

We ‘Hear’ the Structure Our Mind Creates

Hisao Tokizaki

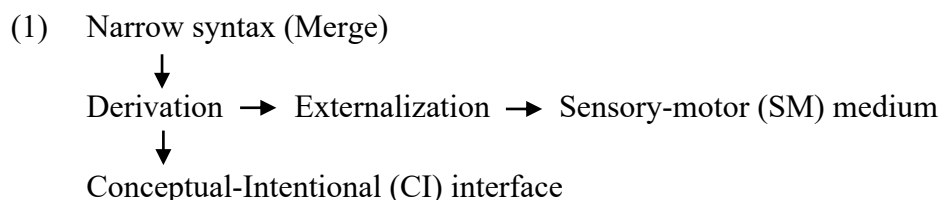
Sapporo University

ABSTRACT. Chomsky (2021) argues that children ignore what they hear (linear order) and rely on the structure they never hear. This paper claims that an utterance contains prosodic features such as pause, pitch and stress, which the hearer uses in interpreting the hierarchical structure the speaker intends. In this sense, we ‘hear’ the structure our mind creates. I point out the fact that a pause and a boundary tone may well appear at the end of a phrase and tell the hearer that there is a constituent break there. It is also argued that phrasal stress, which is assigned to the bottom of a structure, helps the hearer to construct the hierarchical structure the speaker intends. The syntax-phonology interface is explained in terms of a mapping rule and Set Stress.*

Keywords: externalization, prosody, linearization, pause, stress

1. Externalization in the architecture of grammar

In the minimalist program of linguistic theory, it has been argued that narrow syntax derives expressions, which are externalized to the sensori-motor (SM) medium (Chomsky 2021 among others), as shown in (1).



To illustrate this architecture, let us consider some examples. If we have two syntactic objects X and Y (words or phrases) in the Workspace, we can merge them together to make a set {X, Y}, which has no linear order. This set is externalized to the sensori-motor system in the form of a sequence X Y or Y X.

* This article is based on my presentation at the 13th workshop on the Phonological Externalization of Morphosyntactic Structure: Theory, Typology and History held online, February 12, 2023. I would like to thank the participants of the workshop. This work was supported by JSPS KAKENHI (Grant Number 20H01269).

- (2) a. X, Y
 b. {X, Y} Merge
 c. X Y or Y X Externalization = Linearization

If there is another syntactic object Z in the Workspace, Z can merge with the set {X, Y} to make a larger set {{X, Y} Z}, which has no linear order in the syntax. This set is externalized in one of the orders shown in (3c).

- (3) a. X, Y, Z
 b. {{X, Y}, Z} Merge
 c. X Y Z, Y X Z, Z X Y or Z Y X Externalization = Linearization

In this standard view of externalization shown in (2) and (3), the Externalization of a hierarchical structure amounts to the linearization of syntactic objects.

However, I point out that Externalization contains more information than the linear order of syntactic objects shown in (2c) and (3c). Specifically, I argue that the speaker encodes the hierarchical structure in (3b) with prosody including pause, pitch and stress as shown in (4c), where stressed objects are in bold.¹

- (4) a. X, Y, Z
 b. {{X, Y}, Z} Merge
 c. **X** Y Z, Y **X** Z, Z **X** Y or Z Y **X** Externalization = Linearization + Prosody

Here, adjacent syntactic objects can be separated by a pause. I will discuss the role of pause, pitch and stress in encoding/decoding hierarchical structure in section 3.

2. Agreement and structure

Chomsky (2021: 8) argues that agreement is based not on the linear order but on the structure. He points out that the sentence in (5) is “mastered by children as young as 30 months of age or less as experiment has shown:”²

¹ Generally, the order YX Z in (3c) is not acceptable because it represents the structure [_{NP} [_{NP} Y X] Z], which violates the Final-Over-Final Constraint (Biberauer et al. 2014).

