# 言語単位とコミュニケーション能力
# Language Units and Communicative Competence

カンジェミ　フランチェスコ CANGEMI, Francesco
● 東京大学グローバル教育センター特任講師，国際基督教大学教育研究機構研究員
Project Assistant Professor, Center for Global Education, The University of Tokyo
Research Fellow, Institution of Educational Research and Service, International Christian University

**Keywords** 発話単位，文法，言語学の歴史，言語能力の評価，言語教育
Speech units, grammar, history of linguistics, assessment of competence, language education

## ABSTRACT

1950年代以降，言語能力はより小さな単位をより大きな構造に正しく組み合わせるために必要な理想化された知識として伝統的に理解されてきた（§1）。したがって，言語能力モデルは言語単位を定義しなければならないが，会話音声に適用する場合は特に厄介となる（§2）。発話単位は普遍的な根拠に基づいて動機づけられるが，発話単位化のいくつかの側面は言語固有であり，言語能力と密接に関係する（§3）。本論文では，伝統的な能力の概念に黙示的または明示的な挑戦を投げかける初期の文献を紹介し，特にパフォーマンスに帰属する要素（発話速度など §4），文より大きな構造に関連する言語知識（段落など §5），文法的な正しさを超えた他の要素（適切さなど §6）を取り入れ，言語産出の評価を行う。本論文では，過去50年間，言語教育の研究や言語評価の実践において大きな役割を果たした言語能力に対するコミュニカティブ・アプローチの台頭の背景を説明する（§7）。

Since the 1950s, linguistic competence has been traditionally understood as the idealized knowledge that is required to combine smaller units (e.g. words, phrases) into greater structures (e.g. clauses, sentences) in a correct way (§1). Models of competence therefore need to deal with the problem of the definition of linguistic units, which is especially thorny when applied to conversational speech (§2). While the emergence of speech units can be motivated on universal grounds, some aspects of speech unitization are language-specific, and thus tightly linked to linguistic competence (§3). In this study, I provide a documented account of the early literature (1900～1975) which poses implicit or explicit challenges to the traditional notion of competence, and notably by incorporating factors usually ascribed to performance (e.g. speech rate, §4), linguistic knowledge relative to structures larger than sentences (e.g. paragraphs, §5), and other dimensions for the evaluation of language production beyond grammatical correctness (e.g. appropriateness, §6). This brief account offers some context for the rise of communicative approaches to linguistic competence, which in the last 50 years played a major role in research on language education and in language assessment practices (§7).

## 1. The problem of competence

It is surprisingly difficult to define what it means to be able to speak a language. And yet, the issue of competence lies at the heart of many aspects of research and practice in linguistics. On the applied end of the spectrum, for example, defining linguistic competence is vital for designing standardized assessments, for monitoring progress in language education, and for setting goals in rehabilitation and speech therapy. On the theoretical end of the spectrum, different opinions on what counts as "knowledge of a language" can lead to the development of massively divergent theories and approaches. Indeed, the notion of competence can be used to illustrate both continuity and turning points in the history of linguistics.

For example, in the generativist tradition since the 1950s, linguistic competence is captured and represented by grammars. This is achieved by specifying how words can be legally combined into phrases, and how phrases can be combined into correct, grammatical sentences (Waslow 2003: p300). In this sense, a crucial aspect of linguistic competence lies in the knowledge required to create larger constituents, or "climb up" the levels of a syntactic hierarchy. This view opened opportunities for theoretical refinements in two directions.

A first line of investigation explored the outcome of the process, namely correct sentences. For example, the notion of correctness was questioned by challenging its categorical nature. Lakoff (1973: p277) proposed instead to use continuous values for well-formedness, citing difference in the grammaticality of sentences with preposed adverbs. While time adverbs prepose freely (e.g. *Tomorrow Sam will leave town* can become *Tomorrow it's likely that Sam will leave town*), manner adverbs do not (e.g. *Carefully Sam sliced the salami* cannot become *Carefully it's likely that Sam sliced the salami*). Similarly, grammaticality or correctness can be seen as only one of the aspects of linguistic knowledge. After all, even a posterchild of nonsensical well-formedness such as *Colorless green ideas sleep furiously* could be seen as absolutely meaningful and valid, for example in poetry and in a variety of other contexts (Erard 2010). Therefore, Coseriu (1985: pxxxiv) advocated for the consideration of additional dimensions besides well-formedness and grammaticality, such as congruence and appropriateness.

