government and Binding theory (Chomsky, 1993; Haegeman, 1994). Theta theory seeks to account for how semantic roles, called theta roles or thematic roles, are assigned to syntactic constituents within a sentence. Each argument associated with a verb receives exactly one theta role, and each theta role is assigned to exactly one argument, this is the Theta-Criterion (Chomsky, 1993).

1. Common Theta Roles and Their Semantic Content

While there is some debate regarding the precise inventory of theta roles, several roles recur across languages and have become widely accepted in syntactic literature (Carnie, 2012; Jackendoff, 1990; van Valin, 2012). Some of the most common theta roles include:

- a. Agent: The initiator or intentional performer of an action.
- b. Theme: The entity that undergoes or is affected by the action.
- c. Experiencer: An entity that perceives, feels, or cognitively experiences something.
- d. Goal: The endpoint or destination of a movement or transfer.

- e. Source: The origin of a movement or event.
- f. Instrument: The means by which an action is performed.
- g. Location: The place where the event occurs.
- h. Beneficiary: The entity that benefits from the action.

The assignment of these roles is governed by the lexical semantics of the verb. For example, verbs like *hit*, *see*, *know*, and *give* come with specific theta grids that must be satisfied syntactically for the sentence to be grammatical.

2. English Examples: Thematic Mapping in Practice

Consider the following declarative clause:

The hawk spotted the drone.

In this sentence, the verb *spotted* assigns two theta roles:

- a. Agent: *The hawk* – the initiator of the act of seeing.
- b. Theme: *The drone* – the entity that is seen.

The theta roles assigned by the verb are realized syntactically as arguments: *the hawk* is the subject (in Spec-TP), while *the drone* is the object (within the VP, as the complement of V). This one-to-one mapping respects the Theta-Criterion.

A more cognitively oriented example is:

The child imagined a world made of jelly.

Here, the verb *imagined* involves:

- a. Experiencer: *The child* – the one doing the imagining.
- b. Theme: *A world made of jelly* – the imagined content.

This construction illustrates that even non-action predicates (i.e., verbs of mental activity) enforce strict argument structure based on their theta grids. The subject is not an agent in the traditional sense, but rather an experiencer of an internal psychological state.

3. Indonesian Parallels: Argument Realization and Lexical Matching

Indonesian, though structurally different from English in terms of morphology and word order flexibility, reflects the same basic principles of theta role assignment. Consider the corresponding translation:

Anak itu membayangkan dunia dari agar-agar.

“The child imagined a world made of jelly.”

Here, the verb *membayangkan* (“to imagine”) assigns:

- a. Experiencer: *Anak itu*—the individual who does the imagining.
- b. Theme: *Dunia dari agar-agar*—the imagined entity.

Despite differences in surface form, such as the presence of the verbal prefix *mem-* and the use of syntactically uninflected noun phrases, the underlying theta grid is functionally identical to the English counterpart. This suggests that the semantic structure of predication is universal, even when the surface syntax varies (Musgrave, 2001; Sneddon et al., 2010).

Moreover, Indonesian verbs also exhibit predictable argument structures. For example:

Memberi (“to give”) assigns:

- a. Agent: the giver
- b. Theme: the thing given
- c. Goal: the recipient

These roles can surface in various syntactic configurations, including those involving clitic doubling or prepositional phrases (e.g., *kepada dia* – “to him/her”).

4. Thematic Hierarchies and Syntactic Realization

The mapping of theta roles onto syntactic positions is influenced by thematic hierarchies, ordering principles that reflect prominence or syntactic accessibility. A commonly cited hierarchy is:

Agent > Experiencer > Goal/Recipient > Theme > Location

This hierarchy often correlates with surface word order: more prominent roles appear in higher syntactic positions (subject), while less prominent roles occur as objects or adjuncts. For instance, in both English and Indonesian, agents and experiencers typically surface as subjects, while themes and goals appear in complement or oblique positions.

5. Cross-Linguistic Insights and Theoretical Implications

The uniform assignment of theta roles across English and Indonesian supports the hypothesis that argument structure is a universal component of the human language faculty. While surface syntax is subject to variation, the underlying principles that govern semantic-syntactic mapping appear to be stable across languages. This is in line with the theory of Universal Grammar (Baker, 2003; Chomsky, 1993), which posits that all languages share a core set of structural properties, including how predicates relate to their arguments.

Understanding theta roles is essential for analyzing syntactic phenomena such as passivization, control constructions, raising, and ditransitives, all of which rely on knowing which arguments a verb licenses and how they are realized in the syntax.

B. Thematic Hierarchies

In the generative tradition, particularly within the framework of Government and Binding Theory and subsequent minimalist syntactic models, thematic hierarchies serve as an organizing principle for understanding the interface between argument structure and surface syntactic realization. While theta roles are assigned based on the lexical semantics of predicates, their positioning in the clause, especially in languages with flexible or non-configurational word order, is influenced by hierarchical rankings among those roles. This ranking, often referred to as a thematic hierarchy, reflects both cognitive salience and syntactic prominence (Grimshaw, 1992; Jackendoff, 1990; van Valin, 2012).

1. A Canonical Thematic Hierarchy

A widely accepted thematic hierarchy, based on syntactic and semantic observations across typologically diverse languages, is given as follows:

Agent > Experiencer > Goal/Recipient > Theme > Location

This ordering is not merely theoretical; it has real consequences for syntactic configuration, including subject selection, argument realization, passivization, and the ordering of arguments and adjuncts within the clause. The position of an argument in the syntactic tree often corresponds to its position in the thematic hierarchy, higher-ranked roles are typically mapped to higher syntactic positions.

For example, agents, which are initiators of action and typically sentient and volitional, are prototypically realized as subjects across languages. Lower roles like themes or locations tend to appear in object or adjunct positions, respectively. Thematic hierarchies thus reflect a gradient of syntactic accessibility, influencing processes such as topicalization, passivization, and argument omission.

2. English Examples: Agent Prominence and Linear Order

English syntax strongly reflects thematic hierarchy in its canonical word order and subject selection. Consider the sentence:

The developer redesigned the plaza for tourists.

In this construction:

- a. Agent: *The developer*—mapped to the subject position (Spec-TP)
- b. Theme: *The plaza*—mapped to the object position (complement of V)
- c. Goal/Beneficiary: *For tourists*—expressed as a PP adjunct

This linear order mirrors the thematic hierarchy: the most prominent role (Agent) appears first, followed by the Theme and then the Goal/Beneficiary. Even in more marked

constructions such as passives or clefts, thematic hierarchy often persists in interpretation, if not in surface order.

This pattern aligns with broader cross-linguistic generalizations such as Subject = Highest Theta Role (Baker, 1988; Grimshaw, 1992), reinforcing the idea that syntax is sensitive to semantic prominence.

3. Indonesian Parallels: Role Hierarchy Without Morphosyntactic Agreement

Although Indonesian lacks overt tense and agreement morphology, the organization of arguments in the clause still reflects the same underlying thematic hierarchy. Consider the equivalent sentence:

Pengembang itu mendesain ulang plaza untuk turis.

“The developer redesigned the plaza for tourists.”

Here, the thematic roles align as follows:

- a. Agent: *Pengembang itu* (the developer) as the subject
- b. Theme: *Plaza* as the object
- c. Goal/Beneficiary: *Untuk turis* (for tourists) as an adjunct PP

Despite structural differences, such as the use of derivational morphology (*mendesain ulang*) rather than inflection, the hierarchical relationship among arguments is preserved. The Agent occupies the subject position, the Theme appears as the verbal complement, and the Goal is syntactically realized as an adjunct introduced by a preposition (*untuk*).

Indonesian’s relatively free word order in informal or poetic contexts (e.g., *Untuk turis, plaza itu didesain ulang oleh pengembang*) does not disrupt the interpretive force of the thematic hierarchy. The preferred default order, especially in written or formal registers, aligns with the prominence-based ordering of semantic roles. This observation confirms that even in analytic, morphologically impoverished languages, hierarchies of thematic prominence

influence grammatical structure (Musgrave, 2001; Sneddon et al., 2010).

4. Thematic Hierarchy and Syntactic Processes

Thematic hierarchies are also relevant to syntactic operations such as:

- a. Passivization: Typically demotes the Agent and promotes the Theme to subject position.

- 1) Active: *The developer redesigned the plaza.*

- 2) Passive: *The plaza was redesigned (by the developer).*

The passive subject must be the highest non-agentive theta role, usually the Theme.

- b. Causatives: Introduce an external causer above an existing argument structure.

- 1) English: *The manager made the team cancel the event.*

- 2) Indonesian: *Manajer itu membuat tim membatalkan acara itu.*

In both cases, the causer (new Agent) outranks the original Theme and is syntactically promoted to subject.

- c. Control Structures: Subjects of infinitival clauses are interpreted as Agents or Experiencers, depending on their theta roles assigned by the matrix verb.

- 1) English: *She promised to leave.* (*She* = Agent of both verbs).

- 2) Indonesian: *Dia berjanji untuk pergi.* (same thematic mapping).

These syntactic behaviors confirm that thematic prominence influences not only argument order, but also argument selection and movement, reinforcing the interaction between lexical semantics and syntactic structure (Grimshaw, 1992; Levin & Hovav, 1995).

5. Cross-Linguistic and Theoretical Implications

The presence of similar hierarchical preferences in languages as typologically distinct as English and Indonesian suggests that thematic hierarchies are not language-specific artifacts but likely reflect cognitive universals in human language. They inform learners' expectations about argument roles, guide sentence processing, and help disambiguate surface forms.

Moreover, within the minimalist framework (Chomsky, 1995), thematic roles are introduced in the lexical domain (vP), while syntactic operations manipulate constituents above this level. Thematic hierarchy thus bridges lexical conceptual structure and syntactic derivation, making it a vital component of contemporary theories of argument structure and predicate composition.

C. Case Assignment and the Extended Projection Principle (EPP)

A central tenet of syntactic theory in the Government and Binding framework and its minimalist successors is that noun phrases (NPs) must be licensed within a clause. One of the primary licensing mechanisms is Case assignment. Case Theory posits that all overt noun phrases must receive abstract Case at the level of syntax (Chomsky, 1986, 1993; Radford, 2012). This abstract Case, distinct from morphological case, is required to satisfy formal feature-checking within a derivation. Failure to receive Case results in ungrammaticality, making Case assignment a crucial constraint on well-formedness.

1. Structural Case: Nominative and Accusative

In English, structural Case is assigned as follows:

- a. Nominative Case is assigned by finite Tense (T) to the subject in Specifier of TP (Spec-TP).
- b. Accusative Case is assigned by transitive verbs to their internal arguments (typically direct objects) within the VP.

Consider:

- a. *Ziya has vanished.*

Here, *Ziya* is assigned nominative Case by the finite auxiliary *has* located in T.

- b. *The scientist launched the probe.*

1) *The scientist*: subject, receives nominative Case from T.

2) *The probe*: object, receives accusative Case from the verb *launched*.

The licensing of noun phrases through Case assignment ensures that each DP (Determiner Phrase) is both semantically interpretable and syntactically well-formed. It also explains why certain noun phrases cannot appear in positions where no Case is available (e.g., *Him likes ice cream* is ungrammatical because *him* cannot receive nominative Case).

2. The Extended Projection Principle (EPP)

In addition to Case assignment, the Extended Projection Principle (EPP) mandates that every finite clause must have a subject in Spec-TP, even if semantically vacuous (Chomsky, 1981). The EPP accounts for the use of expletive subjects, syntactically required but semantically empty elements, in sentences that lack an overt thematic subject.

For example:

- a. *It seems that Ziya has vanished*

The verb *seems* does not assign an external theta role, yet English grammar demands that Spec-TP be filled. The expletive "it" is inserted to fulfill the EPP requirement, even though it contributes no meaning to the sentence. The clausal complement *that Ziya has vanished* contains the real subject but cannot satisfy EPP directly.

- b. *There appeared to be a comet over the forest*

Similarly, *there* serves as a quasi-argumental expletive, allowing the real subject (*a comet*) to remain in a lower position for Case and thematic assignment. The DP *a comet* later raises or checks Case in the derivation

depending on the syntactic framework adopted (e.g., through Agree or movement).

These constructions reveal how Case assignment and the EPP operate independently but interactively: Case ensures nominal licensing, while the EPP ensures structural completeness of the clause.

3. Indonesian: Case Without Morphology and Flexible EPP Satisfaction

Indonesian, unlike English, lacks overt morphological case and is often classified as a non-configurational or analytic language. However, Case assignment is still active at an abstract level and is observable through syntactic distribution and argument licensing (Musgrave, 2001; Sneddon et al., 2010).

Compare:

Sepertinya Ziya sudah menghilang.

Literally: "Seems-like Ziya already disappeared."

This sentence lacks an overt subject like English *it*, but is fully grammatical. The preverbal expression *sepertinya* ("seemingly") performs a discourse-marking function and allows the verb *menghilang* ("to disappear") to surface without requiring an expletive subject. The DP *Ziya* satisfies the thematic and structural subject role without movement to a canonical subject position.

This construction shows that Indonesian does not enforce the EPP in the same rigid manner as English. Rather than requiring a filled Spec-TP with an expletive, Indonesian permits null or adverbial-like expressions to satisfy discourse and structural requirements. Moreover, since Case is not morphologically marked, argument licensing relies more heavily on word order, prepositions, and verbal morphology for interpretation.

Another key comparison involves existential constructions:

- a. English: *There is a dog under the table.*
- b. Indonesian: *Ada anjing di bawah meja.*
(Literally: "Exists a dog under table")

Here, Indonesian employs the existential predicate *ada* ("there is") without requiring an expletive element like *there*. The subject *anjing* ("dog") receives its interpretation and abstract Case within the VP domain, with no need to raise to Spec-TP for EPP satisfaction. This illustrates that Indonesian satisfies structural requirements through predicate-internal licensing, contrasting with English's T-domain licensing.

4. Theoretical Implications

From a minimalist perspective (Chomsky, 1995, 2001), Case is a result of feature valuation within the Agree system: T probes for a DP with unvalued Case features, and once an appropriate DP is found, Case is assigned. Similarly, the EPP is theorized as either a formal feature of T (requiring a specifier) or the reflex of a phasal boundary condition, ensuring that syntactic structure is complete before transfer to the interfaces.