² Chomsky put a footnote here: “According to some experimental work, down to 18 months. Shi et al. (2020).” In fact, Shi et al. (2020) use the French sentences in (i) for their experiments.

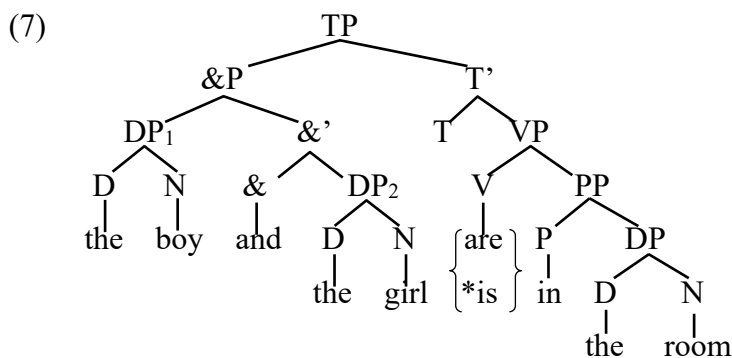
- (i) a. [La *banane*_{FEM} dans le *chapeau*_{MAS}]_{FEM}, *elle*_{FEM}/**il*_{MAS} V ...
 the banana in the hat it

(5) the boy and the girl are/*is in the room

He argues that the child does not use adjacency but relies on the structure, as shown in the citation in (6) (underscore HT).

(6) To determine agreement, the child does not use the simplest computational rule, adjacency. Instead, the child reflexively relies on something it never hears: the structure its mind creates. The child then assigns plurality by virtue of the nature of this abstract structure. (Chomsky 2021: 8)

The structure of the sentence (5) can be represented as in (7).



In (7), the verb should agree with &P *the boy and the girl*, not with DP₂ *the girl*, hence the verb form should be *are* rather than *is*. I basically agree with the idea that this example shows that the child relies on the structure, not on the linear adjacency.

However, some questions arise. Does the child only hear the words linearized in a sequence? Aren't there any prosodic cues in (7) that makes the verb agree with the &P rather than DP₂? Can we say that DP₂ is really 'adjacent' to the verb even if they are separated by a pause? Below I argue that we 'hear' the prosody that the speaker uses in the Externalization of syntactic structure.

3. The structure encoded in prosody

3.1 Pause duration

First, we 'hear' silence between words and phrases. When we utter the sentence in (5) above, we are likely to put a pause of some length between the subject *the boy and the girl* and

b. [La banana_{FEM} et le chapeau_{MAS}]_{MAS}, ils_{MAS}/*elles_{FEM} V ...
 the banana and the hat they

the verb *are*, as shown in (7) where x shows a certain length of pause (cf. Selkirk's (1984) silent demibeat).

(7) The boy and the girl x are in the room.

In an interview in 2023, Chomsky talked about the subject-verb agreement using an example similar to (7).³

(8) John and Bill are/*is here.

His utterance of this sentence in the interview is analyzed as Figure 1 with the Praat software.⁴

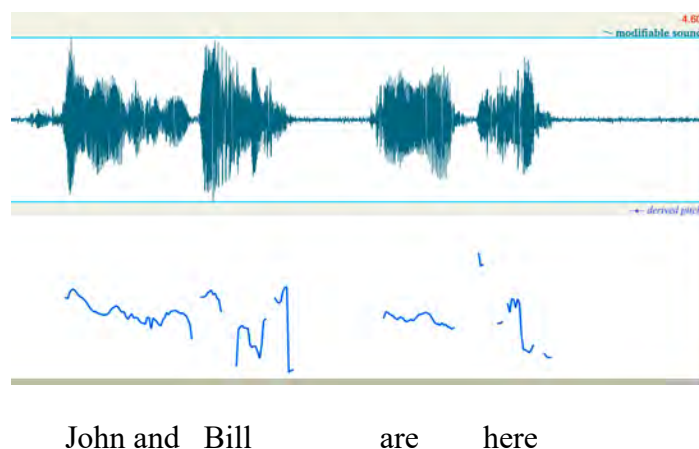


FIGURE 1

Here *Bill* and *are* are separated by a long pause, which might be due to some emphasis on the verb form. However, if we compare a complex subject with a simple subject, the pause between the subject and the verb is likely to be longer in a sentence with a complex subject than in a sentence with a simple subject, as shown in (9).