A second line of investigation focused on the input to the process, namely the language units which are combined into larger structures. For example, besides the traditional focus on combining words into sentences, it was recognized that linguistic knowledge also includes a representation of how individual sounds combine into larger structures like syllables. Greenberg & Jenkins (1963: p158) show that, in English, these structures can be seen as grammatical (e.g. /strʌk/), unattested (e.g. /strib/) or unacceptable (e.g. /gvsurs/), opening the field for the study of phonotactics.

Theoretical and descriptive linguistics offered different views of competence, based on (i) the size and type of the units that are combined, for example sounds or words, (ii) the properties of the produced structures, for example correctness or appropriateness and (iii) the way these properties are scaled, for example in a binary or continuous way. What these different approaches have in common, however, is a tendency to employ *exempla ficta* and introspection (e.g. /strib/, *Tomorrow it's likely that Sam will leave town*, *Colorless green ideas sleep furiously*). This top-down view was perhaps the only efficient approach in the decades before the advent of large-scale, deeply tagged and easily searchable digitalized corpora (Hopper 2012: p302). However, it only partially meets the needs of the applied end of the linguistics spectrum. Language teachers and speech therapists, in fact, need to assess the material

actually produced by their learners.

Such a bottom-up view is comparatively understudied. Especially when using speech data, researchers using actual productions for the study of linguistic knowledge run into immediate hurdles, notably at the stage of delimiting the relevant analysis units (Bowie & Popova 2020: §2.3). Therefore, as a stepping stone towards the development of such a bottom-up model of competence, in this study I present a brief collection of passages from the psychological and linguistic literature in the first three quarters of the 20th century, that is until competence became in the mid-1970s a hotly debated topic in language education. I focus in particular on the introduction to comparative phonetics published by Passy in 1906, a seminal text which has the advantage of predating the debate on competence and performance, thus offering a fresh perspective on the issue. Taken individually, most of these passages might seem to deal with the problem of defining linguistic units – which, as we have seen, is the first, proximal hurdle that the analyst faces. However, taken as a whole, the collection aims to illustrate how linguistic units relate to linguistic domains, which are then organized into hierarchies, which in turn embody a specific view of competence. These different views can be used to shape divergent linguistic theories, analytical frameworks, and educational goals – in other words, the ultimate, distal aim of generating insight in linguistics.

## 2. Speech unitization

At an abstract level, to characterize a sentence (or word) as grammatical (or as appropriate) we perform two operations: (i) focusing our attention at the level of analysis of the sentence (or word), thus delimiting its boundaries, and (ii) applying relevant tests to determine whether (or to what degree) they exhibit some properties, such as grammaticality (or appropriateness). In their abstract form, these two operations are not unique to linguistics. Early work in qualitative psychology sees them as underlying the general procedure of *coding*:

> The transformation of qualitative data obtained in interviews, autobiographies, free-answer responses to open-ended questions, projective materials, and observation of group situations into a form which renders them susceptible to quantitative treatment constitutes coding. The clinician and social psychologist increasingly use coding procedures to obtain more rigorous statistical demonstration of their hypotheses. […] The coding of qualitative data involves two operations, that of separating the qualitative material into units, and that of establishing category-sets into which the unitized material may be classified. The fruitfulness of the transformation depends upon the ingenuity and insight with which the experimenter chooses his units and category-sets. The reliability of the coding depends upon the accuracy with which the unitizing and subsequent classifying are carried out. (Guetzkow 1950: p47)

The quote highlights the importance of unitization, that is the separation of continuous action into units that can be evaluated and categorized. Compared to other scientists in the humanities, it might seem that linguists are in a privileged position, since "Of all forms of behavior the formal aspects of speech are most easily divided into units and measured objectively" (Goldman-Eisler 1958: p60).

In truth, this operation is never trivial, since it builds upon implicit and explicit assumptions. While this is true for all forms of language production, including written texts, this is particularly the case for conversational speech:

> Although the concept of the unit of a dialog or

> conversation is essential both for description and for analysis, there is no single accepted definition for the unit. For example, grammatical units can be used as a unit of analysis. They, however, seem inadequate for spontaneous conversations because in such conversations there are various obstacles to grammatical constructions, such as repairs, hesitations, and interruptions by other speakers, which make the judgment of grammatical units considerably difficult. Speaking turns can also be the unit of analysis. The concept of "turn," however, has been defined by many researchers in many different ways. [...] Thus, the identification of turns in real conversations is also very difficult. […] More objective definitions of the unit of analysis have been provided by some researchers, who have employed pauses as delimiters. […] Pauses are physically detectable as particular regions in speech based on energy measurements, provided that high-quality recorded speech materials are available. This resolves difficulty in objectively and reliably defining the unit. (Koiso et al. 1998: p298)

