

Richard McGinn
Ohio University

0.0 Introduction

This paper is concerned with the problem of interpreting ellipsis in a generative grammar without deletion rules and transformations. The theoretical framework is the Simpler Syntax Hypothesis presented in Culicover and Jackendoff (2005). The model accepts the goal of traditional generative grammar (Chomsky 1965, Chapter One), namely: to account for the native speaker's intricate and detailed *knowledge of his or her language*. However, the authors depart from what they label "mainstream generative grammar" over the question of how such knowledge arises in the individual. For Chomsky, it is *nature* (not nurture) that plays a scientifically interesting role through the agency of the LAD (language acquisition device) which is part of the human genome hard-wired in the brain. For Culicover and Jackendoff, on the other hand, child language *learning* (nurture) also plays a huge and scientifically interesting role in the acquisition of a language. In this paper, we develop the position that Culicover and Jackendoff's model is not only tenable and attractive on theoretical grounds, but is also practical and useful from the point of view of the field linguist whose goal is to contribute to the store of known facts about the diversity of the world's languages.

1.0 The Data: An Elliptical Utterance in Rejang

The divide between the two approaches can be characterized by a single utterance from Rejang which occurred in a conversation recorded in my field notes in 1974.

- (1) Coa adé majea jurusan.
not any ACT-teach subject
'not teach any subject'

The utterance is highly elliptical but perfectly interpretable in at least two very different ways according to two very different theories of generative grammar. On the one hand, following mainstream generative grammar, the utterance can and must be interpreted by constructing a deep structure representing "perfect" knowledge of a complete sentence. This is accomplished, in part, by supplying a structure derived mostly from the human genome (Universal Grammar) supplemented by certain parameterized language particulars such as lexical items and their narrowly allowable underlying patterns of arrangement. Moreover, since the genome does not license missing subjects and objects, the deep structure must provide them. Thus, underlying elliptical (1) is allegedly a full sentence including a subject *si* 'he' and an indirect object *magea ku* 'to me'.

- (2) Si coa adé m-ajea jurusan magea ku.
He not any ACT-teach subject to me.
'He doesn't teach any subjects to me.'

Any linguist, whether generative or not—and if generative, whether mainstream or not—has to know “beforehand”, as it were, that elliptical (1) implies something like (2) based on the discourse facts, including the fact that elliptical (1) was uttered in response to a question, namely (3).

- (3) Jerusan jano n-ajea ne ko, Bi?
subject what pass.teach by-him to-you, Bi

‘What subjects does he teach you, Bi?’

Paraphrasing Pike (1947), the lexical and semantic/pragmatic association linking (1), (2) and (3) constitutes a “discourse prerequisite” to syntactic analysis. And yet, amazingly, within mainstream generative grammar it is considered a “mystery” how the speaker managed to utter (1) having the interpretation (2) in mind while composing his response to the question (3). The mystery derives from the fact that the theory excludes the concept of “short-term memory” while at the same time claiming to be “mentalist”. It is not allowed to state formally that the speaker uttered (1) while holding (3) in short-term memory. In other words, mainstream generative grammar MAKES USE OF discourse information in analyzing an utterance and constructing a deep structure along the lines of (2), but systematically rules out the discourse when accounting for the speaker’s formal knowledge of his language.

But there are numerous reasons to reject this narrow focus of mainstream generative grammar. For one thing, there is a significant relationship between short-term memory and the linguistic function of choice of possible interpretations of an utterance in a given discourse context. As will be shown in detail in section § 3, the verb *-ajea* ‘teach’ can license either one, two or three arguments; therefore, an elliptical utterance containing this verb is potentially ambiguous. In fact, however, the speaker’s elliptical response (1) is anything but ambiguous. The meaning is definitely not ****HE DOESN’T TEACH**, nor even ****HE TEACHES NO SUBJECTS**—either of these possibilities is rejected on grammatical and pragmatic grounds; moreover, both imply, falsely, that the agent is not a teacher. The actual meaning of the elliptical utterance (1) is **HE TEACHES NO SUBJECTS TO ME**. The participants in the conversation know this because of the semantic structure of the question (3) that elicited the elliptical response, namely: ‘What subjects does HE teach YOU?’.

Crucially, understanding elliptical (1) requires knowledge of the seven-sentence discourse of which it is a part. The following seven-sentence discourse was recorded in 1974 at Dusun Duku Ilir, Curup, Bengkulu, Indonesia, during a conversation that lasted several hours. The participants were Abdul Gani, a farmer aged 54 (= Speaker A), Zainubi Arbi, then a 19-year-old medical student (= Speaker B) and several others, including yours truly, a foreign linguist recording the conversation on a tape recorder.