- (9) a. The boy and the girl x are in the room. (= (7))
 b. The girl is in the room.

Cooper and Paccia-Cooper (1980) observe that the more complex the syntactic constituent, the more likely a pause is to appear. Then, the hearer 'hears' the silence between the subject and the verb in (9a) but not in (9b).

³ Noam Chomsky on Theories of Linguistics (Part 2) Closer To Truth Chats (4:13)

<https://youtu.be/E4KhK3kktcM?t=264>

⁴ <https://www.fon.hum.uva.nl/praat/>

How can we explain the effect of syntactic complexity on pause? In Tokizaki (1999, 2008), I propose a syntax-phonology mapping rule shown in (10).⁵

(10) Interpret a syntactic bracket [or] as a phonological boundary /.

This rule maps the syntactic structure of the sentences in (7) onto the prosodic structure as shown in (11) and (12), where the boundaries between the subject and the verb are underscored.⁶

(11) a. [TP [&P [DP the boy] [&' and [DP the girl]]][VP are [PP in [DP the room]]]

b. /// the boy // and / the girl /// are / in / the room ///

(12) a. [TP [DP the girl][VP is [PP in [DP the room]]]

b. // the girl // is / in / the room ///

Between the subject and the verb, there are four prosodic boundaries in (11b) and two boundaries in (12b). Thus, we can correctly predict that the pause between subject and the verb is longer in (9a) (= (11)) than in (9b) (= (12)).

In this sense, *the girl* and the verb are not adjacent as claimed by Chomsky in (5), but are separated by a pause, which I claim is a phonetic realization of the phonological boundaries. In other words, the long pause between *the girl* and *are* in (9a) (= (11)) tells the hearer that *the girl* and the following VP are not sister constituents in the sentence. The pause helps the hearer to find that the intended subject is *the boy and the girl* rather than *the girl*.

This effect of pause is also seen in the high/low attachment of relative clauses in psycholinguistics. The sentence in (13) is ambiguous in that the relative clause modifies the high NP *the servant of the actress* or the low NP *the actress*.

(13) Someone shot the servant of the actress who was on the balcony.

It has been pointed out that this sentence can be disambiguated by placement of pause (#) as shown in (14) (cf. Schafer 1997, Fodor 2002, Jun 2003, Hwang et al. 2011).

(14) a. Someone shot the servant # of the actress who was on the balcony.

b. Someone shot the servant of the actress # who was on the balcony.

⁵ In Tokizaki and Okazaki (2022: 28-29), I discuss how to reinterpret the mapping rule (10) as an Externalization rule in the minimalist program.

⁶ In (11a) and (12a), I omitted the head T and its intermediate projection T' between *girl* and *are/is* because T does not contain any phonetic features (cf. Tokizaki 1999, 2008).

The listener is likely to interpret (14a) as the low attachment and (14b) as the high attachment. I argue that this preference of interpretation is due to the difference in syntactic boundaries as shown in (15), where I omit the boundaries irrelevant to the discussion.

- (15) a. Someone shot the servant of [the [actress [who [was [on [the balcony]]]]]]
 b. Someone shot [[the [servant [of [the actress]]] [who [was [on [the balcony]]]]]]

The mapping rule (10) applies to the low attachment (15a) and the high attachment (15b) and gives (16a) and (16b), respectively.