Languages vary in how these principles are realized:

- a. English enforces the EPP strictly and uses expletives to satisfy it.
- b. Indonesian allows discourse elements, null subjects, or predicate-internal arguments to fulfill the same structural function without overt syntactic devices.

These differences reinforce the notion that universal grammatical principles manifest in language-specific ways, allowing for parametric variation in how Case and EPP requirements are satisfied (Adger, 2003).

D. Summary

This unit explored how syntactic structures reflect the semantic roles that verbs assign to their arguments, known as theta roles. It introduced common theta roles such as Agent,

Theme, Experiencer, and Goal, and explained how verbs encode specific expectations about the number and type of arguments they select. Theta theory was linked to syntactic structure through argument realization and the requirement that all arguments be assigned a theta role. Thematic hierarchies were discussed as influencing surface word order, though constrained by syntactic principles. The unit then introduced Case Theory, explaining that all overt noun phrases must receive abstract Case to be licensed. It covered nominative and accusative Case assignment in English, as well as the Extended Projection Principle (EPP), which requires that every clause have a subject. Indonesian data were used to show how Case and theta roles are interpreted even in the absence of overt inflectional morphology. The unit concluded that theta roles and abstract Case form essential components of sentence well-formedness, guiding both syntactic structure and semantic interpretation.

GLOSSARY

- **Argument Structure** – Mapping between roles and syntactic positions.
- **Case (Abstract Case)** – NPs must get nominative/accusative, etc.
- **EPP** – Clause must have a subject in Spec-TP.
- **Experiencer/Stimulus** – Roles in psychological predicates.
- **Indonesian Voice Morphology** – meN-/di- patterns interact with roles.
- **Theta Criterion** – One θ -role per argument; each θ -role assigned once.
- **Theta Role (θ -role)** – Semantic role assigned by a predicate (Agent, Theme, Goal).
- **Unaccusative vs. Unergative** – Intransitives with Theme vs. Agent subjects.

UNIT 11 | BINDING THEORY

A. Anaphors, Pronouns, and R-expressions

One of the major components of generative syntax is Binding Theory, a system proposed within the Government and Binding framework (Chomsky, 1986, 1993) to account for the distribution and interpretation of nominal expressions, particularly anaphors, pronouns, and referential expressions (or R-expressions). Binding Theory is crucial in explaining coreference relations, that is, when two expressions refer to the same individual or entity, and determining which configurations are syntactically acceptable.

Binding Theory identifies three categories of noun-like elements or nominal expressions, each governed by distinct syntactic principles that determine how and where they may or may not be interpreted in a sentence.

1. The Three Categories of Nominal Expressions

- a. Anaphors: These are expressions such as *himself*, *herself*, *themselves*, or their language-specific equivalents. Anaphors must be bound within a specific local domain. To be bound means that the anaphor must have an antecedent, another expression that refers to the same entity, within the same clause or binding domain.
- b. Pronouns: These include items like *he*, *she*, *it*, *them*. Pronouns must be free in their binding domain. That is, they cannot be coindexed with another expression in the same local syntactic domain.

- c. R-expressions (referential expressions): These are full noun phrases such as *Ravi*, *the teacher*, *the woman in the red dress*. R-expressions must be free everywhere; they cannot be coindexed with any other expression that c-commands them.

Each of these categories is governed by a distinct principle:

- a. Principle A: An anaphor must be bound within its local domain.
- b. Principle B: A pronoun must be free within its local domain.
- c. Principle C: An R-expression must be free everywhere.

2. Illustrative Examples in English

These principles can be clearly observed through carefully constructed examples.

a. Anaphor Binding (Principle A):

Nora painted herself with glowing ink.

In this sentence, the anaphor *herself* must be bound by *Nora*, which is within the same clause and c-commands *herself*. The binding is successful, and the sentence is grammatical. The coreference is permitted because the antecedent *Nora* satisfies both the locality and c-command conditions required by Principle A.

b. Pronoun Independence (Principle B):

Juno believed her friend doubted her.

In this case, the pronoun *her* may or may not refer to *Juno*. If it does refer to *Juno*, it must not be bound within the same clause; the pronoun and its potential antecedent are in separate binding domains. The sentence is therefore ambiguous but grammatical. If *her friend* and *her* refer to different people, there is no binding relation. If they refer to the same person, binding does not violate Principle B due to the clause boundary created by *believed*.

However, a violation would occur if a pronoun is bound within the same clause where the antecedent c-commands it:

**Juno saw her in the mirror.*

(if *her* is intended to mean *Juno*—ungrammatical under this reading)

c. R-expression Independence (Principle C):

The astronaut said that the astronaut would return.

In this case, the repetition of *the astronaut* forces independent reference. Principle C prohibits coindexation between a c-commanding antecedent and an R-expression. The two instances of *the astronaut* must refer to different individuals unless the second does not fall under the c-command domain of the first. If we attempt to coindex both, the result is pragmatically odd and syntactically constrained.

Consider:

He said that Ravi would return.

This is grammatical only if *he* ≠ *Ravi*, due to Principle C.

3. Binding Domains and C-command

To apply these principles effectively, one must understand the notion of binding domain and c-command.

- a. The binding domain is usually defined as the smallest clause containing the nominal expression and a subject.
- b. C-command refers to a structural relationship where a node A c-commands node B if the first branching node dominating A also dominates B, and A does not dominate B itself.

Anaphors must be bound by an antecedent that c-commands them within their local domain. Pronouns must avoid being c-commanded by a coindexed expression within the same domain. R-expressions must avoid being c-commanded by any coindexed antecedent at all.

4. Indonesian Parallels: Anaphors and Pronouns

In Indonesian, the binding principles show interesting parallels with English, though expressed through different morphological forms.

a. Anaphor:

Nora melukis dirinya dengan tinta bercahaya.

"Nora painted herself with glowing ink."

The reflexive *dirinya* corresponds to English *herself* and is bound by *Nora* within the same clause, satisfying Principle A.

b. Pronoun:

Juno percaya bahwa temannya meragukan dia.

"Juno believes that her friend doubted her."

Here, *dia* (she/he) functions like an English pronoun. It can be interpreted as either coreferential with *Juno* or someone else. The clause boundary (marked by *bahwa*) prevents a Principle B violation.

c. R-expression:

Astronot itu mengatakan bahwa astronot itu akan kembali.

"The astronaut said that the astronaut would return."

Just like in English, the repetition of the same noun phrase implies disjoint reference. If the speaker intended to refer to the same individual, the sentence would be semantically awkward, violating Principle C if the antecedent c-commands the second occurrence.

5. Theoretical and Cross-Linguistic Significance

Binding Theory not only explains coreference within and across clauses but also provides strong evidence for the hierarchical and abstract nature of syntactic structure. The distribution of pronouns and reflexives is not determined by linear proximity or semantic plausibility but by formal structural relationships such as c-command, locality, and clausal boundaries. These constraints are universal but may

manifest differently across languages depending on available forms, clause structure, and morphological marking.

While English uses distinct forms for reflexives and pronouns, Indonesian relies heavily on syntactic position and context, making it a valuable comparative language for testing the universality and parameters of Binding Theory (Landau, 2000; Safir, 2004).

B. Principles A, B, and C

Within Binding Theory, the syntactic distribution and interpretation of noun phrases, especially anaphors, pronouns, and R-expressions, are governed by three central principles originally formulated by Chomsky (1993) in the Government and Binding (GB) framework. These principles are formulated in terms of syntactic relations such as binding, c-command, and local domain.

Each principle constrains how a noun phrase (NP or DP) may be interpreted relative to another within a sentence. When these principles are violated, the result is ungrammaticality or semantic infelicity. Understanding these rules allows syntacticians to explain how reference is structured in natural language and why some coreference patterns are acceptable while others are strictly prohibited.

1. The Three Binding Principles

- a. Principle A:** An anaphor must be bound in its local domain.

This means that an anaphor (e.g., *himself*, *herself*, *themselves*) must be c-commanded by a coindexed antecedent within the smallest clause or syntactic unit that contains a subject. If the antecedent is outside the anaphor's local domain, or does not c-command it, the structure becomes ungrammatical.

- b. Principle B:** A pronoun must be free in its local domain.

In contrast to anaphors, pronouns (e.g., *he*, *she*, *him*, *them*) cannot be bound (i.e., c-commanded by a coindexed NP) within the same local domain. They must instead

refer to entities introduced earlier in discourse, or outside the immediate clause.

- c. **Principle C:** An R-expression must be free everywhere.

R-expressions (referential noun phrases such as *Ali*, *the teacher*, *my brother*) must be interpreted independently. They cannot be bound by any coindexed expression that c-commands them, whether inside the same clause or not. This reflects their referential autonomy in syntax and semantics.

2. Grammatical and Ungrammatical Illustrations

Let us now examine examples that violate each principle and analyze the syntactic reasons for their unacceptability.

- a. **Violation of Principle A (Anaphor not properly bound)**

**Rico admired himself's story.*

In this sentence, the reflexive *himself* is used as a possessor, but reflexives in English generally cannot function as possessors of noun phrases. The possessive construction creates a domain boundary that blocks binding. Since *himself* is not bound in its local domain (within the DP *himself's story*), the sentence violates Principle A. A grammatical alternative would be:

Rico admired his own story.

Here, *his own* is a possessive form that obeys binding conditions because *his* is a pronoun and *own* marks the reflexive meaning.

- b. **Violation of Principle B (Pronoun improperly bound)**

**Leo told him that Leo would win.*

In this structure, the pronoun *him* is in the same clause as the subject *Leo* and is c-commanded by it. If *him* is coindexed with *Leo* (i.e., referring to the same person), then Principle B is violated because the pronoun is not free within its local domain. Binding is forbidden here.

To make the sentence grammatical with a coreferential reading, the structure must be adjusted:
Leo told himself that he would win.

Now, *himself* is appropriately bound within its domain, satisfying Principle A.

Alternatively, if *him* is meant to refer to someone other than *Leo*, the original sentence is grammatical with disjoint reference.

c. Violation of Principle C (R-expression bound by pronoun)

**He thinks that Ali will win.*
 (with *he* = *Ali*)

This sentence is only problematic if *he* and *Ali* are coindexed. In this case, the pronoun *he* c-commands *Ali*, and if they refer to the same person, the sentence violates Principle C. R-expressions like *Ali* must be free, they cannot be syntactically dependent on a coindexed antecedent.

A structurally acceptable version would be:
Ali thinks that he will win.
 (with *he* optionally coindexed or disjoint)

Here, *Ali* c-commands *he*, and thus there is no violation of Principle C. Coreference is permitted depending on pragmatic context.

3. Summary Table of the Principles

Category	Principle	Requirement	Example (✓ = grammatical, ✗ = ungrammatical)
Anaphors	Principle A	Must be bound in local domain	✓ <i>Maria blamed herself.</i> ✗ <i>Maria blamed himself.</i>
Pronouns	Principle B	Must be free in local domain	✓ <i>Leo told him that he would win.</i> ✗ <i>Leo told him that Leo would win.</i>

Category	Principle	Requirement	Example (✓ = grammatical, ✗ = ungrammatical)
R-expressions	Principle C	Must be free everywhere	✓ <i>Ali thinks that he will win.</i> ✗ <i>He thinks that Ali will win.</i>

4. Additional Theoretical Remarks

The significance of Binding Theory extends beyond descriptive grammar. These principles are taken as evidence for abstract syntactic structure and the universality of hierarchical relationships in human language. Binding constraints are not reducible to linear order or surface meaning. Instead, they reflect deep configurational principles such as c-command, structural domains, and syntactic dependency.

Moreover, cross-linguistic studies (e.g., Landau, 2000; Safir, 2004) have shown that these principles manifest with variation across languages. For instance, some languages allow possessive reflexives (*himself*'s equivalents), while others (like Indonesian) rely more heavily on discourse factors to establish binding relationships.

Nonetheless, the core logic remains consistent: syntax imposes structural limits on coreference, ensuring that sentences not only express meaning but also adhere to constraints grounded in cognitive grammar.

C. Locality and Binding Domains

A critical component of Binding Theory is the concept of locality, the idea that syntactic binding is constrained by the clausal or structural domain in which a nominal expression appears. Binding is not merely a matter of semantic interpretation or discourse context; rather, it is governed by hierarchical syntactic structure, particularly c-command and the binding domain. Understanding locality is key to distinguishing grammatical from ungrammatical configurations involving anaphors and pronouns.

1. Definition of Binding Domain

In most generative frameworks, especially those following the Government and Binding model (Chomsky, 1986, 1993), the binding domain is defined as the smallest TP (Tense Phrase) or clause that contains:

- a. The anaphor or pronoun in question, and
- b. A subject that could potentially serve as a binder.

This domain restricts where an antecedent can legally occur to bind a reflexive or a pronoun. If a potential antecedent lies outside this minimal domain, it cannot bind the anaphor without violating Principle A (for reflexives) or Principle B (for pronouns).

2. English Case: Reflexive Binding Blocked by Locality

Consider the following English sentence:

The cat near him feared himself.

This sentence is ungrammatical under the reading where *him* and *himself* are coindexed (i.e., referring to the same individual). To understand why, we must examine the syntactic structure.

In this case:

- a. *Him* is embedded within a prepositional phrase (PP) (*near him*), which is in turn part of the subject NP (*The cat near him*).
- b. *Himself* appears as the *object of the verb *feared*.
- c. Crucially, *him* does *not c-command *himself*, because *him* is embedded within a PP modifier of the subject, not in a syntactic position that allows it to bind anything outside the PP.

C-command is a necessary precondition for syntactic binding. Since *him* fails to c-command *himself*, binding fails, and the sentence violates Principle A, which states that an anaphor must be bound in its local domain.

A grammatical alternative would be:

The cat feared himself.

In this case, the subject *the cat* c-commands *himself* within the same TP, and binding is licensed.

3. Indonesian Comparison: *Dirinya* and Domain Sensitivity

Indonesian reflexives, especially the form *dirinya*, display similar sensitivity to locality, although certain constructions may be more flexible due to the language's analytic nature and interaction with discourse prominence. Consider the example:

Kucing di dekatnya takut pada dirinya.

Literally: "The cat near him was afraid of himself."