- (4) A: Abis! Jerusan jano n-ajea ne ko, Bi?
B: Coa adé m-ajea jerusan. Si yo dosen UNSRI.
Uku belo o nak Akademi Perawat. T-en-unjuk ne baso Inggeris.
Uku galak t-em-unjuk si baso Ejang o.
- A: Enough! What subject does he teach you, Bi?
B: He doesn’t teach me any subjects. He’s a university instructor.
I was in the Nursing Academy. He shows me English. I enjoy showing him the Rejang language.

The opening exclamation *Abis!* ‘Enough’ indicated to all participants that the previous topic was not to be continued¹. Abdul Gani (speaker A) then opened a new topic by asking Zainubi (speaker B) the question which we introduced above in (3). Zainubi (referred to as Bi in the opening line of the dialog) responded with five sentences beginning with the elliptical utterance introduced in (1) above.

Perhaps the most primitive semantic fact about the elliptical utterance is that it has a corrective function. Abdul Gani’s question (2) presupposes a counterfactual, and the elliptical utterance (1) corrects it. An information-theoretic analysis of elliptical (1) expresses this corrective function.

(5) Information Structure: Focus, Topic, Ground

There is no x, x a school subject, such that HE (the linguist) teaches x to ME (Zainubi).

Notice that the TOPIC and GROUND constitute the false presupposition implicit in Abdul Gani’s question (‘What subjects does he teach you, Bi?’) which the FOCUS constituent of Zainubi’s elliptical response corrects. The remaining four sentences in Zainubi’s five-sentence response complete the process of correction by filling in the factual details, namely, the relationship between Zainubi and yours truly was not one of student-teacher, but of linguist-informant.

Crucially from the perspective of the child language learner and the field linguist, there is no real “mystery” about Zainubi’s elliptical utterance (1) that requires deep structure analysis. What the child and the field linguist need to focus on instead is the fact that the adult participants in the conversation seem to have gleaned from the previous discourse that the missing subject pronoun in (1) is *si* ‘he’ and refers to the field linguist passively standing by; and that the missing indirect object is *magea ku* ‘to me’ and refers to the speaker (Zainubi) to whom the question was addressed. Such knowledge of the discourse context is essential to the analysis of the elliptical utterance. This much is acknowledged in any theory; the only question is whether the discourse context is merely “pretheoretical” as claimed by mainstream generative grammar, or included as part of the linguistic description, as claimed by the Simpler Syntax model.

2. A Field Worker’s Brief for *Simpler Syntax* (2005)

This section presents the case for the Simpler Syntax Hypothesis by arguing that it is a generative theory that best satisfies the field worker’s desiderata, namely, the following.

- (6) a. Comprehensive – encompasses Phonology, Syntax and Semantics
- b. Multi-tiered—allows for direct interaction among major components of grammar, e.g. Phonology & Semantics/Pragmatics/Discourse
- c. Bi-directional – is neutral between speaker (meaning → sound) and hearer (sound → meaning)

Crucially, bidirectionality restores the option of following the path of discovery modeled on the child language learner and mirrored by the field linguist. Culicover and Jackendoff express this virtue in more abstract terms: “A competence theory of

¹ The previous topic had touched on a taboo subject, namely, tigers. A man had been killed by a tiger just two weeks previously in a village nearby.

syntax-semantics interface is ideally bidirectional, so that it can suit the needs of both language production and language perception.” (*Simpler Syntax*, p. 151)

Simpler Syntax offers a generative model that is data-driven. But the authors do not merely present an alternative to MGG; they also offer reasons why and how MGG has gotten on the wrong track. At bottom is their claim that Chomsky’s bedrock poverty-of-stimulus argument has little empirical support. Humans possess adequate (= large) memory capacity, and powerful general learning strategies. Linguistic theory should consider the evidence available to actual (not just “ideal”) language learners, including irregular verbs and idioms, which differ little from so-called “core” phenomena.

2.1 The Model

Simpler Syntax by Peter Culicover and Ray Jackendoff (Oxford 2005:18) presents a monostratal, multi-tiered, bi-directional model of linguistic structure.