This strategy resulted in the adoption of the "interpausal unit (IPU), which is a stretch of a single speaker's speech bounded by pauses longer than 100ms" (Koiso et al. 1998: p299). Subsequent studies have confirmed the success of interpausal units in the analysis of spontaneous conversation, and suggested using alternative durational thresholds for their delimitation (e.g. 200ms, see Campione & Véronis 2002: p202). Pauses have been used as a cornerstone for speech unitization since the earliest systematizations of phonetics:

> The sounds which combine into speech do not present themselves as an uninterrupted, uniform series. They form groups, which are again divided and subdivided into smaller groups. [...] We all know that one cannot speak continuously for a long time without stopping. We stop for two reasons. First, because it is *impossible* to speak without stopping. [...] We must restock breath, fill the lungs again; and during this time we cannot speak. Second, because we *speak to be understood*, and we would not be, if we did not stop. (Passy 1906: §§48-49)[1]

The passage highlights the functional concerns behind the notion of *breath group*, that is the natural subdivision of speech into phrases. According to Passy, this is due universal constraints, which can be (i) physiological, for the benefit of the producer, namely breathing, or (ii) psychological, for the benefit of the perceiver, namely decoding. These dichotomies, however, should not be overemphasized. This is because every conversationalist is both a producer and perceiver of speech. Therefore, pauses can also serve the management of interaction:

> Let us assume that the speaker is motivated to keep control of the conversational "ball" until he has achieved some sense of completion. He has learned that unfilled intervals of sufficient length are the points at which he has usually lost this control – someone else has leapt into his gap. Therefore, if he pauses long enough to receive the cue of his own silence, he will produce some kind of signal ([m, ər], or perhaps a repetition of the immediately preceding unit) which says, in effect, "I'm still in control – don't interrupt me!" We would thus expect Filled Pauses and Repeats to occur just before points of highest uncertainty, points where choices are most difficult and complicated. (Maclay & Osgood 1959: p41)

The issue of uncertainty shows that, besides addressing the physiological needs of the producer and the psychological needs of the receiver (as in the quotes from Passy above), pauses can also serve the psychological needs of the producer, in

terms of mustering and deploying the resources required for speech planning:

> The experimental evidence […] has shown hesitation pauses to anticipate a sudden increase of information (measured in terms of transition probability); indeed, the close relation found to exist between pauses and information on the one hand and fluency of speech and redundancy on the other, seems to indicate that the interpolation of hesitation pauses in speech is a necessary condition for such an increase. [...] Fluent speech was shown to consist of habitual combinations of words such as were shared by the language community and such as had become more or less automatic. Where a sequence ceased to be a matter of common conditioning or learning, where a speaker's choice was highly individual and unexpected, on the other hand, speech was hesitant. (Goldman-Eisler 1958: p67)

A final way in which pauses interact with speech units is in the ease of articulation for the speaker, and notably in the management of word boundaries: "The frequent phenomena of elision and liaison have the goal of making it easier to pass from a word to the next, without any interruption" (Passy 1906: §55). Indeed, resyllabification phenomena are well-documented in many languages, albeit with differences in their domains of application. Specifically, the French liaison mentioned above refers to "a general process of linking which allowed consonants to survive final consonant deletion" (Durand & Lyche 2008: p34) in certain contexts, for example *petit prince* [pətipʁɛ̃s] versus *petit amie* [pətitami]. In other words, Passy recognizes that the management of pauses and phrasing is not exclusively driven by universal constraints, but that it also includes language-specific mechanisms. And if the construction of speech units is subject to language-specific constraints, then pausing, phrasing and unitization strategies become part of the linguistic competence of speakers.

## 3. Units and grammatical domains

The linguistic consequences of pausing should be considered as part of the knowledge that language learners are expected to acquire. In Passy's playful words:

> In fact, we do not perform liaison when there is a break. Nothing is more ridiculous than an inappropriate liaison. *C'est une idée* uttered as **sɛt, ynide** makes you think of a hiccup. (Passy 1906: §55)

In a similar vein, Passy also offers an early cross-linguistic example of how fine phonetic detail can serve as a cue to word segmentation:

> While in French *les zônes* and *les aunes* are pronounced exactly the same **le´zo**:**n**, English distinguishes *a name* **ə´neim** from *an aim* **ən´eim**. (Passy 1906: §59)