- (16) a. Someone shot the servant of / the / actress who / was / on / the balcony // // // //
 b. Someone shot // the / servant / of / the actress who / was / on / the balcony // // // //

Between *actress* and *who*, there is only one prosodic boundary in the low attachment (16a) and four prosodic boundaries in the high attachment (16b). The listener is likely to interpret the pause between *actress* and *who* in (14b) as a sequence of boundaries in (16b) rather than a boundary in (16a). Then, we can ‘hear’ the structure in this case as well as in the subject-verb agreement case in (7) and (8).

3.2 Pitch

In addition to pause, pitch signals the structure the speaker creates in their mind. Let us look at the pitch contour of the utterance in (8).

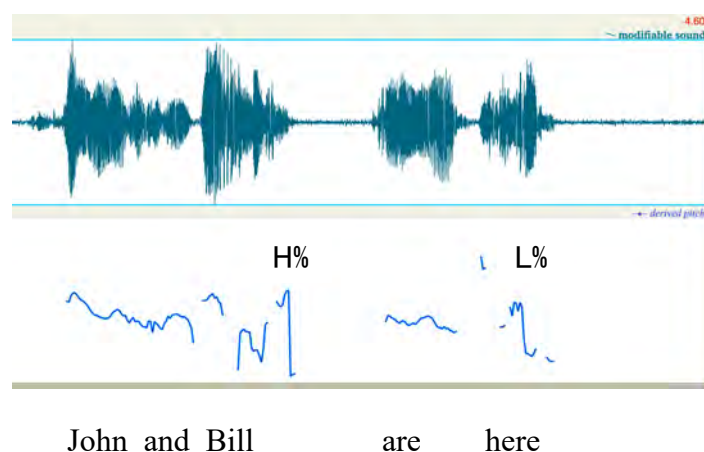


FIGURE 2

The pitch contour shows that there is a H% boundary tone at the end of the subject NP and a L% boundary tone at the end of the sentence. If we assume that a boundary tone appears at the end of an intonational phrase (IntP), we can represent the intonational phrasing in (8) as (17).

(17) (_{IntP} John and Bill) (_{IntP} are here)

The hearer hears the boundary tones as well as pause in (8) to construct the prosodic structure in (17). This prosodic structure helps the hearer to build the syntactic structure (18a) rather than (18b).

- (18) a. [[John and Bill] [are here]]
 b. [John and [Bill are here]]

If we apply the mapping rule (10) to (18a) and (18b), we get the phonological representations in (19a) and (19b), respectively.

- (19) a. // John and Bill // are here //
 b. / John and / Bill are here //

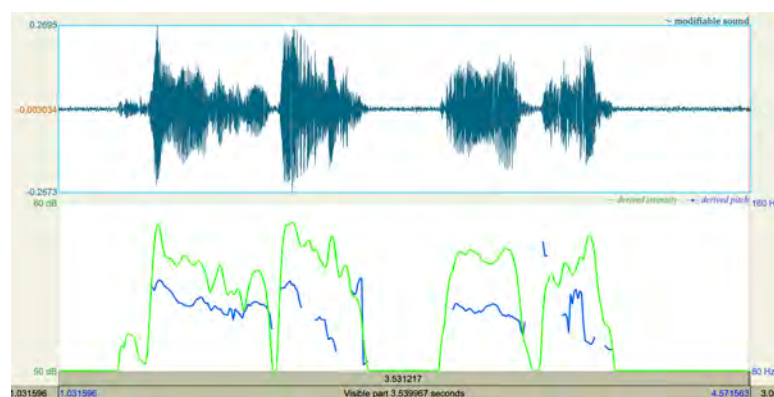
The intonational phrasing in (17) matches (19a) rather than (19b) because there are two boundaries between *Bill* and *are* in (19a). Thus, the hearer hears the intonational phrasing in (17) and recognizes that the structure the speaker intended is (18a) rather than (18b). In this sense, the hearer 'hears' the structure through the syntax-phonology mapping mechanisms.