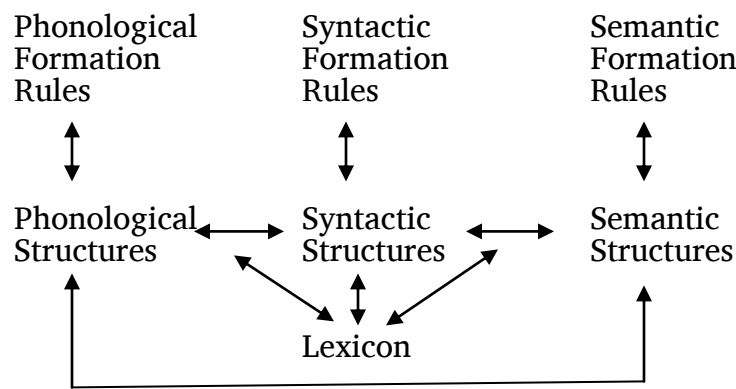


Figure 1: *Simpler Syntax* model

Unlike many modern theories, SS does not reject UG with a wave of the hand, but merely assigns to UG a much reduced—but still vital—role in the theory. For example, SS retains syntactic categories (N, NP, V, etc.) and grammatical relations (Subject, Object) as essential constituents of UG. What is most distinctive about SS in contrast to MGG is the recognition of dynamical connectivity between semantics and other components of a language, including intonation, syntax, and pragmatics as well as morphology and phonology. Formally, these connections are represented by Association Lines, a device borrowed from autosegmental phonology, as illustrated below.

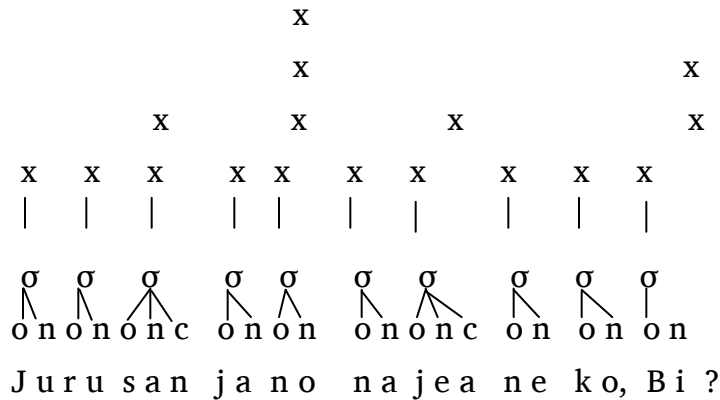


Figure 2: Metrical, Syllabic, and Segmental Tiers

Figure 2 represents formal links among conceptually distinct levels or tiers, namely: phonemes, syllables, and metrical patterns. Likewise, Figures 3-4 represent formal links among semantic, pragmatic and discourse-theoretic levels or tiers.

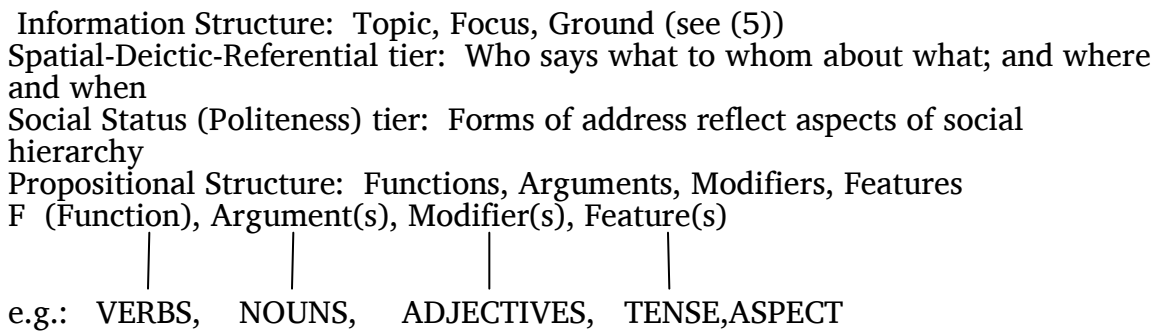


Figure 3: Some Semantic Tiers

The correspondence rules of the *Simpler Syntax* model allow direct links between Phonology (Intonation) and Semantics (Information Structure) as illustrated by the English examples below, where capitalized words indicate intonation-peak phonologically, and FOCUS semantically. (Notice there is no need to manipulate the syntax to express the relevant relationships in English.)

- (7) The BEAR chased a lion.
 The bear CHASED a lion.
 The bear chased a LION.

The data in (6) can be analyzed formally using association lines linking semi-independent tiers showing syntactic, phonological and discourse-theoretical structures, as shown below.