The shift in perspective from universal phonetic constraints to language-specific mechanisms is finally evident in his early discussion of vowel harmony:

> In Yakut, a Siberian language, […] the same ending takes four different forms, depending on the timbre of the preceding vowel (**ɑgɑ** 'father', **ogo** 'child', **æsæ** 'bear', **dœrœ** 'strap'). In Yakut it must be often possible to guess word boundaries without understanding them. (Passy 1906: §59)

The quotes above show how the continuous stretches observed in actual speech can be related to the application of various language-specific

phenomena. Words (e.g. Yakut harmony) and pauses (e.g. French liaison) play a crucial role in this respect. However, pausing can be modulated by speech tempo. After introducing the idea of breath groups as the natural division of speech due to universal constraints (see §2 above), Passy suggests that breath groups also tend to be isomorphic with units of meaning, since "*A breath group corresponds to the expression of a simple idea*, or in other words to an *elementary phrase*" (Passy 1906: §51). However, and crucially to our purposes:

> It often happens that two or three phrases are unified into a single breath group […] especially in familiar conversation, when we talk to people who are used to our way of speaking. On the contrary, in speech, lectures, etc., the same phrase can be split into several groups. The correspondence between breath group and elementary phrase takes place in familiar slow pronunciation. (Passy 1906: §51)

In other words, according to Passy, breath groups are at the intersection of phenomena that we could call of universal nature (e.g. emerging from breathing constraints), of language-specific competence (e.g. conditioning the application of liaison), and of situation-specific performance (e.g. regrouped according to the audience). This state of affairs might seem particularly confusing, especially from the point of view of formal linguistic theories that emerged half a century after the publication of Passy's work. In truth, as we will see in the next section, the line between performance factors and competence factors in the emergence of linguistic hierarchies can sometimes be blurred.

## 4. Domains, performance and constituent hierarchies

As we have seen above, the link between speech units and linguistic phenomena is not straightforward. For example, among other things, French liaison is modulated by the presence or absence of pauses, which is in turn related to the regrouping of phrases, which is in turn based on situational aspects such as familiarity with the audience or speech rate. This casts a shadow on the possibility of neatly defining the domain of application of various linguistic rules. Such a shadow was acknowledged in early work in the generative tradition, for example in studies on the phonological structure of Mandarin. Among the many phenomena discussed by R. Cheng (1966) is the fact that a word carrying a dipping tone (e.g. *měi* 美 'beautiful') is uttered with a rising tone instead, if followed by another word with a dipping tone (e.g. *méi jǐu* 美酒 'beautiful wine'):

> Syntactic information, however, is not the sole linguistic supra-syllabic factor in determining the phonetic reality of sentences. A factor such as speed seems to play an important role in determining whether some morphophonemic transformations will take place or not (among which are transformations from one tone to another). To incorporate this complexity the syntactic component can provide a basis for setting up several degrees of closeness (in terms of syntactic relationships) between each pair of syllables, and let the factor of speed decide what degree(s) of syntactic closeness (or depth as it is called by Wang, 1965) are eligible for or excluded from the same morphophonemic transformation. (Cheng 1966: p150)

Specifically, it is suggested that in long sequences of dipping tones, the application of the sandhi rule

that transforms a tone into another will depend on speed, or speech rate. This is because, just as in the example from Passy above, speech rate ultimately modulates how constituents are grouped. The idea of incorporating performance factors into the grammar of Mandarin tone sandhi has further been explored by C. Cheng (1970), who went as far as proposing the inclusion of attitudinal factors into tree diagrams.

Figure 1 links levels of casualness (i.e. numbers in the left branch) with the occurrence of pauses (i.e. numbers between the words in the terminal nodes). This suggests that, in fast speech, and thus at high levels of casualness, only few pauses are inserted, and thus fewer but longer constituents are generated, which in turn results in a wider application of the tone sandhi rule. For example, if the sentence *lǎo lǐ mǎi hǎo jǐu* 老李买好酒 'old Li buys good wine' is uttered at level of casualness 3, only a pause between NP (*lǎo lǐ*) and VP (*mǎi hǎo jǐu*) is expected, and thus the resulting expected utterance will be *láo lǐ / mái háo jǐu*, that is with the last words in each phrase (*lǐ*, *jǐu*) still maintaining the original dipping tone. On the other hand, in extremely careful speech (at levels of casualness 0 or 1), each word would be followed by a pause, and thus no tone substitution would occur (i.e. *lǎo lǐ mǎi hǎo jǐu*).