Syntactically, this sentence mimics the problematic English structure. *Di dekatnya* contains the pronoun *-nya* ("him"), which, like its English counterpart *him*, is embedded within a PP inside the subject NP. The reflexive *dirinya* appears in the complement position of the verb *takut* ("was afraid"), typically requiring a c-commanding antecedent within the same TP to satisfy Principle A.

Under a strict syntactic interpretation, *-nya* does not c-command *dirinya*, and thus the reflexive would be expected to be **unbound**, violating Principle A. However, the acceptability of this construction in Indonesian may vary, and some speakers may allow it depending on discourse salience, context, and pragmatic recoverability.

This variation suggests that while Indonesian generally adheres to Binding Theory constraints, it may tolerate certain violations in favor of interpretive flexibility, especially when *dirinya* is used reflexively but influenced by topicality or contrastive focus. This aligns with research that highlights the influence of discourse context in binding constructions in non-configurational or analytic languages (Cole et al., 2006; Musgrave, 2001).

4. Implications for Locality in Syntax

The behavior of reflexives in both English and Indonesian supports the idea that syntactic binding is not reducible to proximity or linear order, but rather determined by structural relationships such as c-command and clause-internal binding domains. Violations of locality constraints lead to ungrammatical or degraded sentences.

Moreover, cross-linguistic variation in the treatment of binding domains illustrates how Universal Grammar provides the foundational principles (e.g., c-command, locality, subject orientation), while language-specific rules and discourse pragmatics can modify or extend these principles' application.

In syntactic theory, especially within the minimalist program (Chomsky, 1995; Safir, 2004), locality constraints are often reanalyzed as the result of feature checking, phase boundaries, or movement constraints, but the essential insight remains: binding is locally constrained and structurally dependent.

D. Summary

This unit focused on Binding Theory, a foundational component of generative syntax that explains how noun phrases, pronouns, and anaphors relate to each other structurally within a sentence. It introduced the three key principles of Binding Theory, Principle A, Principle B, and Principle C, each of which governs the distribution and interpretation of different types of referential expressions.

Principle A requires that anaphors, such as *himself* or *herself*, be bound within their local domain. Principle B dictates that pronouns like *him* or *her* must be free within their local domain, and Principle C states that referential expressions (R-expressions), such as *Ali* or *the teacher*, must be free everywhere. These principles were illustrated with both grammatical and ungrammatical English examples, showing

how violations of structural configurations result in interpretation errors.

The unit emphasized the importance of c-command and local domain in determining whether a binding relationship is permitted. Tree structures were used to clarify how syntactic positions affect the possibility of coreference or coindexation. The discussion extended to Indonesian, where reflexives such as *dirinya* exhibit binding behaviors that, while shaped by discourse factors, still reflect the same underlying syntactic constraints.

In addition, the unit addressed locality effects, showing how an anaphor cannot be bound by an antecedent that is too far removed in the structural hierarchy. Cross-linguistic evidence supported the claim that despite differences in morphological marking and surface forms, Binding Theory reveals deep universals in how languages manage reference and dependency.

Altogether, the unit reinforced the notion that syntactic structure governs interpretation in precise and predictable ways, and that Binding Theory offers a unified account of how different referential expressions behave within the grammar of natural languages.

GLOSSARY

- **Anaphor** – Reflexives/reciprocals requiring local antecedent.
- **Binding Domain** – Minimal clause/DP governing binding.
- **C-command in Binding** – Structural prerequisite for co-reference.
- **Principle A** – Anaphors bound in their local domain.
- **Principle B** – Pronouns free in their local domain.
- **Principle C** – R-expressions free everywhere (no c-commanding coreference).
- **Pronoun** – Refers independently; must avoid certain local antecedents.
- **R-expression** – Referring NP (e.g., proper name) must be free.



PART V:
SYNTAX IN ACTION



UNIT 12

ELLIPSIS AND EMPTY CATEGORIES

A. VP Ellipsis and Pseudogapping

Ellipsis refers to a syntactic phenomenon in which certain constituents are phonologically absent but semantically interpreted from the surrounding context. Among the various types of ellipsis, Verb Phrase (VP) Ellipsis and pseudogapping are two commonly discussed constructions within the theoretical syntax literature, especially in English. These ellipsis strategies raise crucial questions about constituency, identity, and the mechanisms of sentence derivation in the generative grammar framework (see Carnie, 2012; Johnson, 2001; Merchant, 2001; Radford, 2012; Sag, 1976).

1. VP Ellipsis in English

VP Ellipsis typically occurs when a verb phrase is omitted in the second conjunct of a coordination structure or a comparative clause, but the meaning is fully recoverable from the antecedent clause. This kind of ellipsis is dependent on the presence of an auxiliary verb that carries tense or aspectual information, while the main verb and its complements are elided.

Consider the example:

Mira could climb the tower, and Kito could ___ too.

In this sentence, the second conjunct lacks the verb phrase *climb the tower*, but the interpretation is recovered by copying the VP from the first conjunct. The ellipsis site is

structurally licensed by the auxiliary *could*, which syntactically supports the empty VP.

Structurally, VP ellipsis supports the idea of derivational identity between the elided VP and its antecedent. Theoretical treatments differ in whether the ellipsis site is filled at PF (Phonological Form) through copying or left empty with interpretation assigned at LF (Logical Form). Johnson (2001) and Merchant (2001) argue that VP ellipsis is constrained by syntactic isomorphism and identity conditions.

2. Pseudogapping

Pseudogapping is a more restricted and marked form of ellipsis in which only part of the VP is deleted, while certain constituents—typically objects or complements, remain overt. It usually appears in comparative or contrastive contexts and is often licensed by auxiliary verbs.

Example:

She would accept the offer, and he would __ the counter-proposal.

Here, the verb *accept* is omitted in the second clause, but the object *the counter-proposal* remains. This kind of ellipsis is more complex than full VP ellipsis because it requires movement of the remnant (*the counter-proposal*) out of the ellipsis site before deletion. Analyses by Lasnik (1999) and Merchant (2001) suggest that pseudogapping involves both movement and ellipsis, illustrating the interplay between syntax and PF-deletion.

Pseudogapping tends to exhibit rigid structural conditions. For instance, the remnant must contrast with an element in the antecedent clause, and the ellipsis site must correspond to a well-formed constituent.

3. Indonesian Comparison: Ellipsis and Productivity

In Indonesian, ellipsis phenomena are present but differ in frequency, productivity, and structural realization compared to English. While full VP ellipsis and NP ellipsis are attested and pragmatically licensed, pseudogapping is

significantly more restricted and less productive. Indonesian typically favors repetition or use of pro-forms over partial deletion.

Consider the example:

Mira bisa naik menara itu, dan Kito juga bisa __.

“Mira can climb that tower, and Kito can too.”

This sentence illustrates VP ellipsis in Indonesian. The auxiliary verb *bisa* (“can”) licenses the omission of the main verb phrase *naik menara itu* (“climb the tower”). The meaning is recoverable due to parallel structure and semantic context. While Indonesian lacks overt tense or agreement morphology, aspectual or modal markers such as *bisa* and *sudah* often serve as ellipsis licensors.

However, pseudogapping constructions are rare and often avoided in Indonesian:

**Mira akan menerima tawaran itu, dan Kito akan __ usulan balasan itu.*

Intended: “Mira will accept the offer, and Kito will [accept] the counter-proposal.”

This sentence, while modeled after English pseudogapping, feels unnatural to native speakers unless modified to repeat the verb or use a resumptive construction: *Mira akan menerima tawaran itu, dan Kito akan menerima usulan balasan itu juga.*

Such preferences suggest that Indonesian syntax avoids remnant movement out of ellipsis sites, possibly due to typological constraints on movement operations and a general reliance on discourse-level anaphora rather than syntactic ellipsis (Musgrave, 2001).

4. Theoretical Implications

The contrast between English and Indonesian regarding ellipsis underscores the parametric variation in how languages implement ellipsis mechanisms. English exhibits robust VP ellipsis and pseudogapping, supported by rich verbal inflection and auxiliary stacking, whereas

Indonesian exhibits limited ellipsis, constrained by its analytic structure and reliance on pragmatic recoverability.

In theoretical syntax, such patterns inform debates over:

- a. The locus of ellipsis (PF vs. LF)
- b. The role of auxiliaries and functional heads
- c. The interaction between movement and deletion
- d. Cross-linguistic constraints on ellipsis licensing (Goldberg, 2005; Merchant, 2001)

B. Null Subjects and PRO

In the study of syntax, especially within the framework of Government and Binding (GB) theory and its descendants, null subjects refer to syntactic positions that are structurally present but phonologically silent. In English and many other languages, null subjects are typically found in non-finite (infinitival) clauses. These positions are filled by a theoretical element known as PRO (pronounced as "big PRO" or /pro/), which plays a crucial role in control constructions, sentences where the subject of an embedded non-finite clause is understood but not overtly pronounced.

This section examines the behavior of PRO in English, compares it to subjectless constructions in Indonesian, and discusses the syntactic licensing conditions and interpretive constraints that govern such structures.

1. PRO and Control in English

In control constructions, a verb in the matrix (main) clause semantically controls the subject of an embedded infinitival clause. Since infinitival clauses in English lack tense, they do not license nominative case. As a result, the subject position of these clauses cannot be occupied by an overt NP or pronoun. Instead, this position is filled by PRO, an invisible pronoun that receives its reference from an antecedent in the matrix clause.

Consider the following sentence:

Ina tried [PRO to finish the puzzle before midnight].

Here, PRO functions as the subject of the embedded clause *to finish the puzzle before midnight*. Although not phonetically realized, PRO is syntactically projected in the subject position of the non-finite TP and receives its interpretation from the antecedent *Ina*. In this case, *Ina* is both the subject of the matrix verb *tried* and the understood subject of the embedded verb *finish*. This is known as a subject control configuration.

In terms of structure, the embedded clause has the following abstract representation:

[VP tried [CP [TP PRO to finish the puzzle before midnight]]]

Theoretical analyses (e.g., Chomsky, 1993; Landau, 2000) argue that PRO must be ungoverned or unbound by syntactic heads, hence why it appears in non-finite contexts where case assignment is unavailable. This constraint distinguishes PRO from other null elements like *pro* (in pro-drop languages) or traces left by movement.

2. Types of Control: Subject vs. Object

While **subject control** is more common, English also allows for object control, where the controller of PRO is the object of the matrix verb:

Ina persuaded Kito [PRO to leave early].

Here, *Kito* is the understood subject of *to leave early*, making it a case of object control. The verb *persuade* assigns a thematic role to both *Kito* and the embedded clause, and syntactic and thematic mechanisms identify *Kito* as the controller of PRO.

3. Cross-Linguistic Comparison: Indonesian Control Structures

Indonesian also exhibits control constructions in which the subject of the embedded clause is unexpressed and interpreted via context or the matrix argument. The use of infinitival constructions is marked by modals or purpose expressions like *untuk* ("to" or "in order to"), which function similarly to English *to* in non-finite contexts.

Consider the Indonesian equivalent:

Ina mencoba [untuk menyelesaikan teka-teki itu sebelum tengah malam].

“Ina tried to finish the puzzle before midnight.”

In this sentence, the subject of *menyelesaikan* (“to finish”) is not overtly stated, yet the interpretation is unambiguous: *Ina* is the understood subject of both clauses. Structurally, Indonesian does not insert an overt pronoun in the embedded clause, nor does it use morphological case marking to signal control. However, like in English, the subject of the matrix clause semantically controls the embedded verb’s subject position.

The parallel between English PRO and Indonesian null subjects in control contexts has been widely discussed in syntactic literature (see Landau, 2000; Musgrave, 2001). Indonesian may not overtly mark PRO-like elements, but the underlying syntactic dependency still reflects the properties of control.

4. Structural Licensing of PRO

In both languages, the presence of PRO is conditioned by:

- a. Clause type: Non-finite clauses that lack tense or agreement.
- b. Thematic dependency: Matrix verbs that assign a thematic role to the subject of the embedded clause.
- c. Syntactic position: PRO occupies the subject position of the embedded TP.
- d. Interpretive control: The antecedent is determined by the control verb and the argument structure.

These constraints align with the idea that PRO is syntactically projected but semantically dependent, governed not by movement or overt agreement but by lexical semantics and binding theory.

5. Theoretical Significance

The study of PRO enriches our understanding of argument structure, null elements, and the syntax-semantics interface. It also raises questions about modularity in grammar, how syntax and semantics interact to produce coherent interpretation even when elements are not pronounced.

Moreover, the comparison with Indonesian shows how similar control relations can emerge in languages with different morphosyntactic properties. Indonesian lacks overt subject agreement or case marking but still supports control interpretations, demonstrating that PRO-like structures can be parameterized across languages (see Landau, 2004; Polinsky & Potsdam, 2006).

C. Control vs Raising

One of the most fundamental distinctions in the theory of non-finite clause structure is the difference between control constructions and raising constructions. Though both involve infinitival complements and null subjects, they are syntactically and semantically distinct. This section explores their core differences, structural representations, diagnostic tests, and theoretical implications, especially within the Government and Binding (Chomsky, 1993) and Minimalist (Chomsky, 1995) frameworks.

1. Control Constructions

In control constructions, the matrix verb assigns a thematic role (theta-role) to the subject of the embedded infinitival clause. The subject of the non-finite clause is PRO, a null pronoun that receives its interpretation from an argument of the main clause. The type of control is determined by the argument structure of the matrix verb, leading to either subject control **or** object control.

Example of subject control:

Ari promised [PRO to call the doctor].

In this sentence, *Ari* is both the subject of the matrix verb *promised* and the understood subject of the embedded verb *call*. The embedded clause lacks an overt subject, and the null subject PRO is controlled by *Ari*. Crucially, the verb *promise* assigns a theta role to both the matrix subject (*Ari*) and the embedded clause. The embedded TP is a complement of the verb, and PRO occupies the embedded subject position.

Structurally:

[TP *Ari* [VP *promised* [TP PRO to call the doctor]]]

The presence of PRO in control constructions is mandated by the fact that the embedded clause does not receive case, but still requires a syntactic subject for interpretability (EPP). The controller of PRO is selected by the control verb according to lexical semantics (Landau, 2000).