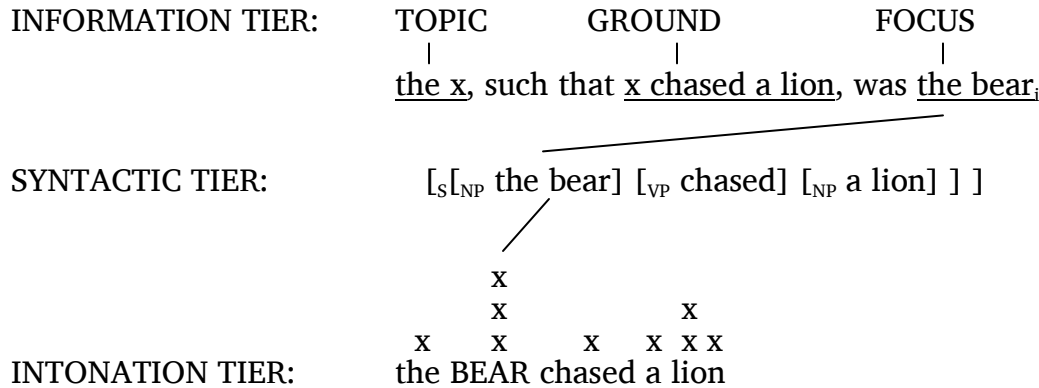


Figure 4: Syntax and Intonation

The Simpler Syntax model is “simpler” mainly because there is there is no central component to be served—neither phonology (as with the post-Bloomfieldians), nor syntax (as with MGG). Consequently, there is no need for any covert or overt syntactic manipulation to account for the phonology-semantic interface in (6) apart from simply linking a syntactic phrase (*the bear*) with (a) the FOCUS constituent of the Information tier and (b) the INTONATION PEAK of the metrical-phonological tier.

A crucial practical implication for field linguistics is that since languages are actually learned by the child (not merely dropped by the genome), then the role of field work is likewise invigorated to discover the data (and the rules) of the language, alongside the child.

Field work in mainstream generative grammar is theory-driven: the goal is to refine the theory of the “language organ” by discovering and studying *relevant* data which, by definition, can be used to test the Universal Grammar Hypothesis. By contrast, many field workers (and, presumably, children acquiring their native language) toil under far less restrictive goals, motivated by curiosity about the language and culture in question. Field workers, in particular, often seek to justify their interests in terms of a theory that is *data-driven*—meaning more balanced between facts (data) and expectations (theory) than is MGG.

2.2. Contrasting Views of the Nature of Language

The fundamental theoretical issue is neatly summarized Culicover & Nowak (2003:11) as displayed in (8).

(8)

	<u>Basic</u>	<u>Derived</u>
MGG (Chomsky)	Constituent Structure	Linear Order
Simpler Syntax	Linear Order	Constituent Structure

These contrasting visions of what language is “like” are concerned with the tension between innate vs. learned knowledge. MGG assumes that the child already “knows” (cognizes) the constituent structure of every sentence in advance of the analysis, hence the only task is to learn the word order. In other words, an elliptical sentence such as (1) over which a field worker might puzzle for years seeking the right constituent analysis, presents to the child, according to MGG, only a word order puzzle and not a constituent analysis puzzle. By contrast, the Simpler Syntax Hypothesis assumes that the child, like the field worker, is presented with data (word order) and must discover the constituent structure.

3.0 Back to the Rejang Data: Ellipsis and Discourse

To return to the theme of this paper, Culicover and Jackendoff claim that MGG's deep structure/derivational approach to ellipsis shows not just a weakness to be corrected but a fundamental flaw in the enterprise. The basic argument appeals to Occam's Razor ("Do not multiply theoretical entities without necessity.") After all, what could be simpler than a theory that says, in effect, pay attention to the preceding sentence (in this case, a question) to find the particular source of "knowledge of language" needed to interpret *this* sentence (namely, the answer to the question). The theoretical mechanism the authors introduce has a fancy name: Indirect Licensing. According to them:

- (9) "Indirect Licensing is the grammatically primitive process of interpreting fragmentary utterances in terms of the pragmatics of the discourse and nonlinguistic context."
--*Simpler Syntax*, p. 539

Upon this stone is built the theoretical edifice of *Simpler Syntax*. A second, and more familiar, kind of licensing is called Direct Licensing. The following contrasts the two kinds of licensing within the model, with reference to the elliptical utterance (1).

- (10) Direct Licensing: (a) lexical-conceptual structure of verb *ajea* 'teach', (b) double-object construction, (c) active morphology, (d) fixed expression *coa adé si* 'not ...any (NP); none at all'.

Indirect Licensing: (a) Referential content of "missing" pronouns in the fragment inferred from the discourse, i.e. from the question that elicited the fragment as a response, (b) corrective function of the elliptical utterance (1) effectuated by use of fixed expression + ellipsis.

3.1 Licensing of Abdul Gani's Passive Wh- Question

This section will elucidate the structure of the stimulus question (3) asked by Abdul Gani which elicited Zainubi's elliptical response.