Consonantal alternations in Spanish offer another example of the early generative interest in how speech styles interact with the application of phonological phenomena. According to Harris (1969: p9), "certain alternations can be correctly accounted for only by recognizing descrete [sic] levels of style", such as "*Andante*: slow, careful, but unaffected" and "*Allegretto*: moderately fast tempo". This is, for example, the case of the alternation between initial stops and non-initial continuants:

> Let us clarify the statement that *b*, *d*, and *g* occur as stop [b, d, g] "initially": an utterance such as *Beatriz babea*, 'Beatriz slobbers', occurs as both [beatrizbaβea] and [beatrizβaβea]. The former represents the more careful pronunciation, and the latter the more casual pronunciation. Let us say that the former is Andante and the latter is Allegretto. Thus, for Andante, "initially" means both [after a silence and after a word boundary…]. In Allegretto, on the other hand, "initially" means only [after a silence]. (Harris 1969: p42)

The interplay between phonological rules, domains, unit boundaries and speech styles is then seen as defining phonological representations, which

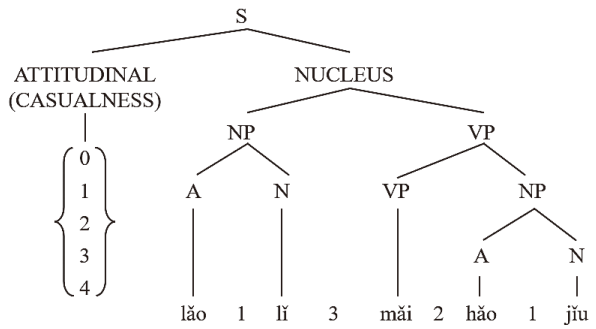


**Figure 1**
The grammar of Mandarin Tone 3 sandhi includes attitudinal factors.
Recreated from Cheng (1970).

are an important part of linguistic competence:

> ...one of the formal correlates of the stylistic distinction Andante/Allegretto is not the appearance of optional boundary elements in the rules of one style but not of the other, but rather that boundary elements are deleted in phonological representations at higher level of derivation in Allegretto than in Andante. (Harris 1969: p67)

These early attempts to build a bridge between sandhi phenomena on one side and pauses or tempo on the other side show the need for a hierarchy of constituents that are not merely defined in phonological or syntactic terms. In the following years, work on prosodic hierarchies and on the prosody-syntax mapping has indeed dealt with this very issue. Consistently with our interest in the early literature, however, I will not delve into proposals concerning the prosody-syntax interface. Instead, I will highlight yet another way in which stylistic factors were incorporated as a systematic modulating force into the knowledge required for unitizing speech. In his study of colloquial standard Japanese, Bloch shows that even facultative, stylistic, and otherwise meaningless pauses follow an implicational hierarchy. In other words, when pauses occur at a certain type of boundary (e.g. after direct object constituents), they are also expected to occur at a different type of boundaries (e.g. after certain topic constituents):

> All pauses within a sentence are preceded by level intonation. Every such pause is facultative: repeated utterances of the same sentence (by the same speaker or by different speakers) will show the pause sometimes present, sometimes absent, without any change in meaning. The presence or absence of pauses within a sentence depends partly on stylistic factors (with more pauses in emphatic or affective speech), partly on the tempo and the care of utterance.
>
> However, some facultative pauses are more constant than others, appearing more consistently in repeated instances of the same sentence. It is enough to distinguish two ranks of facultative pauses: higher (more constant), marked with a semicolon; and lower (less constant), marked with a comma. Where one or more lower-ranking pauses are followed by a higher-ranking pause, the lower-ranking pauses appear only if the higher-ranking pause is present. The sentence *Kono goro wa, yoohuku o, kaú no ni wa; okane ga, takusañ; irimásu yo* 'It takes a lot of money to buy a suit these days' will occur with all, some, and none of the facultative pauses actually present; but it will never occur with pauses after *yoohuku o* and *okane ga* unless there are also pauses after *kaú no ni wa* and *takusañ*. (Bloch 1946: p201)

While native speakers might feel that certain pauses in the example above could be interpreted as belonging to a different rank, this would not challenge the general idea of the existence of different ranks of optionality. Therefore, the quotes in this section show that linguistic and speech units are organized into constituents, which are then organized into hierarchies, which in turn can be seen as an aspect of language competence. However, these hierarchies are also modulated by factors that are usually not ascribed to competence, such as speech tempo.