3.3 Stress

Next, let us consider the function of stress in the interpretation of structure. The sentence in (8) is most likely to be uttered with stress on *Bill* and *here*.

(20) John and **Bill** are **here**

This stress placement can be seen in an analysis of the utterance (8) shown in Figure 3.



John and Bill are here

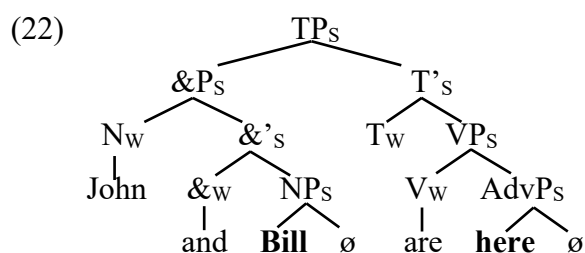
FIGURE 3

Here we can tell high intensity on *Bill* and *here* from the wave form in the upper part and the green line in the lower part.⁷

It has been argued that stress is assigned to the bottom of a syntactic tree diagram (Cinque 1993, cf. Szendrői 2001, Reinhart 2006). In order to formalize this idea in the minimalist grammar, Tokizaki (2018, 2020a, b) proposes a rule Set Stress (21).

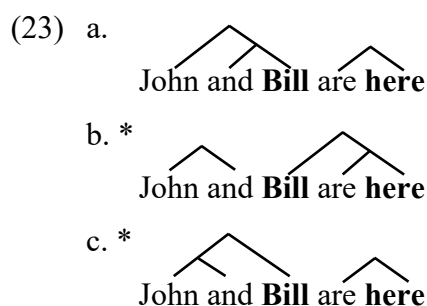
(21) Assign Strong to a set and Weak to a terminal.

This rule applies to the sentence in (20) to derive a representation in (22).⁸



Here, I assume that the words *John*, *and*, and *are* are non-branching terminals while $\{Bill, \emptyset\}$ and $\{here, \emptyset\}$ are sets consisting of a word and an unpronounced modifier (\emptyset) (see Tokizaki (2018, 2020a, b) for the problem of the first Merge). *Bill* and *here* receive the main stress in the subject and in the predicate because they are dominated only by S. Thus, we can explain the connection between stress and structure in terms of the rule (21).

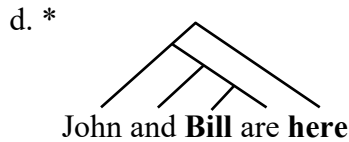
The hearer can use the stress location for parsing the sentence structure. In (20), the stress on *Bill* and *here* helps the hearer to choose the structure intended by the speaker (23a) rather than the other structures, some of which are shown in (23b, c, d).⁹



⁷ In Figure 3, *are* also has some intensity, which is due to emphasis on the verb form.

⁸ In (22), both $\&P_s$ and T'_s are assigned S. This would cause stress clash between these constituents. In Tokizaki (2018), I argue that this problem does not occur if we assume Transfer of VP to the phonological component.

⁹ In (23), I omit zeros (\emptyset) after *Bill* and *here* in (20) for the sake of simplicity.



The stress is correctly assigned to the bottom of a structure in (23a), but not in (23b, c, d). In (23a), the stressed words *Bill* and *here* are at the bottom of the structure containing them. In (23b), neither *John* nor *and* receives stress, and *Bill* is assigned a stress even though *Bill* is not at the bottom of the structure dominating *Bill are here*. In (23c), *Bill* is assigned a stress even though it is not at the bottom of the structure dominating *John and Bill*. In (23d), a stress should fall on *are* rather than *Bill*, and *here* should not receive stress because *here* is not at the bottom of the whole structure. Thus, the hearer can correctly build the structure the speaker intended with the help of stress as well as pause and pitch.