- (11) ABDUL GANI'S QUESTION

Jerusan jano n-ajea ne ko, Bi?
school subject what PASS-teach by-him to-you (name)

'What subject does he teach you, Bi?'

The most basic licenser is lexical. For example, in (11) the verb *ajea* 'teach' licenses the links at the interface of Phonology, Syntax and Conceptual Structure, as shown in Figure 5.

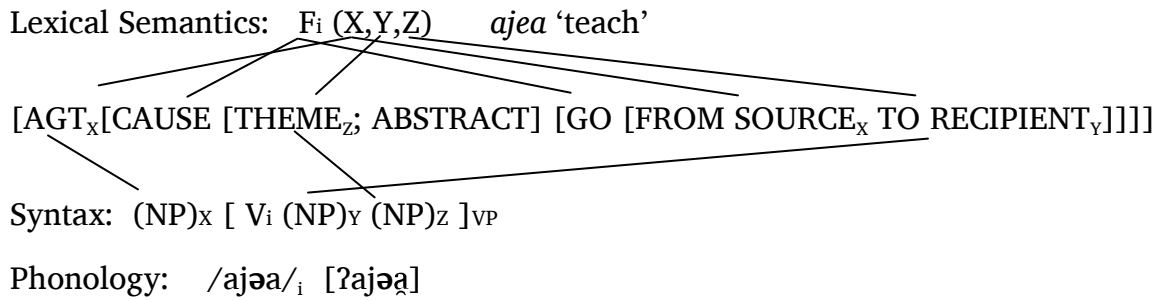


Figure 5: Lexical Licensing Based on Verb *ajea* 'teach'

Figures 5 and 6 analyze Abdul Gani's question in terms of its syntax, grammatical function structure (GF), and conceptual structure (CS—which includes thematic structure (θ-structure)).

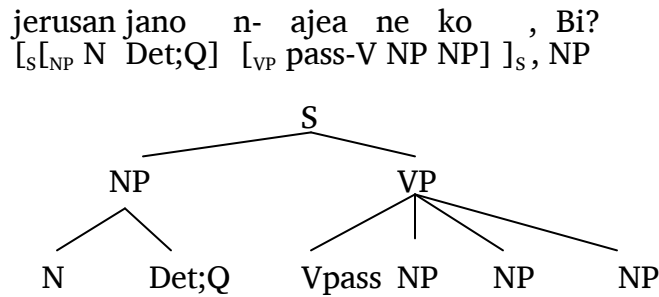


Figure 6: Syntactic Tier

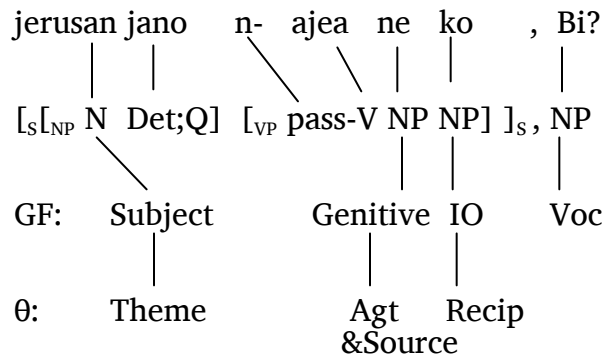


Figure 7: Thematic and GF Tiers

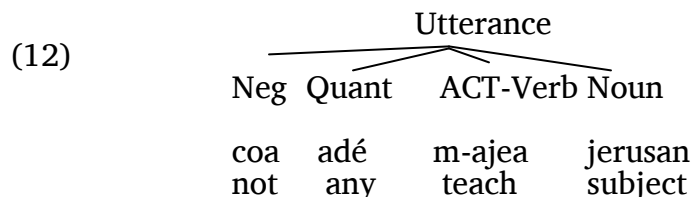
Like most Western Austronesian languages, Rejang has a robust passive that directly licenses a Wh- constituent as subject in clause-initial position (McGinn 1983).

3.2 Choice of Pronouns

Finally, the pronouns occurring in Abdul Gani’s question bear not only grammatical function and thematic role content as determined by lexical (direct) licensing, but also lexico-semantic features of Person, Number, Case and Social Status, which are licensed indirectly by the discourse context, and in particular, by the Spatial-Deictic-Referential tier of semantic structure. Thus the agentive pronoun *ne* ‘by-him’ refers to the linguist standing passively by, and the pronoun *ko* ‘you’ refers to Zainubi. Given these referents, both pronouns are masculine in Gender (a covert category in Rejang) and singular in Number. As for case features, *ne* (Genitive) can be either genitive or agentive—here the passive syntax determines the latter because *ne* represents the passive agent; likewise, *ko* ‘you’ can serve as either subject, direct object or indirect object—here again the passive syntax and the position of *ko* determines that *ko* is the Indirect Object.²

4. Toward a Formal Account of Zainubi’s Elliptical Response

Whereas Abdul Gani’s passive question (11) = (3) has a rich syntactic structure illustrated in Figure 6, Zainubi’s elliptical response contains hardly any syntactic structure at all. To recognize this fact constitutes an important feature of the Simpler Syntax Hypothesis. The flat sparseness of (12) below contrasts sharply with the rich semantic structures that license it.