2. Raising Constructions

In contrast, raising constructions involve no theta-role assignment by the matrix verb to the subject. Instead, the subject originates in the embedded clause and raises to the subject position of the main clause. The matrix verb in these constructions is semantically light or non-thematic, meaning it does not assign a theta role to its subject. These verbs often describe epistemic or evidential judgments (e.g., *seem*, *appear*, *happen*).

Raising example:

Ari seems [to know the answer].

Here, *Ari* is understood as the subject of the verb *know*, not of *seem*. The DP *Ari* originates in the embedded clause but raises to satisfy the EPP feature in the matrix TP. The embedded clause does not include PRO, since there is an overt DP subject that moves.

Structure:

[TP *Ari* [VP *seems* [TP *t* to know the answer]]]

The trace *t* marks the position from which *Ari* is raised. The raising verb *seem* does not assign a theta role to *Ari*; rather, *Ari* receives its role from the verb *know* in the

embedded clause. This makes raising verbs structurally and semantically distinct from control verbs.

3. Diagnostic Test: Expletive Insertion

A classic syntactic test to distinguish control from raising is the possibility of inserting an expletive subject (e.g., *it*, *there*) in the matrix clause. Since raising verbs do not assign a theta role to the subject, they can occur with non-referential expletives, which do not need theta roles. In contrast, control verbs require referential subjects because they assign theta roles.

Compare:

a. *It seems [that Ari knows the answer].*

✓ Grammatical: *seems* allows expletive subject.

b. *It promised [that Ari would call].*

X Ungrammatical: *promise* requires a referential subject to receive the agent theta-role.

This test reveals that raising verbs are structurally compatible with expletives, while control verbs are not. This provides empirical support for the claim that raising verbs do not select their subjects, while control verbs do.

4. Cross-Linguistic Comparison: Indonesian

In Indonesian, the distinction between control and raising is less morphologically marked due to the language's analytic structure and lack of overt agreement or case marking. However, semantic and syntactic differences between control-like and raising-like constructions are still observable.

Control example:

Ari berjanji [untuk menelepon dokter itu].

"Ari promised to call the doctor."

The verb *berjanji* ("promise") selects *Ari* as the controller of the null subject in the embedded clause, aligning with English subject control. The embedded verb *menelepon* ("call") has no overt subject, but the interpretation is controlled by *Ari*.

Raising-like example:

Ari tampaknya [tahu jawabannya].

“Ari seems to know the answer.”

The verb *tampaknya* (“seems”) does not assign a thematic role to *Ari*, which semantically belongs to the embedded predicate *tahu* (“know”). This structure corresponds to a raising construction, though Indonesian lacks overt movement evidence such as agreement or inflection.

Indonesian thus reflects the same conceptual distinction between thematic control and non-thematic raising, even though the diagnostic tools are less syntactically explicit than in English. Researchers such as Musgrave (2001) and Landau (2000) have argued that Indonesian relies more on lexical semantics and discourse cues to distinguish these structures.

5. Theoretical Implications

The distinction between control and raising supports a modular view of grammar, where syntax, argument structure, and interpretation interact but are governed by distinct principles. Control involves semantic selection and the presence of PRO, while raising is driven by morphosyntactic requirements such as the EPP and lacks theta-role assignment.

This distinction is also central to debates about movement, locality, and the interface between syntax and semantics. Raising structures often provide evidence for movement operations and locality constraints, while control structures reveal the interaction of lexical properties and binding.

D. Summary

This unit explored the syntax of ellipsis and empty categories, focusing on how languages like English and Indonesian handle structures that include omission, null elements, and invisible syntactic dependencies. In English, VP

Ellipsis involves the omission of an entire verb phrase that is recoverable from a previous clause. It is typically licensed by auxiliaries and is syntactically productive. A related but more complex phenomenon is pseudogapping, where the main verb is omitted but complements remain overt. Pseudogapping requires specific syntactic movement and is more restricted in distribution.

In contrast, Indonesian permits full VP or NP ellipsis, especially when context or discourse enables recovery of the missing material. However, pseudogapping is not productive in the language, and speakers tend to favor repetition or the use of pro-forms instead. These patterns reflect typological differences between English and Indonesian and underscore the role of syntax in organizing information structure and discourse economy.

The unit also addressed control and raising constructions, both of which involve embedded non-finite clauses. In control constructions, the embedded subject is an invisible pronoun called PRO, whose reference is determined by an argument in the matrix clause. The choice between subject control and object control depends on the lexical semantics of the matrix verb. Theoretical constraints on PRO include its occurrence in non-finite clauses, its inability to receive case, and its requirement for control by a higher argument.

By contrast, raising constructions involve movement of a DP from the embedded subject position to the matrix subject position. Crucially, raising verbs do not assign a theta role to their surface subjects, unlike control verbs. One diagnostic for this distinction is the use of expletives: raising verbs (e.g., *seem*) can occur with non-referential subjects like *it*, whereas control verbs (e.g., *promise*) cannot.

Although Indonesian lacks overt markers such as infinitival "to" or morphologically encoded case, similar control and raising distinctions are identifiable through semantic interpretation and argument structure. Indonesian speakers interpret null subjects in embedded clauses based on pragmatic

and structural cues, suggesting that PRO-like behavior and control relations are universal, even in languages with different surface morphologies.

Together, these constructions, ellipsis, PRO, control, and raising, highlight how syntactic theory accounts for invisible yet rule-governed structures. They reinforce the idea that syntax operates below the level of surface expression, mediating between grammatical constraints, meaning, and cross-linguistic variation.

GLOSSARY

- **Control** – Subject understood from matrix controller.
- **Indonesian Nulls** – Subjects may be contextually omitted in discourse.
- **Null Categories** – Unpronounced elements licensed by syntax.
- **PRO** – Null subject in non-finite/control clauses.
- **PRO vs. pro** – Big PRO (English control) vs. little pro (null subject languages).
- **Pseudogapping** – Ellipsis leaving a focused remnant.
- **Raising** – NP moves from embedded predicate (seems [t] tired).
- **VP Ellipsis** – Deleting VP under identity (She will $\emptyset$, and I will $\emptyset$ too).

UNIT 13

APPLIED PROBLEM SOLVING IN SYNTAX

A. GB Syntax Problems: Assignments and Commentary

Applied syntax is not merely about memorizing theoretical terms, it involves diagnosing abstract structure, identifying invisible elements, and testing theoretical claims against real-world data. In this section, we take a hands-on approach by analyzing syntactic phenomena using principles from the Government and Binding (GB) framework (Chomsky, 1993; Haegeman, 1994). Through careful examination of sentence structures, we apply concepts such as X-bar theory, theta roles, movement, case theory, and binding.

Each problem below includes a novel English sentence, its phrase structure, and an analysis of movement, argument structure, and syntactic licensing.

1. Problem 1: Raising and Embedded Predicates

a. Sentence:

The drone in the sky appeared to malfunction.

This sentence contains a raising verb (*appeared*) and an embedded infinitival clause. The apparent subject (*the drone in the sky*) is semantically interpreted as the agent of the verb *malfunction*, but syntactically occupies the subject position of the matrix clause. This is a classic example of subject-to-subject raising, where the subject originates in the embedded clause and moves into the higher clause to satisfy the Extended Projection Principle (EPP) and receive nominative case from the matrix tense.

b. Structural Analysis:

- 1) The matrix verb *appeared* is a non-thematic verb: it does not assign a theta-role to its surface subject.
- 2) The subject *the drone in the sky* originates as the subject of *malfunction*, an embedded infinitival predicate.
- 3) Since infinitives do not assign case, movement is required for case licensing.
- 4) The embedded clause (*to malfunction*) lacks tense, which blocks nominative case assignment.

c. X-bar Representation Summary:

Matrix TP:

- 1) DP: *The drone in the sky* (raised from embedded TP)
- 2) T': T [+past] + VP
 - a) V: *appeared*
 - b) Complement: [TP t to malfunction]

d. Theoretical Implications:

This structure reflects a raising-to-subject configuration. The subject DP raises out of the embedded infinitival clause because it cannot receive case within the non-finite domain. The raising operation satisfies both the EPP and the case filter. There is no PRO involved in this structure, as *appeared* does not assign a theta-role.

Relevant literature has long supported this analysis (Chomsky, 1993; Radford, 2012). The possibility of replacing the subject with an expletive confirms the raising nature of the verb:

It appeared to malfunction. ✓

The expletive *it* fills the subject position, indicating that the matrix verb *appeared* does not assign a semantic role.

2. Problem 2: Control and Passive within Infinitival Clauses

a. Sentence:

They asked to be given another attempt.

This sentence showcases a more complex structure involving:

- 1) Control: A null subject (PRO) in the embedded clause is interpreted as referring to *they*.
- 2) Passive infinitival construction: *to be given* is a passive VP embedded within an infinitival complement.

b. Analysis:

- 1) The verb *asked* is a subject control verb when used without an overt object.
- 2) The subject *they* controls PRO in the embedded clause.
- 3) The embedded clause *to be given another attempt* contains a passive verb phrase, where the subject PRO is the recipient of the action.

c. Syntactic Tree Summary:

Matrix TP:

- 1) Subject: *They*
- 2) T': T [+past] + VP
 - a) V: *asked*
 - b) Complement: [TP PRO to be given another attempt]

d. Case and Argument Structure:

- 1) *Asked* assigns a theta-role (experiencer) to its subject.
- 2) The embedded clause is an infinitival TP that cannot assign case.
- 3) *Be given* is a passive construction that lacks an external agent; the subject PRO receives the goal role.
- 4) The DP *another attempt* is the theme, assigned internal argument status within the VP.

e. Theoretical Implications:

This sentence demonstrates the interaction between control theory, theta role assignment, and voice alternations (active/passive) within infinitival clauses.

The presence of PRO confirms that the structure is a control configuration, and the embedded VP retains passive morphology despite being non-finite.

It also illustrates the modular nature of GB syntax, where:

- 1) PRO is interpreted via control (binding theory),
- 2) Case is regulated by structural position (case theory),
- 3) Voice affects argument realization (theta theory),
- 4) And movement is constrained by locality principles.

Unlike raising, this structure cannot accommodate an expletive subject:

It asked to be given another attempt. X

This sentence is ungrammatical because *asked* requires a referential subject to receive a theta-role, making it incompatible with non-referential expletives.

3. Pedagogical Exercises (Optional)

To reinforce understanding, students can:

- a. Diagram both sentences using X-bar trees and show movement paths.
- b. Identify theta roles, case assignments, and binding relationships.
- c. Compare these constructions with Indonesian equivalents (e.g., raising with *tampak* or control with *meminta* or *berusaha*).
- d. Test sentences using expletives or topicalization for further analysis.

B. Derivational Analyses and Tree Drawing

One of the most effective ways to internalize syntactic theory is through derivational analysis, a method in which one constructs the abstract syntactic structure of a sentence through a step-by-step series of operations, beginning with the underlying representation and ending with the surface form. This approach, rooted in the tradition of Government and Binding Theory (Chomsky, 1993) and carried forward in the Minimalist Program (Chomsky, 1995), enables learners to

visualize hierarchical structure, movement operations, case checking, and phrase projection.

In this section, we apply this technique to a representative English sentence, decomposing it through successive syntactic stages. Students are encouraged to draw accompanying X-bar-style trees for each stage and annotate them with movement paths, category labels, and structural boundaries such as TP (Tense Phrase), VP (Verb Phrase), and CP (Complementizer Phrase).

1. Sentence for Analysis

Those students seem to enjoy quantum riddles.

At first glance, this appears to be a simple declarative sentence. However, a closer syntactic analysis reveals a raising construction, where the subject of the matrix clause *those students* actually originates as the subject of the embedded infinitival clause. The matrix verb *seem* is a classic raising verb that does not assign a thematic role to its surface subject. The analysis therefore requires tracing movement and evaluating feature-checking processes across clause boundaries.

2. Step 1: Underlying Structure

In the underlying representation (D-structure or base structure), the subject DP *those students* is generated as the specifier of TP in the embedded infinitival clause headed by *to*. The embedded clause contains the verb *enjoy* and its complement, the DP *quantum riddles*. At this stage, *those students* has not yet moved to the matrix clause; it receives its theta-role (experiencer) from the embedded predicate *enjoy*. Structure (abstract):

[VP seem [TP *those students* to enjoy quantum riddles]]

This structure is not yet grammatical because the DP *those students* has no case, and the matrix TP lacks a subject to satisfy the EPP requirement, which demands a subject in the specifier of TP.

3. Step 2: Raising for Case and EPP

To satisfy case requirements and the EPP, *those students* must raise from the embedded subject position to the specifier of the matrix TP. This is an instance of A-movement (movement to an argument position), as the DP moves to receive nominative case from the finite T head in the matrix clause.

The derivation proceeds as follows:

- a. T (in matrix TP) has uninterpretable ϕ -features and EPP.
- b. *Those students* raises to Spec-TP in the matrix clause to check these features.
- c. The embedded subject position is left with a trace (t), preserving the interpretive link between the two positions.

Resulting structure:

[TP *those students* [T' T [+pres] [VP *seem* [TP t to enjoy quantum riddles]]]]

This tree satisfies all structural requirements:

- a. Theta-role is assigned by *enjoy* to *those students* in the embedded clause.
- b. Nominative case is assigned by matrix T to the moved subject.
- c. EPP is satisfied by having a DP in Spec-TP.
- d. The matrix verb *seem* remains thematically inert, which is consistent with its raising behavior.

4. Step 3: Tree Drawing with Functional Projections

To visually represent the derivation, we construct a tree with X-bar projections, labeling all categories. The tree must reflect the following hierarchical organization:

Matrix CP (optional, often covert in simple declaratives)

Matrix TP

- a. Specifier: DP *those students*
- b. T': T (e.g., [+present])
- c. VP: V *seem*

Complement: TP (embedded clause)

- 1) Specifier: trace of *those students*

- 2) T': T *to*
- 3) VP: V *enjoy*
- 4) DP: *quantum riddles*

Key tree features:

- a. The embedded TP is a non-finite clause, unable to license nominative case.
- b. The matrix TP is a finite clause, responsible for case assignment.
- c. Movement is upward and local, complying with locality constraints such as the Minimal Link Condition.