4. Conclusion

As Chomsky (2021) claims, it is an amazing fact that we rely on structure rather than linear order. However, utterances are not just a linear order of meaningful elements; they also contain prosody such as pause, pitch and stress, into which we encode information about the structure according to the rules of syntax-phonology interface, such as boundary mapping (10) and Set Stress (21). The hearer hears the prosody of utterances and uses it in parsing the sentence structure. In this sense, we (including children) 'hear' the structure our mind creates.

Of course, I am not arguing against Chomsky's claim that we rely on the structure our mind creates. The point is that structure is encoded in prosody, which we can 'hear' and use in learning and using languages.

References

- Biberauer, Theresa, Anders Holmberg and Ian Roberts. 2014. A syntactic universal and its consequences. *Linguistic Inquiry* 45, 169–225.
- Chomsky, Noam. 2021. Minimalism: Where are we now, and where can we hope to go. *Gengo Kenkyu* 160, 1–41. doi: 10.11435/gengo.160.0_1
- Cinque, Guglielmo. 1993. A null theory of phrase and compound stress. *Linguistic Inquiry* 24, 239–298.
- Cooper, William E. and Jeanne Paccia-Cooper. 1980. *Syntax and speech*. Cambridge, MA: Harvard University Press.
- Fodor, Janet D. 2002. Prosodic disambiguation in silent reading. *Proceedings of the North East Linguistic Society* 32, 113–132. Amherst, MA: GLSA Publications.

- Hwang, Hyekyung, Moti Lieberman, Heather Goad and Lydia White. 2011. Syntactic ambiguity resolution: Effects of prosodic breaks and prosodic length. *Proceedings of the 28th West Coast Conference on Formal Linguistics*, ed. Mary Byram Washburn et al., 267–274. Somerville, MA: Cascadilla Proceedings Project.
- Jun, Sun-Ah. 2003. Prosodic phrasing and attachment preferences. *Journal of Psycholinguistic Research* 32, 219–249.
- Reinhart, Tanya. 2006. *Interface strategies: Optimal and costly computations*. Cambridge, MA: MIT Press.
- Schafer, Amy. 1997. Prosodic parsing: The role of prosody in sentence comprehension. Doctoral dissertation, University of Massachusetts, Amherst.
- Selkirk, Elisabeth O. 1984. *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
- Shi, Rushen, Camille Legrand, and Anna Brandenberger. 2020. Toddlers track hierarchical structure dependence. *Language Acquisition* 27, 397–409.
- Szendrői, Kriszta. 2001. Focus and the syntax-phonology interface. Doctoral dissertation, University College London.
- Tokizaki, Hisao. 1999. Prosodic Phrasing and Bare Phrase Structure. *Proceedings of the North East Linguistic Society* 29, vol. 1, 381–395. Amherst, MA: GLSA Publications.
- Tokizaki, Hisao. 2008. *Syntactic structure and silence: A minimalist theory of syntax-phonology interface*. Tokyo: Hituzi Syobo.
- Tokizaki, Hisao. 2018. Obligatory Contour Principle and minimalist syntax. *Phonological Externalization*, volume 3. ed. by Hisao Tokizaki, 73–83. Sapporo University. <http://id.nii.ac.jp/1067/00007527/>
- Tokizaki, Hisao. 2020a. Recursive strong assignment from phonology to syntax. In Kuniya Nasukawa (ed.), *Morpheme-internal recursion in phonology*. 369–392. Berlin: Mouton de Gruyter.
- Tokizaki, Hisao. 2020b. Is the mapping to PHON complex? *Phonological Externalization*, volume 5. ed. by Hisao Tokizaki, 25–39. Sapporo University. <http://id.nii.ac.jp/1067/00007712/>
- Tokizaki, Hisao. 2022. Parameters in phonology and morphosyntax. *Phonological Studies* 25, 133–142. Tokyo: Kaitakusha.
- Tokizaki, Hisao and Masao Okazaki. 2022. *On-inron-to ta-no bumon-tono intaafaisu* [Phonology and its interfaces with other components], Tokyo: Kaitakusha.