² There is further semantic content that must be included in a full description, for example, the age-dependent social-status implications associated with Abdul Gani’s choice of the familial pronoun *ko* instead of honorific *kumu* to address the student Zainubi in 1974. If this discourse took place today, a Rejang speaker would likely choose honorific *kumu* instead of *ko* when addressing Dr. Zainubi.

Syntactically, (12) is little more than a string of words. By contrast, as we have seen, whereas the bulk of Abdul Gani's passive question (3) is licensed directly by lexical items, linear order and passive morphology, Zainubi's elliptical response is licensed by an interesting mix of direct and indirect licensers. Consider the licensers outlined in (13).

(13)

Direct licensers: (a) semantic and syntactic properties of lexical items, including the three-argument verb *ajea*, (b) active morphology, (c) word order.

Indirect licensers: (a) corrective function of *coa adé si* + ellipsis; (b) pro-drop; (c) association lines linking semantic (referential) features referential) from Abdul Gani's question to Zainubi's elliptical response (e.g. *ne* 'by-him' in the question and *si* 'he' implicit in the elliptical response refer to the same person), and (d) shifting of deictic features appropriate to the shift of participant role. Thus the discourse-participant Zainubi undergoes a shift from *hearer* of the question to *speaker* of the elliptical response, with the linguistic effect such that *ko* 'you' in the question and the implicit argument *ku* '(to) me' (as inferred from the semantics of elliptical response) refer to the same person.

We begin with the simpler and grammatically more "primitive" process, Indirect Licensing (IL), which plays a crucial role in accounting for the referential content of two missing pronouns in the fragment. The following quotation is worth repeating for clarity of exposition.

(9) "Indirect Licensing is the grammatically primitive process of interpreting fragmentary utterances in terms of the pragmatics of the discourse and nonlinguistic context."

--Simpler Syntax, p. 539

4.1 Corrective Use of Ellipsis

As noted at the outset of this paper (see (5) above), Zainubi's elliptical response (1) functions in part to correct a false presupposition entailed by Abdul Gani's question (3). This corrective function is determined by the concurrence of (a) the lexicalized expression (idiom) *coa adé si* meaning 'there is no NP'; 'it doesn't exist' and (b) the ellipsis in which both the subject and the indirect object are elided, but not the direct object *jerusan* 'field of study; school-subject'.

To see how *coa adé si* + *ellipsis* manages to convey the corrective function, it is helpful to introduce an analogy based on an interesting use of the English conjunction *but*. As pointed out by David Bell (personal communication), English *but* has a corrective function when used together with subject ellipsis, whereas *but* without the ellipsis lacks this corrective function. Consider the following contrast.

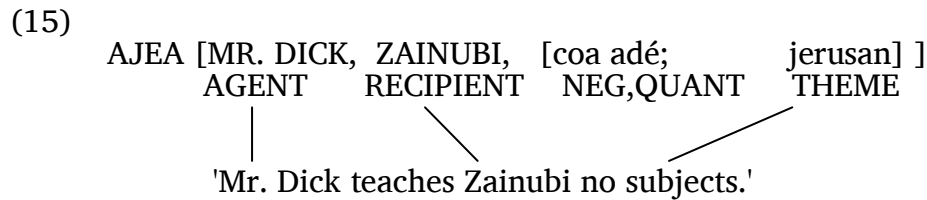
(14) Corrective vs. non-corrective use of *but*:

- a) He's not an economist but a businessman.
(not + *but* + elided subject: 'a businessman is not qualified')
- b) He's not an economist, but he is a businessman.