5. Summary and Pedagogical Reflection

This derivational analysis illustrates multiple foundational principles of GB-style syntax:

- a. Argument structure: The subject DP receives its theta-role in the embedded clause.
- b. Movement: A-movement occurs from embedded subject to matrix subject.
- c. Case theory: The DP requires movement to receive nominative case from matrix T.
- d. EPP satisfaction: Every finite TP requires a subject in its specifier.
- e. Tree geometry: Phrase structure rules and X-bar theory govern how projections are constructed.

By tracing the derivation step by step, learners can deepen their understanding of invisible syntactic dependencies, appreciate how structure underlies surface form, and gain confidence in syntactic reasoning.

C. Argumentation and Hypothesis Testing

Syntactic analysis is not merely about describing sentence structure; it is fundamentally an exercise in scientific reasoning. In linguistics, and especially within formal syntax, one of the most essential skills students must acquire is the ability to formulate hypotheses, make predictions, and test those predictions against native speaker intuitions or typological data.

This process parallels the scientific method in other disciplines and lies at the heart of how theories such as Government and Binding (GB) and the Minimalist Program are constructed and evaluated (Chomsky, 1993, 1995).

In this section, we illustrate this method through the example of raising versus control verbs, showing how formal diagnostics can be used to test hypotheses about the underlying structure of sentences.

Syntactic Question: Is "seem" a Raising or Control Verb?

Let us consider a syntactic inquiry: what type of verb is *seem*? This question is theoretically significant because the classification of a verb as either raising or control has substantial consequences for theta-role assignment, argument structure, movement, and null subject licensing.

1. Initial Hypothesis

The verb *seem* is a raising verb rather than a control verb. This hypothesis is grounded in the observation that *seem* does not assign a semantic role (e.g., agent, experiencer) to its grammatical subject. Instead, the subject receives its interpretation from the embedded predicate.

2. Theoretical Background

Raising verbs are characterized by their inability to assign theta roles to the subject position of the matrix clause. This contrasts with control verbs, which semantically select their subjects or objects as controllers of PRO in embedded infinitival clauses. According to standard analyses (see Carnie, 2012; Radford, 2012), *seem* lacks a thematic subject and instead relies on syntactic movement from the embedded clause to fill its subject position.

3. Empirical Test: Expletive Insertion

A powerful diagnostic for distinguishing raising verbs from control verbs involves the use of expletive subjects such as *it* or *there*. Since expletives are non-referential and cannot receive thematic roles, their compatibility with a given verb suggests that the verb does not assign a theta role to its subject. If an expletive is grammatical with the verb in

question, then that verb likely does not semantically select its subject and is therefore a raising verb.

Example 1 (Raising verb with expletive):

It seems that the answer is correct. ✓

This sentence is fully grammatical. The expletive *it* occupies the subject position of the matrix clause, but it does not bear any semantic content. This implies that *seem* does not assign a thematic role to its subject, consistent with the hypothesis that it is a raising verb.

Example 2 (Control verb with expletive):

It promised to leave early. ✗

This sentence is ungrammatical because *promise* assigns an agentive theta role to its subject. The expletive *it* cannot receive that role, and thus the structure fails.

The contrast in grammaticality between these examples strongly supports the claim that *seem* is a raising verb, whereas *promise* is a control verb. This test is widely accepted in syntactic diagnostics (Adger, 2003; Haegeman, 1994) and forms part of standard syntactic training.

4. Additional Supporting Evidence

a. Theta-role test:

Raising verbs like *seem*, *appear*, and *happen* do not assign semantic roles to their subjects. Their subjects are interpreted solely via the embedded clause.

b. PRO test:

Raising verbs do not license PRO in the embedded clause because they do not semantically select the subject. In contrast, control verbs require PRO, and the matrix subject or object semantically controls it.

1) *They promised [PRO to leave early].* ✓ Control

2) *They seemed [PRO to leave early].* ✗ Uninterpretable; no controller assigned

c. Passivization test:

In raising constructions, the subject can often be passivized within the embedded clause while maintaining a grammatical matrix structure.

Kai seems [to have been given a medal]. ✓

The subject of the passive embedded clause (*Kai*) is raised to the matrix subject position.

5. Application to Indonesian

Although Indonesian lacks overt expletives and rich morphological case marking, semantic and syntactic cues allow similar distinctions. For instance, the verb *tampak* ("seem") does not assign an agentive role to the subject and can host clauses in which the subject is interpreted via the embedded predicate:

Tampaknya [Rina tahu jawabannya].

"It seems that Rina knows the answer."

The subject *Rina* semantically belongs to the embedded verb *tahu*, not to *tampak*.

In contrast, control verbs such as *berjanji* ("promise") require an agentive subject, and inserting an expletive equivalent (e.g., generic *ada* or *itu*) yields unacceptable constructions.

These parallels illustrate the cross-linguistic robustness of raising versus control distinctions, even in languages with divergent morphosyntactic systems (Landau, 2000; Musgrave, 2001).

D. Summary

This unit focused on the application of syntactic theory through structured problem-solving and data analysis. It guided students through the process of analyzing sentence structure using principles from Government and Binding (GB) theory and the Minimalist framework, encouraging them to move beyond memorization toward active hypothesis testing and syntactic argumentation.

The unit began with step-by-step analyses of English sentences, demonstrating how to identify phrase structure, argument positions, and movement operations using derivational trees. It emphasized the importance of tracing syntactic elements from their base-generated positions through various stages of derivation to their surface forms. Through case studies, students practiced identifying raising and control structures, ellipsis patterns, and clause boundaries, while drawing on diagnostics such as theta roles, c-command, and locality constraints.

Derivational tree drawing was introduced as a method for visualizing syntactic dependencies and testing structural predictions. The unit showed how structural representations help reveal the internal organization of sentences, clarify ambiguous constructions, and distinguish between different types of movement (e.g., WH-movement vs. head movement).

In addition, the unit presented real-world syntax problems designed to train students in linguistic inquiry. These included interpreting ungrammaticality, evaluating constituency, and testing whether a verb licenses certain types of embedded clauses. Students were encouraged to formulate and revise hypotheses in response to new data, mirroring the process of syntactic research.

The comparative dimension was maintained through occasional references to Indonesian data, highlighting how familiar sentence types may obey the same structural principles despite differing in surface form. Overall, the unit reinforced the value of syntactic theory as a scientific tool and equipped students with practical strategies for analyzing both English and cross-linguistic data in a principled, theory-driven manner.

GLOSSARY

- **Argumentation** – Building evidence-based analyses.
- **Competing Analyses** – Compare predictions/explanatory power.
- **Corpus Check** – Real-world usage to triangulate intuitions.
- **Data Set** – Curated examples for hypothesis testing.
- **Derivational Steps** – Documenting movement and agreement checks.
- **Grammaticality Judgment** – Native-speaker acceptability ratings.
- **Tree Drawing Conventions** – Clear, labeled, and consistent diagrams.
- **Typological Comparison** – Cross-linguistic tests of proposals.

UNIT 14 | LANGUAGE VARIATION AND UNIVERSAL GRAMMAR

A. Syntactic Parameters: Word Order, Null Subjects

One of the central tenets of the Universal Grammar (UG) hypothesis, as proposed by Chomsky (1986, 1993), is that while all human languages share an underlying structural blueprint, they may vary along a limited set of syntactic parameters. These syntactic parameters represent points of variation that are fixed during early language acquisition based on linguistic input. This approach helps explain both the profound similarities and systematic differences among the world's languages. Two of the most studied parameters in syntactic theory are the Word Order Parameter and the Null Subject Parameter.

1. Word Order Parameter

A fundamental aspect of syntactic variation lies in constituent ordering, particularly the sequence of Subject (S), Verb (V), and Object (O) in canonical declarative clauses. The Word Order Parameter accounts for this cross-linguistic variation by positing that UG allows different languages to set different default word order templates, such as SVO, SOV, VSO, etc.

English is canonically SVO:

The fox chased the squirrel.

- a. Subject: *The fox*
- b. Verb: *chased*
- c. Object: *the squirrel*

This ordering is rigid in English due to a relatively fixed word order system. Movement for focus or topicalization is limited and marked, and syntactic functions (subject, object) are largely determined by position rather than morphology.

Indonesian, despite being typologically different in terms of morphology and inflection, also follows a basic SVO order in unmarked declarative clauses:

Rubah itu mengejar tupai itu.

- a. Subject: *Rubah itu* (the fox)
- b. Verb: *mengejar* (chased)
- c. Object: *tupai itu* (the squirrel)

This structural similarity suggests that the Word Order Parameter in both English and Indonesian is set to SVO. However, it is crucial to note that Indonesian exhibits greater discourse flexibility, allowing for pragmatically motivated reorderings such as topicalization (*Tupai itu dikejar oleh rubah itu*), especially in passive constructions or cleft sentences. Nevertheless, the default parametric setting remains SVO in both languages (Musgrave, 2001; Sneddon et al., 2010).

2. Null Subject Parameter

Another well-documented locus of syntactic variation concerns the Null Subject Parameter, also known as the pro-drop parameter. This parameter distinguishes languages that require an overt grammatical subject in every tensed clause from those that allow null subjects under certain grammatical or discourse conditions.

In English, the parameter is set to [+subject], meaning that all tensed clauses must have an overt subject. Subject omission in declaratives leads to ungrammaticality:

- a. *Is going to the market.* ✗
- b. *He is going to the market.* ✓

Even in imperatives or casual registers, English allows subject omission only in highly restricted cases, such as commands (*Go now!*), which are licensed by pragmatic context and not syntactic permissiveness.

In Indonesian, however, the system is more flexible. The subject can be dropped in many contexts without affecting grammaticality, especially when its referent is recoverable from discourse. This makes Indonesian [$\pm$ subject], or a partial pro-drop language (Landau, 2000; Musgrave, 2001). For example:

Pergi ke pasar tadi.

“(Someone) went to the market earlier.”

The subject is not pronounced but is pragmatically understood.

This capacity for subject omission is linked not to rich verbal agreement, as in Romance pro-drop languages like Spanish or Italian, but rather to a combination of discourse-pragmatic factors, topic prominence, and semantic recoverability. This aligns Indonesian more closely with topic-drop languages, such as Chinese or Japanese, than with classical null subject languages governed by rich inflectional agreement (Huang, 1984).

3. Theoretical Implications

These parametric differences reinforce the claim that Universal Grammar allows for constrained syntactic variation. Each child acquiring a language encounters linguistic input and sets parameters accordingly. The child exposed to English data will eventually fix the Null Subject Parameter to [$+$ subject], while the child acquiring Indonesian data will recognize the availability of null subjects in pragmatically licensed contexts.

Furthermore, these parameters are not isolated, they interact with other elements of the grammar, including case licensing, tense specification, agreement morphology, and information structure. For instance, in English, the lack of agreement in non-finite clauses still prohibits null subjects

("To go to the store is fun" is acceptable only when *to go* is interpreted as a gerundive NP). In Indonesian, by contrast, even finite clauses can accommodate null subjects due to the topic-oriented discourse structure of the language.

B. Cross-linguistic Syntax: Indonesian Case Only

Cross-linguistic syntax examines the extent to which natural languages conform to universal principles while exhibiting surface-level variation in their morphosyntactic properties. Indonesian provides a rich empirical site for exploring syntactic phenomena due to its analytic morphology, pragmatic flexibility, and structural simplicity in overt inflection. As a member of the Western Malayo-Polynesian branch of the Austronesian family, Indonesian differs significantly from Indo-European languages like English in its syntactic realizations. Nevertheless, the foundational structures proposed by Universal Grammar, such as argument structure, clause projection, and movement operations, still underlie its surface forms.

This section examines several core syntactic features of Indonesian, word order, subject omission, voice alternations, topic structures, and argument realization, while situating them within a cross-linguistic theoretical framework.

1. Basic Clause Structure and Word Order

Indonesian exhibits a basic Subject-Verb-Object (SVO) word order, which it shares with English and many other languages. However, unlike English, this order is not enforced by rich inflectional morphology but rather emerges from information structure and topic-prominence. Consider the following example:

Rina membaca buku itu.

"Rina read the book."

Here, *Rina* functions as the subject, *membaca* as the verb, and *buku itu* as the object. The subject precedes the verb, aligning Indonesian with head-initial languages. However, due to the absence of subject-verb agreement morphology and case marking, constituents can be rearranged under

specific discourse conditions without ungrammaticality. This flexibility distinguishes Indonesian from configurational languages like English (Musgrave, 2001; Sneddon et al., 2010).

2. Null Subjects and Topic Drop

A defining syntactic property of Indonesian is its ability to omit subjects, especially when referents are clear from context. This phenomenon has been analyzed as topic drop, where the topical element is omitted because it is highly accessible in the discourse. While Indonesian is not a canonical pro-drop language (like Italian or Spanish), it supports null subject constructions in both imperative and declarative clauses.

Sedang makan nasi goreng.

“(He/She) is eating fried rice.”

The above construction is grammatically complete and semantically transparent despite lacking an overt subject. This kind of null subject is acceptable in matrix and embedded contexts, provided that pragmatic salience is preserved (Cole et al., 2006). Unlike canonical pro-drop languages, Indonesian does not rely on agreement morphology to license null subjects. Instead, omission is permitted because of pragmatic recoverability, a property shared with other topic-prominent languages such as Chinese and Japanese (Huang, 1984).

3. Voice Alternations and Argument Structure

Indonesian exhibits a morphologically rich **voice system**, distinguishing it from many Indo-European languages. The two primary voices are:

- a. Active Voice, marked by the prefix *meN-* on the verb.
- b. Passive Voice, typically marked by the prefix *di-*.

Examples:

- a. Active: *Dia menulis surat itu.*

“He wrote the letter.”

- b. Passive: *Surat itu ditulis oleh dia.*

“The letter was written by him.”

In active constructions, the subject functions as the agent, while in passive constructions, the theme is fronted to subject position. Importantly, Indonesian allows passive clauses without an overt agent, as in:

Surat itu ditulis kemarin.

“The letter was written yesterday.”

This construction indicates a syntactic and morphological strategy for argument suppression, which aligns with lexical mapping theory (Bresnan & Kanerva, 1989) and voice morphology studies in Austronesian languages (Kroeger, 2004). The passive construction in Indonesian supports the analysis of syntactic functions (subject, object) being determined by syntactic position and voice marking, rather than by agreement morphology.