## 5. At the top of the hierarchies

As seen above, traditional syntactic (or prosodic) hierarchies are best understood when performance factors are taken into consideration. Another way of refining our understanding of hierarchies is by focusing on their top node. In most cases, these

hierarchies culminate into a sentence-sized (or utterance-sized) constituent. Sentences and utterances have always had a privileged position in a variety of linguistic traditions, for example in Firth's insightful critique of phoneme-centered phonology:

> The primary linguistic data are pieces, phrases, clauses, and sentences within which the word must be delimited and identified. […] On the perception side, it is improbable that we listen to auditory fractions corresponding to uni-directional phonematic units in any linear sense. Whatever units we may find in analysis, must be closely related to the whole utterance, and that is achieved by systematic statement of the prosodies. In the perception of speech by the listener whatever units there may be are prosodically reintegrated. We speak prosodies and we listen to them. (Firth 1948: pp127, 152)

The sentence is thus a privileged unit for the linguist because it is also a privileged unit for the language user, since listeners do not attend to individual phonemes, but rather to long and context-rich stretches of information. There is however at least one other additional reason for the success of sentences or utterances as units of analysis:

> Nothing is generally said about the interrelations among whole utterances within a discourse. Now in many, perhaps all, languages there are particular successions among types of utterance within a discourse. This may be seen in a stretch spoken by one speaker (compare the first sentence of a lecture with one of the later sentences), or in a conversation (especially in such fixed exchanges as *"How are you?" "Fine; how are you?"*). Since these are distributional limitations upon the utterances with respect to each other within the discourse, they could be studied with the methods of descriptive linguistics. The amount of data and of analytic work required for such a study would, however, be much greater than that required for stating the relations of elements within single utterances. For this reason, the current practice stops at the utterance. (Harris 1951: p12)

In other words, relations between sentences and within even higher domains should be an integral part of linguistic knowledge. It is only by a restricted understanding of competence that linguists focus on the combination of morphemes into words, words into phrases, phrases into clauses, and clauses into sentences (or of phonemes into syllables, syllables into words, words into phrases, and phrases into utterances). Exploring competence at the discourse level is simply more difficult, and thus less practiced. There are, of course, brave exceptions:

> It is clear that people communicate by isolated sentences only in exceptional circumstances. A message is often long and complicated enough to require that it be cast in paragraph form. […] Sentences within a paragraph may be formally linked by the use of definite and indefinite articles, deictic adverbs and pronouns, sequence of tenses etc. It is hypothesized also that paragraphs possess a suprasegmental structure that indicates the beginning and end of paragraphs and characterizes the body of the paragraph. (Lehiste 1975: p195)

There is an essential similarity between well-known phenomena at the phrase level and understudied phenomena at the discourse level. Consider for example the amply documented relation between pitch reset and the prosodic hierarchy. The seminal study by Ladd (1988) featured sentences with 3 clauses, joined by either

*and* or *but*, resulting in structures like *A but B and C* or *A and B but C* (e.g. "Allen is a stronger campaigner, *and* Ryan has more popular policies, *but* Warren has a lot more money"). Since "these sentences have a natural interpretation in which the but-boundary is stronger" (Ladd 1988: p532), it was hypothesized that their pitch contours would differ, and indeed clause-initial peaks were found to be higher after stronger boundaries. The vertical dotted line in Figure 2 shows the point in the utterance where peaks are reset to a higher pitch and a new pitch declination line.

Extending this line of thinking from relationships between clauses to relationships between paragraphs, Beck & Bennet (2007) investigated oral narratives in Lushootseed, a Salishan language of Washington State. Based on thematic and narrative grounds, they grouped utterances into paragraphs, that is, nodes higher than the utterance in the prosodic structure. Similarly to Ladd (1988), they found that:

> [Phonological Paragraphs] are marked by declination patterns in the F0 Maxima of Utterances, which tend to decline over the length of the [Phonological Paragraph] and then are reset to mark the beginning of a new discourse-level prosodic unit. The phonetic evidence for [Phonological Paragraphs] is supported by morphosyntactic data such as coincidence of [Phonological Paragraph]-boundaries with grammatical particles, topic subject-continuity, and the distribution of syntactically and/or phonologically marked topic shifting structures. (Beck & Bennet 2007: p32)

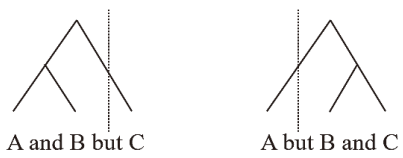


**Figure 2**
Hierarchical relations are indexed by pitch reset.
Recreated from Ladd (1988).

This amounts to stating that phonetic, phonological, morphological, syntactic and pragmatic phenomena indicate that speakers know how to organize language above the level of the utterance or sentence. This suggests that traditional hierarchies should be expanded to include higher levels, such as the paragraph, in order to represent the linguistic knowledge that speakers have of these structures.