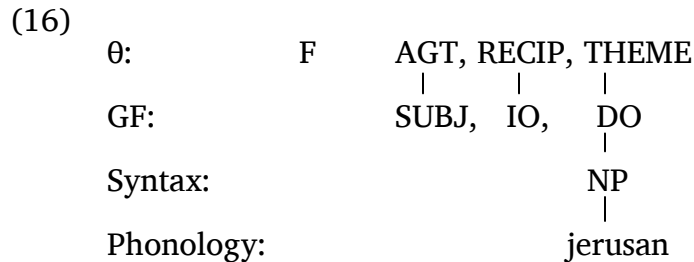
(not + *but* + full subject: ‘a businessman will do just as well as an economist’)

Such a corrective function assigned to an elliptical utterance certainly involves semantic and pragmatic factors which cannot be captured by a deep-structure analysis. And yet, as these examples show, in both Rejang and English, an ellipsis can positively contribute to the meaning of an utterance³. Such meanings can and should be accounted for as part of what the speaker-hearer knows about his language; what a child has to learn about his language; and what a field linguist routinely wants to describe in relation to the form and meaning of utterances in the language.

Continuing along these lines, perhaps the most important syntactic fact about the elliptical utterance is that there is almost no syntactic structure to describe. Almost everything of interest is in the semantics and pragmatics. Consider the richness of the elliptical utterance’s Conceptual Structure (CS) licensed by the verb *ajea* ‘teach’ in contrast to the sparseness of its syntactic structure. This point is illustrated in (15) and (16).



In Zainubi’s elliptical utterance, only one NP is realized phonologically *and* syntactically.



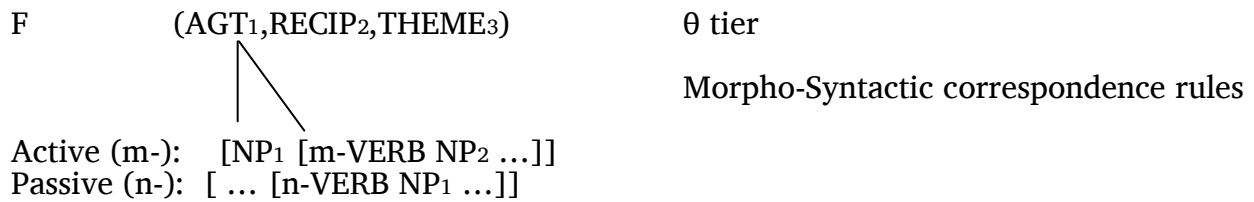
To account for all the information contained in the elliptical utterance, it is necessary to consider further direct licensing “tools” provided by the grammar of Rejang.

4.2 Direct Licensing by Active morphology

GF- and θ-correspondence rules license the Active prefix *m-* in the elliptical response. (The Active prefix *m-* contrasts with passive *n-* which occurred in Abdul Gani’s question).

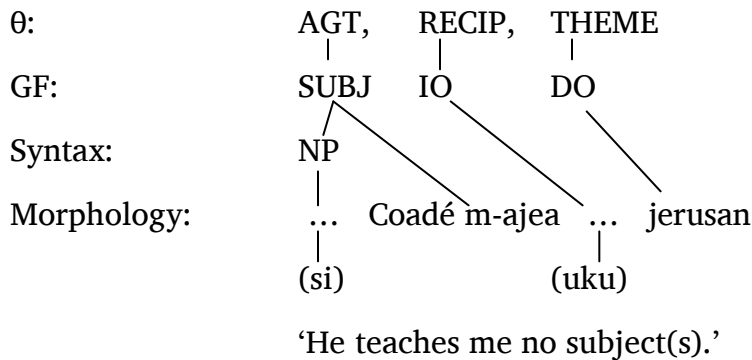
³ cf. Hendriks 2004:1 states that “In a number of cases, ellipsis decreases rather than increases the number of readings of a sentence.”

(17)



Partial interfaces of morphology, syntax, and semantics are displayed below.

(18)

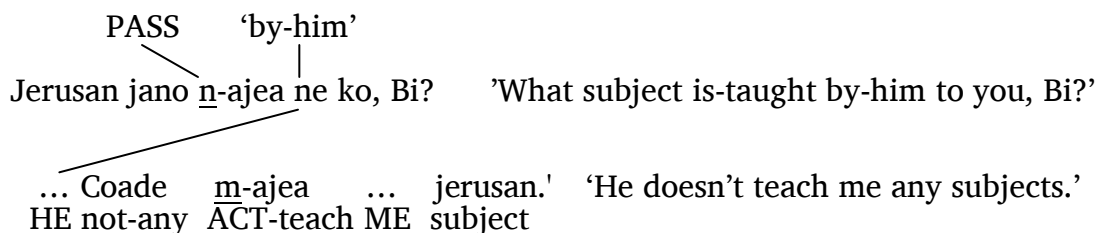


But the missing elements are more than grammatical functions and θ -roles licensed by active *m-ajea*. They also have semantic features not licensed by *m-ajea*. Semantically, they carry Person, Number, Gender, and (Social) Status features. Such features, viewed as knowledge possessed by the adult native speakers participating in the conversation, raise interesting theoretical questions such as: How do they know this? And how do children acquire such knowledge?