4. Topicalization and Focus

Due to its flexible constituent order, Indonesian permits various forms of topicalization and focalization through fronting. These constructions are syntactically well-formed but often pragmatically marked. Consider:

Buku itu Rina baca kemarin.

“That book, Rina read yesterday.”

Here, *buku itu* is topicalized for emphasis or contrast. The movement does not disrupt the syntactic licensing of arguments but reflects the high salience of the topic. Such constructions require the hearer to presuppose the topical referent, further reinforcing Indonesian’s discourse-configurational properties.

Topicalization in Indonesian differs from English cleft constructions, which rely on additional syntactic material (e.g., “It was the book that Rina read”). In Indonesian, such structures are constructed more economically, often without a copula or overt relative marker.

5. Argument Licensing and Case

Indonesian lacks morphological case marking on nouns, yet syntactic arguments are licensed by structural positions and predicate-argument relations. In active constructions, the external argument (agent) precedes the verb, while the internal argument (theme) follows. In passive constructions, the theme becomes the subject, satisfying the Extended Projection Principle (EPP).

The lack of case morphology challenges traditional Case Theory (Chomsky, 1993), but empirical research suggests that abstract case is still assigned covertly. This view is supported by the fact that unlicensed noun phrases are rare and generally avoided in standard Indonesian, suggesting that a syntactic mechanism ensures argument saturation and grammatical completeness (Musgrave, 2001).

C. Implications for Language Acquisition

The study of cross-linguistic syntax, particularly in relation to Universal Grammar (UG), holds profound implications for understanding how children acquire language. It invites fundamental questions: How do infants, equipped with limited cognitive resources and exposure, master the complex and often subtle rules of syntax? How do they resolve variation among languages while still converging on grammatically accurate systems? This section addresses these questions by exploring how parameter setting, input frequency, and functional structure interact during the process of language acquisition, with a focus on Indonesian as a typologically rich case study.

1. The Poverty of the Stimulus and the Role of UG

One of the central arguments for Universal Grammar is the Poverty of the Stimulus (Chomsky, 1980; Pinker, 1994). This argument suggests that the linguistic input available to children is insufficient to explain the rapid, uniform, and largely error-free development of complex syntactic knowledge. Children are rarely given explicit grammatical

instruction, nor do they receive negative evidence (i.e., information about what is ungrammatical). Yet, by the age of five, they produce and comprehend structurally intricate sentences.

Universal Grammar is posited as a biologically endowed cognitive system that constrains the hypothesis space children consider during acquisition. It provides a set of universal principles, such as structure-dependence, locality, and projection, and a finite inventory of parameters, such as Null Subject and Head Directionality, which must be fixed based on the child's linguistic environment (Chomsky, 1993; White, 2012).

2. Parameter Setting and Early Input

As described in earlier sections, Indonesian exhibits properties that differ from rigidly configurational languages like English. For instance, its tolerance of null subjects under discourse-governed conditions and its use of morphological voice alternations present children with surface variability that must be interpreted against underlying grammatical constants.

When acquiring Indonesian, children must determine that:

- a. Overt subject pronouns may be omitted in contexts where the referent is recoverable (Cole et al., 2006).
- b. Verbs do not inflect for agreement features (person, number, gender), but aspectual and voice morphology still constrains argument realization (Musgrave, 2001).
- c. Clause structure still conforms to a hierarchical projection, including TP and CP layers, even if these are not morphologically marked (Sneddon et al., 2010).

Thus, parameter setting in Indonesian requires the child to deduce, for example, that the Null Subject Parameter is set to a topic-drop configuration rather than full pro-drop or obligatory overt-subject grammar. This is not derived from explicit instruction but inferred from consistent patterns in the ambient language.

3. The Acquisition of Functional Categories

Functional projections such as CP (Complementizer Phrase), TP (Tense Phrase), and vP (light verb phrase) play a crucial role in syntax. Although they are not always overt in Indonesian, due to the absence of inflectional affixes or complementizers like "that" in English, their presence is revealed through syntactic behavior, such as clause embedding, word order, and interpretation of tense and aspect.

Children acquiring Indonesian must abstract these invisible layers from the input. For instance:

- a. *Saya pikir bahwa dia pergi.*

"I think that he left."

Here, the complementizer *bahwa* signals the CP layer explicitly. However, it can be omitted in many contexts, especially in colloquial or child-directed speech:

- b. *Saya pikir dia pergi.*

"I think he left."

Even in the absence of *bahwa*, the child must still interpret the embedded clause as a syntactic unit—a challenge that supports the claim that UG guides learners to assume hierarchical, not merely linear, structure (Radford, 2004; O'Grady et al., 2017).

4. Raising and Control: Acquisition Challenges

Complex constructions such as raising and control pose significant challenges in acquisition, especially because they involve non-canonical subject positions, null elements (e.g., PRO), and non-local dependencies.

In English, children often delay mastery of raising verbs (like *seem*, *appear*) and confuse them with control verbs (like *try*, *promise*). Indonesian presents similar complexities, though the surface structure differs. For example, constructions involving *tampak* ("seem") or *coba* ("try") require learners to parse subjects and arguments distributed across clause boundaries.

- a. *Ani mencoba menyelesaikan tugas itu.*
 “Ani tried to finish the assignment.” → Control: PRO subject
- b. *Ani tampak lelah.*
 “Ani seems tired.” → Raising: subject originates in embedded predicate

The child must eventually understand that no theta-role is assigned by *tampak*, while *mencoba* assigns an agentive role to its subject, thus triggering PRO insertion in the embedded clause. These subtle distinctions reinforce the necessity of internal grammatical constraints, as surface cues alone may be insufficient.

5. Indonesian in the Context of UG

Indonesian provides compelling evidence for the theory that even “less inflected” languages conform to the deeper architectural principles of Universal Grammar. Although it lacks overt case marking, agreement, and inflectional tenses, it still:

- a. Distinguishes between grammatical functions using word order and voice morphology.
- b. Supports hierarchical phrase structure with abstract projections (cp, tp, vp).
- c. Obeys constraints such as binding theory, island constraints, and control/raising distinctions.

Therefore, language acquisition in Indonesian involves uncovering not only overt syntactic regularities but also covert structural dependencies, which children appear able to infer with remarkable speed and consistency. These findings underscore the view that UG provides not a fixed grammar but a generative mechanism for deriving language-specific grammars from limited and variable input.

D. Summary

This unit explored the syntactic architecture of Indonesian through the lens of cross-linguistic comparison and Universal Grammar (UG). Although Indonesian lacks the rich morphological agreement systems found in many Indo-European languages, it nonetheless adheres to fundamental syntactic principles posited by generative grammar. Indonesian consistently exhibits SVO word order, but with greater flexibility driven by discourse. This typological variation is captured by the Word Order Parameter, which is set similarly in both English and Indonesian, though realized differently due to differences in morphosyntactic marking.

A central area of contrast lies in subject expression. The Null Subject Parameter distinguishes English, which strictly requires overt subjects ([+subject]), from Indonesian, which permits subject omission in pragmatically recoverable contexts. This makes Indonesian a topic-drop or partial pro-drop language, where syntactic structure interacts dynamically with discourse salience rather than morphologically enforced subject agreement.

Indonesian also features a productive voice system, which allows agent suppression and argument alternation through derivational morphology, especially via the *di-* and *me-* prefixes. Additionally, the language allows topicalization and focalization through fronting constructions that do not violate syntactic constraints, reinforcing the interaction between grammatical structure and information structure.

Despite surface-level differences, Indonesian conforms to deep universal syntactic constraints, such as the licensing of arguments, clause structure, and constituent behavior. It adheres to abstract Case theory even in the absence of overt case morphology and supports hierarchical phrase structures like CP, TP, and vP, albeit often covertly.

In terms of language acquisition, children acquiring Indonesian rely on discourse cues, structural templates, and voice morphology to set parameters for clause structure and

argument realization. The acquisition of control, raising, and null subject constructions reflects the operation of internal syntactic constraints, rather than mere reliance on input frequency. These findings reinforce the idea that Universal Grammar accommodates surface variation while preserving cross-linguistic uniformity in acquisition trajectories.

Taken together, the study of Indonesian syntax offers compelling evidence for the modular architecture of grammar, in which typologically diverse languages instantiate the same universal principles through parameterized settings. Indonesian stands as a valuable empirical case for understanding how deep syntactic structure emerges in languages with minimal overt morphology, yet rich syntactic behavior.

GLOSSARY

- **Acquisition and UG** – How learners set parameters.
- **Implications for Teaching** – Leveraging L1/L2 contrasts in pedagogy.
- **Indonesian vs. English** – Shared SVO with distinct morphology and wh-strategies.
- **Microvariation** – Dialectal/sociolectal differences.
- **Null Subject Parameter** – Languages allowing subject drop.
- **Parameters** – Switch-like settings yielding cross-linguistic variation (e.g., head direction).
- **Polysynthesis/Agglutination/Isolating** – Morphological profiles.
- **Word Order Typology** – SVO/SOV/VSO patterns across languages.

REFERENCES

- Aarts, B. (2008). *English syntax and argumentation* (3rd ed.). Palgrave Macmillan. <https://books.google.co.id/books?id=8MFoAAAAMAAJ>
- Adelaar, K. A., & Himmelmann, N. (2004). *The Austronesian languages of Asia and Madagascar*. Taylor & Francis. <https://books.google.co.id/books?id=OtWdhQ7vc6kC>
- Adger, D. (2003). *Core syntax: a minimalist approach*. Oxford University Press. <https://doi.org/10.1093/oso/9780199243709.001.0001>
- Baker, M. C. (1988). *Incorporation: a theory of grammatical function changing*. University of Chicago Press.
- Baker, M. C. (2002). *The atoms of language*. Oxford University Press. <https://books.google.co.id/books?id=7NDIAAAAMAAJ>
- Baker, M. C. (2003). *Lexical categories: verbs, nouns and adjectives*. Cambridge University Press. <https://books.google.co.id/books?id=EdR5836FdgMC>
- Baker, M. C. (2008). *The syntax of agreement and concord*. Cambridge University Press. https://books.google.co.id/books?id=-_L3MXEDOiuC
- Bresnan, J., & Kanerva, J. M. (1989). Locative inversion in Chicheŵa: a case study of factorization in Grammar. *Linguistic Inquiry*, 20(1), 1-50.
- Carnie, A. (2012). *Syntax: a generative introduction* (3rd ed.). John Wiley & Sons. <https://books.google.co.id/books?id=zWU5Rs2h9nAC>
- Cheng, L. L. S. (1991). *On the typology of wh-questions* [Doctoral dissertation]. Massachusetts Institute of Technology.
- Chomsky, N. (1962). *Syntactic structures*. Mouton. <https://books.google.co.id/books?id=FSxZAAAAMAAJ>

- Chomsky, N. (1965). *Aspects of the theory of syntax* (50th ed.). The MIT Press. <http://www.jstor.org/stable/j.ctt17kk81z>
- Chomsky, N. (1973). Conditions on transformations. In S. Anderson & P. Kiparsky (Eds.), *A Festschrift for Morris Halle* (pp. 232–286). Holt, Rinehart & Winston.
- Chomsky, N. (1980). Rules and representations. *Behavioral and Brain Sciences*, 3(1), 1–15. <https://doi.org/10.1017/S0140525X00001515>
- Chomsky, N. (1986). *Knowledge of language: its nature, origin, and use*. Bloomsbury Academic. <https://books.google.co.id/books?id=b0VZPtZDL8kC>
- Chomsky, N. (1993). Lectures on government and binding. In *The Pisa Lectures*. De Gruyter Mouton. <https://doi.org/10.1515/9783110884166>
- Chomsky, N. (1995). *The minimalist program*. MIT Press. <https://books.google.co.id/books?id=vtPQiYCNpjgC>
- Chomsky, N. (2001). Derivation by phase. In Ken Hale (Ed.), *A Life in Language* (Vol. 36, pp. 1–52). The MIT Press. <https://doi.org/10.7551/mitpress/4056.003.0004>
- Cinque, G. (1999). *Adverbs and functional heads: a cross-linguistic perspective*. Oxford University Press. <https://books.google.co.id/books?id=6SQLAQAAMAAJ>
- Clifton, C., & Frazier, L. (1989). Comprehending sentences with long-distance dependencies. In G. N. Carlson & M. K. Tanenhaus (Eds.), *Linguistic Structure in Language Processing* (pp. 273–317). Springer Netherlands. https://doi.org/10.1007/978-94-009-2729-2_8
- Cole, P., Hermon, G., & Huang, C.-T. J. (2006). Long-distance binding in Asian languages. In *The Blackwell Companion to Syntax* (pp. 21–84). <https://doi.org/10.1002/9780470996591.ch39>

- Cook, V., & Newson, M. (2014). *Chomsky's universal grammar: an introduction*. Wiley-India. <https://books.google.co.id/books?id=3HkU1dMezEMC>
- Dryer, M. S. (2007). Clause types. In T. Shopen (Ed.), *Language Typology and Syntactic Description: Clause Structure* (2nd ed., Vol. 1, pp. 224–275). Cambridge University Press.
- Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The evolution of the language faculty: clarifications and implications. *Cognition*, 97(2), 179–210. <https://doi.org/10.1016/j.cognition.2005.02.005>
- Frazier, L. (1987). Sentence processing: a tutorial review. In M. Coltheart (Ed.), *Attention and Performance XII*. Routledge. <https://doi.org/10.4324/9781315630427>
- Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6(2), 78–84. [https://doi.org/10.1016/S1364-6613\(00\)01839-8](https://doi.org/10.1016/S1364-6613(00)01839-8)
- Goldberg, L. M. (2005). *Verb-stranding VP ellipsis : a cross-linguistic study* [Doctoral dissertation]. McGill University.
- Grimshaw, J. (1992). *Argument structure*. The MIT Press.
- Haegeman, L. (1994). *Introduction to government and binding theory* (2nd ed.). Wiley-Blackwell.
- Halle, M., & Marantz, A. (1993). Distributed morphology and the pieces of inflection. In K. Hale & S. J. Keyser (Eds.), *The view from building* (pp. 111–176). The MIT Press.
- Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science* (New York, N.Y.), 298(5598), 1569–1579. <https://doi.org/10.1126/science.298.5598.1569>
- Heim, I. (1998). *Semantics in generative grammar*. Wiley. <https://books.google.co.id/books?id=tkYomwEACAAJ>

- Huang, C. T. J. (1982). *Logical relations in Chinese and the theory of grammar* [Doctoral dissertation]. Massachusetts Institute of Technology.
- Huang, C. T. J. (1984). On the distribution and reference of empty pronouns. *Linguistic Inquiry*, 15(4), 531–574.
- Huddleston, R., & Pullum, G. K. (2012). *The Cambridge grammar of the English language*. Cambridge University Press. <https://doi.org/10.1017/9781316423530>
- Hymes, D. H. (1972). On communicative competence. In J. B. Pride & J. Holmes (Eds.), *Sociolinguistics: Selected Readings* (pp. 269–293). Harmondsworth: Penguin.
- Jackendoff, R. (1977). *X-bar syntax: a study of phrase structure*. The MIT Press.
- Jackendoff, R. S. (1990). *Semantic structures*. The MIT Press.
- Johnson, K. (2001). What VP ellipsis can do, and what it can't, but not why. In M. Baltin & C. Collins (Eds.), *The Handbook of Contemporary Syntactic Theory* (pp. 439–479). Blackwell Publishers Ltd. <https://doi.org/10.1002/9780470756416.ch14>
- Kroeger, P. (2004). *Analyzing syntax: a lexical-functional approach*. Cambridge University Press. <https://books.google.co.id/books?id=fIAleOSnXpIC>
- Landau, I. (2000). *Elements of control: structure and meaning in infinitival constructions*. Kluwer Academic Publishers. <https://books.google.co.id/books?id=dO5sAAAAIAAJ>
- Landau, I. (2004). The scale of finiteness and the calculus of control. *Natural Language & Linguistic Theory*, 22(4), 811–877. <http://www.jstor.org/stable/4048084>
- Lasnik, H. (1999). *Pseudogapping puzzles* (Fringe Grammar).
- Lasnik, H., Uriagereka, J., & Boeckx, C. (2005). *A course in minimalist syntax: foundations and prospects*. Wiley-Blackwell.