## 6. Competence as forest, not trees

As stated in §1, the ability to climb up the syntactic or prosodic hierarchies and to combine smaller elements into grammatically correct higher structures can be seen as a crucial ingredient of linguistic competence. However, as we have seen in §4, it is necessary to acknowledge that these hierarchies can be modulated by factors that are traditionally seen as outside the core notion of competence, as in the case of speech rate effects on sandhi phenomena. Moreover, as we have seen in §5, since traditional hierarchies culminate in sentences or utterances, they are ill-suited to capture the linguistic competence that language users need to manage discourse. The example above from Lushootseed shows how such competence is required in the management of oral narratives. One might wonder whether other types of language activities and of speech styles would be less dependent on this type of competence at the discourse level. However, a truly encompassing understanding of competence must indeed include a variety of possible language uses. As anticipated in §1, the correct formation of sentences is only a part of what language learners are expected to acquire. In real language use situations, in fact,

sentences are uttered as part of a larger environment, both in the sense of speech (i.e. the discourse context) and in the sense of action (i.e. the social setting):

> In their search for methodological rigor, linguists tended to confine themselves to the internal linguistic patterning of linguistic forms within isolated sentences, ruling out consideration of the broader conversational context or the social settings in which such sentences are embedded. The resulting grammars account for what can be said in particular language [sic], but they make no attempt to specify what constitutes appropriate behavior in particular social circumstances. (Gumperz 1970: p3)

In other words, by equating linguistic competence with the correct generation of grammatical sentences, one cannot capture the full picture of how and why humans learn to use language:

> We have then to account for the fact that a normal child acquires knowledge of sentences, not only as grammatical, but also as appropriate. He or she acquires competence as to when to speak or not, and as to what to talk about with whom, when, where, in what manner [...] The acquisition of such competency is of course fed by social experience, needs, and motives, and issues in action that is itself a renewed source of motives, needs, experience. We break irrevocably with the model that restricts the design of language to one face toward referential meaning, one toward sound, and that defines the organization of language as solely consisting of rules for linking the two. Such a model implies naming to be the sole use of speech, as if languages were never organized to lament, rejoice, beseech, admonish, aphorize, inveigh [...], for the many varied forms of persuasion, direction, expression and symbolic play. A model of language must design it with a face toward communicative conduct and social life. (Hymes 1972: p277)

In the last 50 years, it has become increasingly clear that language education needs to guide learners towards the ability to function in a variety of social contexts. This has led to a variety of insightful critiques of the traditional and reductionist approach to competence in the chomskyan tradition, as well as a number of important proposals on how to assess competence, especially in the context of language education (e.g. Lyons 1977, Canale & Swain 1980, Bachman 1990; see Erton 2017 for a review). Given the focus of this study on the early literature, I will note instead that the American structuralist tradition was already employing a rich, layered understanding of competence:

> The Menomini Indians of Wisconsin, a compact tribe of some 1700 people, speak a language without dialectal differences and have no writing. Yet the Menomini will say that one person speaks well and another badly, that such-and-such a form of speech is incorrect and sounds bad, and another too much like a shaman's preaching or archaic. […] Here is a sketch of the linguistic position of some of the speakers whom I knew best: […] Bird-Hawk, a very old man, who has since died, spoke only Menomini, possibly also a little Ojibwa. As soon as he departed from ordinary conversation, he spoke with bad syntax and meagre, often inept vocabulary, yet with occasional archaism. […] White-Thunder, a man round forty, speaks less English than Menomini, and that is a strong indictment, for his Menomini is atrocious. His vocabulary is small; his inflections are often barbarous; he constructs sentences on a few threadbare models. He may be said to speak no

language tolerably. His case is not uncommon among younger men, even when they speak but little English. (Bloomfield 1927: p436)

In the affectionately scathing description of White-Thunder, we find an implicit definition of competence that encompasses both the generativist focus on correctly managing the syntactic hierarchy ("he constructs sentences on a few threadbare models") and two aspects commonly used in language education (viz. vocabulary size and pronunciation). In the description of Bird-Hawk, however, we also find a consideration of the speaker's ability to function in a variety of speech registers and social situations ("as soon as he departed from ordinary conversation").