4.3 Implicit Arguments

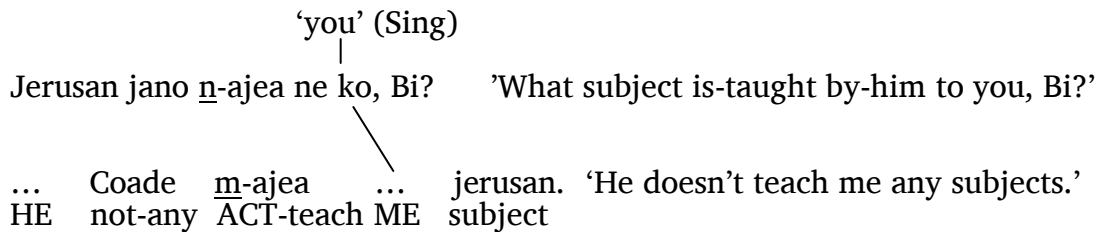
Romance languages like Latin and Spanish allow subject pronouns to be “dropped”. Many Asian languages like Tagalog, Korean, and Rejang allow two arguments to be “dropped” in the same sentence. Here we are concerned to interpret the semantic content of the empty arguments licensed by the verb *ajea* ‘teach’. This is accomplished through Indirect Licensing. As shown by (19), there is an indirect association, or link, between the Agent role in Abdul Gani’s question and Zainubi’s elliptical response which involves deictic and grammatical shifting in the categories of Voice and GF/Case (whereas Gender and Social Status features carry over from the question to the response).

(19)



As shown by (20), the link between the IO/Recipient roles involves a shift in Person (2nd to 1st), whereas Gender and Status features carry over from the question to the response.

(20)



5.0 Conclusion

As illustrated by (12), syntax in *Simpler Syntax* is, well, simpler than in MGG. This fact is most apparent in the description of elliptical sentences. In MGG, licensing of an elliptical utterance like (1) = (12) means constructing a deep structure and deriving the data by transformational rules. By contrast, in the *Simpler Syntax* model, licensing (1) = (12) entails linking implicit arguments to their proper antecedents in the discourse, and accounting for all linguistic-semantic and pragmatic factors, such as the corrective function. Syntax plays a minor role. (And Phonology plays no role at all.)

In conclusion, let us list four advantages of the *Simpler Syntax* approach from the point of view of the field worker interested in the diversity of the world’s languages.

- a. Simplicity – If arguments are present in Semantic structure, why repeat them as ‘empty pronouns’ in the syntax, only to delete them later in the phonology? The burden of proof is to justify carrying empty categories in grammar.
- b. Field Work – the *Simpler Syntax* theory requires the field linguist to go where he wants to go—into Semantics, Discourse, and Culture, in order to license sentence fragments directly via discourse antecedents.
- c. Constraints vs. Transformations – The Constraints-based approach of *Simpler Syntax* is designed to associate elements from different levels (tiers) of language (and the mind), e.g. syntax and intonation (Figure 4). Transformations cannot do this.
- d. Linguistics and Related Fields of Inquiry – the *Simpler Syntax* approach mirrors child language experience, adult language processing, the likely path of the history of the language and, ultimately, the evolution of language, by appealing to “primitive” sources of knowledge of language, such as discourse context. Derivational theories rooted in a “central” and richly elaborated syntactic component, such as MGG, tend to isolate linguistics from other fields, and even attempt to make a virtue out of this tendency.

References

- Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge: MIT Press.
- _____. 1986. *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- _____. 1995. *The minimalist program*. Cambridge: MIT Press.
- Culicover, Peter and Ray Jackendoff. 2005. *Simpler syntax*. Oxford University Press.
- Culicover, Peter and Andrzej. 2003. *Dynamical grammar: Minimalism, acquisition, and change*. Oxford University Press.
- Hendriks, Petra. 2004. Coherence relations, ellipsis and contrastive topics. *Journal of Semantics* 21:133-153
- McGinn, Richard. 1983. *Outline of Rejang syntax*. Jakarta: NUSA.
- _____. 1991 "Pronouns, politeness and hierarchy in Malay." In Robert Blust (ed.), *Currents in Pacific Linguistics: Festschrift in honor of George W. Grace*. Canberra, Australian National University: *Pacific Linguistics* C-117, pp. 197-221.
- _____. 1998. "Anti-ECP effects in the Rejang language of Sumatra. *Canadian Journal of Linguistics* 43(3/4): 359-376.
- Pike, Kenneth L. 1947. Grammatical prerequisites to phonemic analysis. *Word* 3: 155-72.