- Levin, B., & Hovav, M. R. (1995). *Unaccusativity: At the syntax-lexical semantics interface*. The MIT Press.
- Merchant, J. (2001). *The syntax of silence: sluicing, islands, and the theory of ellipsis*. Oxford University Press. <https://books.google.co.id/books?id=dVfl5ZCweiQC>
- Musgrave, S. (2001). *Non-subject arguments in Indonesian subject arguments in Indonesian subject arguments in Indonesian* [Dissertation]. The University of Melbourne .
- O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2009). *Contemporary linguistics*. Bedford/St. Martin's. <https://books.google.co.id/books?id=dshpPgAACAAJ>
- Payne, T. (2006). *Exploring language structure: a student's guide*. Cambridge University Press. <https://books.google.co.id/books?id=LhSQQgAACAAJ>
- Pinker, S. (1994). *The language instinct: how the mind creates language*. Penguin Books.
- Polinsky, M., & Potsdam, E. (2006). Expanding the scope of control and raising. *Syntax*, 9(2), 171–192. <https://doi.org/10.1111/j.1467-9612.2006.00090.x>
- Pollock, J.-Y. (1989). Verb movement, universal grammar, and the structure of IP. *Linguistic Inquiry*, 20(3), 365–424.
- Radford, A. (2009). *Analysing English sentences: a minimalist approach*. Cambridge University Press. <https://books.google.co.id/books?id=oXBwS7q2sFAC>
- Radford, A. (2012). *English syntax: an introduction*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511841675>
- Reinhart, T. (2016). *Anaphora and semantic interpretation*. Routledge. <https://doi.org/10.4324/9781315536965>
- Reinhart, T. M. (1976). *The syntactic domain of anaphora* [Doctoral Dissertation]. Massachusetts Institute of Technology.

- Rizzi, L. (1990). *Relativized minimality*. The MIT Press.
- Rizzi, L. (1997). The fine structure of the left periphery. In L. Haegeman (Ed.), *Elements of Grammar: Handbook in Generative Syntax* (pp. 281–337). Springer Netherlands. https://doi.org/10.1007/978-94-011-5420-8_7
- Rizzi, L., & Cinque, G. (2016). Functional categories and syntactic theory. *Annual Review of Linguistics*, 2(Volume 2, 2016), 139–163. <https://doi.org/10.1146/annurev-linguistics-011415-040827>
- Ross, J. R. (1967). *Constraints on variables in syntax* [Doctoral dissertation]. Massachusetts Institute of Technology.
- Safir, K. (2004). *The syntax of anaphora*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780195166132.001.0001>
- Sag, I. A. (1976). *Deletion and logical form* [Doctoral dissertation]. Massachusetts Institute of Technology.
- Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the full transfer/full access model. *Second Language Research*, 12(1), 40–72. <https://doi.org/10.1177/026765839601200103>
- Sneddon, J. N., Adelaar, K. A., Djenar, D., & Ewing, M. (2010). *Indonesian: a comprehensive grammar*. Routledge. <https://doi.org/10.4324/9780203720882>
- van Gelderen, E. (2002). *An introduction to the grammar of English: syntactic arguments and socio-historical background*. John Benjamins Pub. <https://books.google.co.id/books?id=MrOZ4jOs8mgC>
- van Valin, J. , R. D. (2012). *An introduction to syntax*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139164320>
- White, L. (2012). *Second language acquisition and universal grammar*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511815065>

APPENDICES

Appendix A: Syntax Problem Sets by Unit

Unit 1. What Is Syntax?

1. Define the distinction between descriptive and prescriptive grammar. Provide one example for each from both English and Indonesian.
2. Analyze the imperative sentence 'Run!' in English and its Indonesian equivalent. Discuss the implications for subject representation in syntax.
3. Explain the scientific method as applied in syntax using your own examples.
4. Distinguish between competence and performance using Chomsky's framework. Provide sentence examples that illustrate the difference.

Unit 2. Categories and Parts of Speech

1. Identify the lexical and functional categories in the following sentence: 'The fast fox jumped over the lazy dog.'
2. Provide three Indonesian sentences and classify each word by category (N, V, Adj, Adv, etc.).
3. Using morphological, distributional, and substitutional tests, justify the category of the word 'gedrul'.
4. Explain the differences between open and closed word classes using English and Indonesian data.

Unit 3. Constituency and Phrase Structure

1. Use constituency tests (substitution, movement, coordination) to analyze the sentence: 'The girl in the red coat smiled.'
2. Write phrase structure rules that generate the sentence: 'The old man reads the newspaper in the morning.'
3. Draw a phrase structure tree for: 'The clever student with the glasses solved the puzzle.'
4. Discuss structural ambiguity in the sentence: 'She watched the man with binoculars.' What are the two readings?

Unit 4. X-bar Theory and Sentence Structure

1. Apply X-bar schema to analyze the noun phrase: 'Those ideas about marsupials.'
2. Identify heads, complements, adjuncts, and specifiers in: 'The tall teacher from Canada smiled warmly.'
3. Draw a full CP structure for: 'That Mira saw the painting surprised everyone.'
4. Compare the structure of noun phrases in English and Indonesian using X-bar theory.

Unit 5. Structural Relations in Syntax

1. Define domination, precedence, and c-command. Use a tree to illustrate these relationships.
2. Analyze the binding relations in: 'Lara painted herself with glowing ink.' Identify the antecedent and domain.
3. Explain how c-command accounts for ambiguity in: 'Every guard watched his reflection.'
4. Draw a syntax tree that demonstrates a binding violation and explain why it's ungrammatical.

Unit 6. Clausal Architecture

1. Illustrate the structure of CP and TP using the sentence: 'That she left surprised me.'
2. Compare how English and Indonesian express clause types using complementizers.
3. Analyze the embedded clause in: 'It is strange that Mira found the tunnel.' Provide a full tree.
4. Explain the functional layers of sentence structure and how they relate to clause typing.

Unit 7. Auxiliaries and Structural Evidence

1. Provide examples of auxiliary ellipsis in English and Indonesian. Explain the structure of the elided VP.
2. Arrange the auxiliaries in this English sentence in the correct order: 'She __ have __ being __ observed.'
3. Analyze the morphology in: 'She is walking.' Why is *'She is walk' ungrammatical?

4. Compare auxiliary stacking between English and Indonesian. Discuss flexibility and hierarchy.

Unit 8. Questions and Movement

1. Rewrite the sentence 'Ewan will travel to Mars' as a yes-no question in both English and Indonesian.
2. Analyze the WH-movement in: 'What did she discover?' and compare with Indonesian: 'Apa yang dia temukan?'
3. Identify island constraints in: '*What did you hear the rumor that she bought ___?'
4. Draw a tree for a subject and object WH-question. Explain movement paths.

Unit 9. Transformations and Derivations

1. Contrast 'Jamal sings' and 'Jamal does not sing'. Explain the role of do-support.
2. Discuss head movement and affix hopping in: 'She walks.' Draw the underlying and surface structures.
3. Explain how derivations work in syntax using: 'He ate the sandwich.' What is the deep structure?
4. Compare derivational simplification in English and Indonesian with examples.

Unit 10. Theta Theory and Case

1. Identify the theta roles in: 'The child imagined a world made of jelly.'
2. Explain the thematic hierarchy using an original sentence. Describe its influence on word order.
3. What is the EPP? Analyze: 'There appeared to be a comet over the forest.'
4. Identify Case assignment in: 'It seems that Ziya has vanished.' Compare with Indonesian equivalent.

Unit 11. Binding Theory

1. Label each NP as an anaphor, pronoun, or R-expression: 'Juno believed her friend doubted her.'

2. Test Principles A, B, and C using the following sentence: '*He thinks that Ali will win.'
3. Analyze domain locality in: 'The cat near him feared himself.'
4. Discuss cross-linguistic differences in binding using Indonesian reflexives.

Unit 12. Ellipsis and Empty Categories

1. VP Ellipsis Recognition and Reconstruction

Analyze the following sentences and reconstruct the elided verb phrases:

- a. *Rani will read the proposal, and Arman will __ too.*
- b. *The students have completed the assignment, and the lecturer has __ as well.*
 - 1) Identify the auxiliary verb that licenses ellipsis.
 - 2) What is the underlying full form of each sentence?

2. Pseudogapping Contrast

Distinguish between VP ellipsis and pseudogapping:

Dina would recommend the article, and Zaki would __ the book.

- a. What remains overt in the second clause?
- b. How does this differ structurally from typical VP ellipsis?

3. Null Subjects and Control Structures

Identify the position and interpretation of PRO:

- a. *Tari promised [PRO to revise the abstract].*
- b. *Joko tried [PRO to understand the explanation].*
 - 1) What type of control structure is used in each case (subject or object control)?
 - 2) Is there an overt subject in the embedded clause?

4. Raising vs. Control

Compare raising and control in relation to empty categories:

- a. *It seems [__ to be raining].*
- b. *Bima decided [__ to present early].*
 - 1) Determine whether the subject position of the embedded clause is filled by PRO or a trace.
 - 2) Justify your analysis based on verb properties and subject interpretability.

5. Ellipsis in Indonesian

Provide one Indonesian example of VP ellipsis (or full predicate ellipsis), and explain its interpretation.

- a. Example: *Mira bisa naik menara itu, dan Kito juga bisa __.*
- b. Discuss whether this resembles English VP ellipsis, pseudogapping, or NP deletion.

Unit 13. Applied Problem Solving in Syntax

1. Analyze the sentence: 'The drone in the sky appeared to malfunction.' Identify all projections.
2. Provide a derivational tree for: 'Those students seem to enjoy quantum riddles.'
3. Test whether 'seem' is a control or raising verb using expletive insertion.
4. Generate two ambiguous sentences and propose structural trees to disambiguate them.

Unit 14. Language Variation and Universal Grammar

1. Compare null subject properties in English and Indonesian using three sentence examples each.
2. Provide a syntactic analysis of the SVO order in Indonesian and its alignment with Universal Grammar.
3. Explain how syntactic parameters help define variation. Illustrate with WH-question placement.
4. Analyze how Indonesian syntax supports the concept of Universal Grammar.

Appendix B: Tree Drawing Guide

Tree diagrams are central to syntactic analysis because they visualize the hierarchical structure of phrases and sentences. This appendix provides a step-by-step guide to drawing phrase structure trees using X-bar theory and standard generative grammar conventions.

1. Basic Structure Types

- a. Every phrase (XP) contains a head (X), which may project to intermediate level (X') and maximal projection (XP).
- b. Complements appear within the projection of the head.

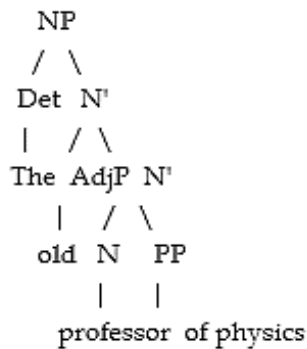
- c. Specifiers appear outside the bar-level projection.
- d. Adjuncts attach recursively to X' and may form multiple X' layers.

2. Notation and Labeling Conventions

- a. $XP \rightarrow (\text{Specifier}) X'$
- b. $X' \rightarrow X (\text{Complement})$
- c. $X' \rightarrow X' (\text{Adjunct})$

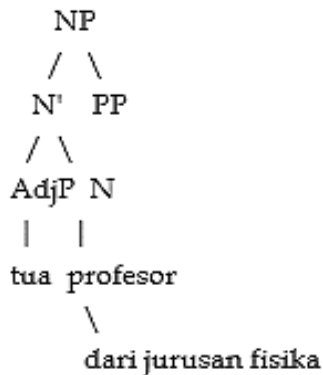
3. Example Tree: English NP

Sentence: "*The old professor of physics*"



4. Example Tree: Indonesian NP

Sentence: "*Profesor tua dari jurusan fisika*"



5. Guidelines for Tree Drawing

- a. Always begin with the sentence (S) or TP as the root.
- b. Ensure that every phrase has a head, and that complements are closely attached to the head.
- c. Represent movement (e.g., WH-movement, raising) with traces or copies if using minimalist representations.
- d. Use arrows or indices to show co-reference or binding relations where applicable.

6. Tools for Tree Drawing

- a. **Software:** TreeForm, Arboreal, Linguistic Tree Constructor, or online syntax tree generators.
- b. **Manual Drawing:** Use labeled brackets and indentation or draw Y-shaped trees with clear branch lines.