## 7. The assessment of competence

To sum up, by the mid-1970s it was already clear that competence is a multilayered construct which requires consideration of both grammatical and social aspects, of both correctness and appropriateness, of both units and their context, of both infra-syllabic and ultra-sentential domains, as well as the inclusion of factors that were not usually ascribed to competence, such as speech rate, hesitation, pausing, and the management of conversation.

If this is the case, how can we hope to translate such a complex concept into an operable construct for the assessment of competence? This question is of immediate relevance for language educators and for the many institutions that rely on language testing for their functioning. The topic deserves its own focused review, since it has attracted a large share of attention in the last 50 years, thus following the period from Passy to Lehiste that has made the object of this study.

## Acknowledgments

This research was conducted while serving as Research Fellow at the Institute of Language Research of the Tokyo University of Foreign Studies. I am grateful for comments from two anonymous reviewers and for feedback at the June 2024 meeting of the Tokyo Circle of Phonologists (Seikei University).

## Note

1 Here and in the following quotes by Passy, I provide my own translation from the French original, while respecting the original emphasis and formatting.

## References

Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford University Press.

Beck, D., & Bennett, D. (2007). Extending the prosodic hierarchy: Evidence from Lushootseed narrative. *Northwest Journal of Linguistics 1*(1), 1-34.

Bloch, B. (1946). Studies in colloquial Japanese II (syntax). *Language 22*(3), 200-248.

Bloomfield, L. (1927). Literate and illiterate speech. *American Speech 2*(10), 432-439.

Bowie, J., & Popova, G. (2020). Grammar and discourse. In B. Aarts, J. Bowie & G. Popova (Eds.), *The Oxford Handbook of English Grammar* (pp. 554-578). Oxford University Press.

Campione E., & Véronis J. (2002). A large-scale multilingual study of pause duration. *Proceedings of the First International Conference on Speech Prosody*, 199-202.

Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics 1*(1), 1-47.

Cheng, R. (1966). Mandarin phonological structure. *Journal of Linguistics 2*(2), 135-158.

Cheng, C. (1970). Domain of phonological rule application. In J. Sadock & A. Vanek (Eds.), *Studies presented to Robert B. Lees by his students* (pp. 39-60). Linguistic Research.

Coseriu, E. (1985). Linguistic competence: What is it really? *The Modern Language Review 80*(4), xxv-xxxv.

Durand, J., & Lyche, C. (2008). French liaison in the light of corpus data. *Journal of French Language*

*Studies 18*(1), 33-66.
Erard, M. (2010). The life and times of "Colorless green ideas sleep furiously". *Southwest Review 95*(3), 418-425.
Erton, İ. (2017). Types of competence in linguistics: A review of processes and their implications in human perception and action. *DTCF Dergisi 57*(1), 157-170.
Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society 47*(1), 127-152.
Goldman-Eisler, F. (1958). Speech analysis and mental processes. *Language and Speech 1*(1), 59-75.
Guetzkow, H. (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology 6*(1), 47-58.
Gumperz, J. (1970). *Sociolinguistics and communication in small groups* (Working Paper No. 33). Language-Behavior Research Laboratory, University of California, Berkeley.
Harris, Z. (1951). *Methods in structural linguistics*. University of Chicago Press.
Harris, J. (1969). *Spanish phonology*. The MIT Press.
Hymes, D. (1972). On communicative competence. In J. Pride & J. Holmes (Eds.), *Sociolinguistics. Selected readings* (pp. 269-293). Penguin.
Koiso, H., Horiuchi, Y., Tsuchiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech 41*, 295-321.
Ladd, R. (1988). Declination "reset" and the hierarchical organization of utterances. *Journal of the Acoustical Society of America 84*(2), 530-544.
Lakoff, G. (1973). Fuzzy grammar and the performance/competence terminology game. *Papers from the Ninth Regional Meeting of the Chicago Linguistic Society*, 271-291.
Lehiste, I. (1975). The phonetic structure of paragraphs. In A. Cohen & S. Nooteboom (Eds.), *Structure and process in speech perception* (pp. 195-203). Springer-Verlag.
Lyons, J. (1977). *Semantics*. Cambridge University Press.
Maclay, H., & Osgood, C. (1959). Hesitation phenomena in spontaneous English speech. *Word 15*(1), 19-44.
Passy, P. (1906). *Petite phonétique comparée des principales langues européennes*. Teubner.
Wasow, T. (2003). Generative grammar. In M. Aronoff, & J. Ress-Miller (Eds.), *The Handbook of Linguistics* (pp. 297-298). Blackwell.