Appendix C: Master Glossary & Quick Reference

A. Master Glossary (A-Z)

1. Adjacency – linear next-to-ness; relevant for affix hopping and cliticization (see Unit.7, Unit.9)
2. Adjunct – optional modifier not selected by the head (see Unit.4)
3. Agreement – feature valuation between functional heads and nps (see Unit.2, Unit.6)
4. Anaphor – reflexive/reciprocal needing a local antecedent (see Unit.11)
5. Argument – phrase required/selected by a head (see Unit.10)
6. Aspect – temporal contour of events (progressive, perfect) (see Unit.7)
7. Binding – dependency between anaphors/pronouns and antecedents (see Unit.5, Unit.11)
8. Case (abstract) – licensing for nps (nominative/accusative) (see Unit.10)
9. C-command – structural relation underpinning binding/scope (see Unit.5)
10. Clause type – declarative, interrogative, imperative, embedded (see Unit.6)

11. Complement – phrase selected by and sister to a head (see Unit.4)
12. Complementizer – C head marking clause force (that/if/bahwa) (see Unit.6)
13. Constituency – grouping of words into units (see Unit.3)
14. Control – PRO's reference fixed by a controller (see Unit.12)
15. Coordination – joining parallel constituents (and/dan) (see Unit.3)
16. Copy/trace – record of moved element in derivations (see Unit.9)
17. Derivation – stepwise construction of a sentence (see Unit.9)
18. Do-support – insertion of do to host tense/negation (see Unit.7, Unit.9)
19. E-language – externalized utterances/texts (see Unit.1)
20. Ellipsis – omission under identity (VP-ellipsis, pseudogapping) (see Unit.12)
21. EPP – clause must project a subject position (Spec-TP) (see Unit.10)
22. Feature – grammatical property (person, number, tense, Case) (see Unit.2)
23. Functional category – grammatical heads (C, T, v, D, P) (see Unit.2, Unit.6)
24. Generativity – finite rules yield infinite sentences (see Unit.1)
25. Government (GB) – relation licensing Case/binding domains (see Unit.11)
26. Head – category projecting the phrase (N in NP, V in VP) (see Unit.4)
27. Head movement – movement of a head to a higher head ($V \rightarrow T$, $T \rightarrow C$) (see Unit.9)
28. Headedness – head-initial vs. Head-final order (see Unit.4)
29. I-language – internal mental grammar (see Unit.1)
30. Island – domain blocking movement (complex NP, wh-island) (see Unit.8)
31. Lexical category – content words (N, V, Adj, Adv) (see Unit.2)

32. Locality – constraints on how far dependencies can stretch (see Unit.5, Unit.8)
33. Minimal pair (syntax) – two sentences differing in one structural point (see Unit.13)
34. Minimalism – framework using Merge/Agree with economy conditions (see Unit.9)
35. Modal – auxiliary encoding mood/possibility/necessity (see Unit.7)
36. Movement – displacement of constituents for feature satisfaction (see Unit.8, Unit.9)
37. NP/DP – noun phrase/determiner phrase analysis (see Unit.4)
38. Null subject (pro) – unpronounced pronominal in null-subject languages (see Unit.12)
39. Parameter – limited points of cross-linguistic variation (see Unit.14)
40. Pied-piping – moving a larger phrase with a wh-word (see Unit.8)
41. PP – prepositional phrase; P takes a DP/NP complement (see Unit.3–4)
42. Predicate – what is said about the subject; often VP/TP (see Ch.6)
43. PRO – null subject of non-finite/control clause (see Unit.12)
44. Pronoun – referring expression with independent reference (see Unit.11)
45. Pseudogapping – ellipsis leaving a focused remnant (see Unit.12)
46. Raising – NP raises from embedded predicate without θ -role there (see Unit.12)
47. Scope – semantic range of operators/quantifiers (see Unit.5)
48. Selection (subcategorization) – requirements a head imposes on complements (see Unit.4)
49. Sentence (clause) – CP/TP structure housing subject and predicate (see Unit.6)
50. Specifier – phrase in Spec position of XP, often subject/DP (see Unit.4)

51. Structural ambiguity – multiple parses yield different meanings (see Unit.3)
52. Subject – external argument in Spec-TP/vp domain (see Unit.6, Unit.10)
53. Subjacency – locality condition limiting movement steps (see Unit.8)
54. Substitution test – replacing a string with a pro-form to show constituency (see Unit.3)
55. T (Tense) – functional head for tense/agreement (see Unit.6)
56. Theta (θ) role – semantic role assigned by a predicate (Agent, Theme, Goal) (see Unit.10)
57. Theta criterion – one θ -role per argument; each θ -role assigned once (see Unit.10)
58. TP – Tense Phrase; clausal spine above vp/VP (see Unit.6)
59. Transformation – earlier term for structure-changing operations (see Unit.9)
60. Tree diagram – hierarchical representation of structure (see Unit.3)
61. UG (Universal Grammar) – innate principles guiding acquisition (see Unit.1, Unit.14)
62. Unaccusative – intransitive with Theme subject; internal argument raises (see Unit.10)
63. Unergative – intransitive with Agent subject (see Unit.10)
64. Voice – active/passive alternations, sometimes morphologized (see Unit.6, Unit.10)
65. VP – verb phrase hosting predicate and complements (see Unit.3–4)
66. VP-ellipsis – deletion of VP under identity with antecedent (see Unit.12)
67. Wh-in-situ – wh-word remains in base position (e.g., Indonesian) (see Unit.8)
68. Wh-movement – wh-phrase moves to Spec-CP in English (see Unit.8)
69. X-bar/X' – schema for universal phrase structure (see Unit.4)

B. Abbreviations & Symbols

1. Category labels – N, V, Adj, Adv, P, D, T, C, v, n, a
2. Phrase labels – NP/DP, VP, PP, TP, CP, vP
3. Features – [person], [number], [gender], [Case], [tense], [aspect]
4. Null elements – Ø (unpronounced), PRO (null control subject), pro (null pronominal)
5. Judgments – * ungrammatical; ? Marginal/questionable; ?? Highly marked; ✓ acceptable
6. Notations – ‘ ’ gloss quotes; “ ” textual quotes; /.../ phonemic; [...] features; <...> copy
7. Movement marks – t (trace), ↑ movement to, → mapping, ... ellipsis
8. Operators – ∀ universal, ∃ existential, ¬ negation
9. Indonesian particles (common) – sudah (perfect-like), akan (future-like), sedang (prog-like), yang (rel/relativizer)

C. Feature Notation Cheatsheet

1. Person – 1, 2, 3
2. Number – sg, pl
3. Tense – pst (past), prs (present/non-past), fut (future-like via auxiliaries/particles)
4. Aspect – prog (progressive), perf (perfect), hab (habitual)
5. Mood – ind (indicative), imp (imperative), subj (subjunctive – rare in eng.)
6. Case – nom (nominative), acc (accusative), gen (genitive), dat (dative)
7. Agreement – t[φ:person/number] probes subject for valuation
8. Epp – t[epp:+] requires a subject in spec-tp
9. Indonesian note – verbs typically lack overt φ/tense morphology; temporal/aspectual interpretation via particles/adverbs/context

D. Glossing Conventions (Leipzig-style, adapted)

Common glosses used with Indonesian-English examples:

1. 1, 2, 3 – first/second/third person
2. Sg, pl – singular, plural
3. Poss – possessive
4. Acc, nom – accusative, nominative
5. Pst, prs – past, present/non-past (when applicable)
6. Prog – progressive
7. Perf – perfect
8. Pass – passive
9. Act – active
10. Comp – complementizer (bahwa/that)
11. Rel – relativizer (yang)
12. Neg – negation
13. Top – topic marker (when annotated)
formatting:
14. Indonesian source line
15. Word-by-word gloss line in small caps (e.g., 3.sg act-eat)
16. Free translation in quotes

Example:

1. Dia sudah makan.
2. 3.sg perf eat
3. “He/She has eaten.”

E. Tree & Movement Conventions

1. Phrase labels – use TP/CP/vp over S/INFL for uniformity
2. Head direction – show head-complement order; annotate if language-specific (Eng head-initial; many Indo phrases also head-initial with post-nominal D)
3. Specifiers – left of X'; one Spec per projection unless noted
4. Movement – draw arrows or annotate with t and indices (Who_i did she see t_i)
5. Copies – optionally show <who> in base position under Minimalist copy theory
6. Agreement/Case – annotate with feature checking (T[φ] ↔ DP[φ]; v assigns acc)

7. Ungrammaticality – mark with *; mild degradation with ?
8. Ellipsis – mark elided constituents with brackets [VP $\emptyset$] and identity pointer $\rightarrow VP_a$

F. Indonesian-English Term Map (selected)

1. Kategori leksikal – lexical category
2. Kategori fungsional – functional category
3. Frasa nomina (fn) – np/dp
4. Frasa verba (fv) – vp
5. Frasa preposisional (fp) – pp
6. Inti/kepala frasa – head of phrase
7. Penyerta/intisari/pelengkap – complement
8. Pewatas/keterangan – adjunct
9. Pengikat/c-command – c-command
10. Peran tematik – theta role
11. Perjanjian/keberjanjikan – agreement
12. Kasus (abstrak) – abstract case
13. Pengendalian – control
14. Pengangkatan – raising
15. Penghilangan/elipsis – ellipsis
16. Pemindahan/gerak – movement
17. Klausa/tataran klausa – clause level (tp/cp)
18. Penanda komplemen – complementizer (bahwa/that)
19. Kata tanya in-situ – wh-in-situ
20. Parameter – parameter
21. Tata urutan kepala – head-directionality

Appendix D: Sample Answers to Problem Sets

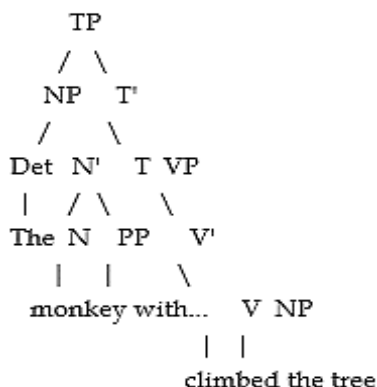
1. Problem 1: Phrase Structure Tree

Sentence: "*The monkey with the red hat climbed the tree.*"

Answer:

- a. NP: [The monkey with the red hat]
- b. PP: [with the red hat]
- c. VP: [climbed the tree]
- d. S/TP = NP (subject) + VP (predicate)

Tree Overview:



2. Problem 2: Binding Theory

Sentence: "*Lina thinks that she hurt herself.*"

- Principle A:** *herself* must be bound within its domain (satisfied).
- Principle B:** *she* must be free within its domain (satisfied).
- Antecedents:** *Lina* binds *she* and *herself* across clause boundary (depending on interpretation).

Appendix E: Further Reading

Suggested readings for students who wish to deepen their knowledge in syntax and related linguistic fields.

1. Introductory Texts

- Radford, A. (2004). *English Syntax: An Introduction*. Cambridge University Press.
- Van Gelderen, E. (2004). *An Introduction to the Grammar of English*. John Benjamins.
- Carnie, A. (2013). *Syntax: A Generative Introduction* (3rd ed.). Wiley-Blackwell.

2. Intermediate to Advanced Readings

- Adger, D. (2003). *Core Syntax: A Minimalist Approach*. Oxford University Press.
- Chomsky, N. (1995). *The Minimalist Program*. MIT Press.
- Haegeman, L. (1994). *Introduction to Government and Binding Theory* (2nd ed.). Blackwell.

- d. Jackendoff, R. (1977). *X-bar Syntax: A Study of Phrase Structure*. MIT Press.

3. Indonesian Syntax and Austronesian Studies

- a. Sneddon, J. N., Adelaar, A., Djenar, D. N., & Ewing, M. C. (2010). *Indonesian: A Comprehensive Grammar* (2nd ed.). Routledge.
- b. Musgrave, S. (2001). *Non-subject Arguments in Indonesian*. PhD Dissertation, University of Melbourne.
- c. Adelaar, A., & Himmelmann, N. (2005). *The Austronesian Languages of Asia and Madagascar*. Routledge.

4. Syntax and Language Acquisition

- a. White, L. (2003). *Second Language Acquisition and Universal Grammar*. Cambridge University Press.
- b. Pinker, S. (1994). *The Language Instinct*. Penguin Books.
- c. O'Grady, W., Archibald, J., Aronoff, M., & Rees-Miller, J. (2017). *Contemporary Linguistics: An Introduction* (7th ed.). Bedford/St. Martin's.

AUTHOR'S BIOGRAPHY

Yanuarius Yanu Dharmawan is a lecturer and researcher in linguistics with a specialization in generative syntax and its pedagogical applications. With over a decade of experience teaching English syntax at the university level, he has guided students through the complexities of syntactic theory using comparative, discovery-based, and research-informed approaches.

His work focuses on demystifying abstract grammatical principles by situating them within the linguistic realities of learners in Indonesia. He is especially passionate about comparative syntax, employing English-Indonesian parallels to make theoretical concepts tangible and contextually relevant.

This book, *English Syntax Onset*, is the culmination of years of reflective teaching and academic writing. It is designed not only to introduce students to generative grammar but also to empower them to think like syntacticians, posing hypotheses, testing structures, and drawing insights from both English and their native linguistic intuitions. By emphasizing active inquiry and structured problem-solving, the author hopes to nurture a new generation of linguists and language educators who are confident in engaging with the syntax of English and beyond.

Dini Hidayati is a lecturer in English Education at Universitas Panca Sakti Bekasi with a strong passion for language learning, academic writing, and research in the field of English as a Foreign Language (EFL). Her scholarly work explores areas such as translanguaging, pragmatics, speaking pedagogy, and the integration of digital tools in language learning. She has co-authored studies on multimodal learning, project-based instruction, and grammar innovation, with recent works investigating the impact of platforms like YouTube, TikTok, and Google Translate in classroom contexts. Beyond her academic life, Dini is also an avid writer and freediver. For her, writing and diving serve as twin acts of reflection, one allows her to dive into thoughts and ideas, while the other immerses her in the quiet depths of the ocean and self. Both passions continue to shape her perspective as an educator, researcher, and creative spirit.