Growth Points, Catchments, and Contexts

David McNeill

The growth point is a theoretical unit of language and thought proposed as the initiating cognitive unit in the production of speech. Growth points are inferred from speech-gesture synchrony and co-expressivity, and incorporate elements from the context of speaking by virtue of what Vygotsky (1987) called ‘psychological predicates.’ The psychological predicate in any given case can be known only in relation to its local context, from which it is the point of differentiation. Thus growth points, as psychological predicates, incorporate context as an inherent property. The first part of the paper consists of a detailed case study of the growth point of a recorded English utterance. The aim is to demonstrate how this growth point was inferred and how it, and no other, and the full sentence it was unpacked into, could have been products of contextual differentiation. The growth point ‘predicts’ a context, which must be observed in order to validate the growth point in each given case. An essential tool of analysis is the concept of a catchment — a thematic unit in gesture revealed through partial form and space recurrences. In our case study, three catchments are involved. The second part of the paper consists of a comparison of the growth point to Kita’s information packaging hypothesis. One difference is that the growth point defines psychological units whereas the information packaging hypothesis, despite its name, does not. This part of the paper also presents an analysis of gestural foreshadowing as seen in a cross-linguistic example (Turkish). Foreshadowing at first glance seems problematic for the growth point, because synchrony appears to break down, but in fact it leads to new insights into mechanisms of growth point composition. The paper concludes with the statement that language is inseparable from imagery — a concept not deducible from the tradition of text-based linguistic analyses. Imagery requires finding new theoretical concepts about language in non-traditional domains.

Keywords: gesture, language and thought, context, speech, cross-linguistic comparison

Introduction — Gestures

1) Preparation of this paper was supported by grants from the Spencer Foundation and the National Science Foundation (STIMULATE and KDI programs). I wish to thank Sotaro Kita for many valuable suggestions.

University of Chicago.
a lack of language-defining properties, idiosyncratic form-meaning pairings, and a precise synchronization of meaning presentations in gestures with co-expressive speech segments. This stable of characteristics shows that gestures and speech are systematically organized in relation to one another. The gestures are meaningful. They form meaningful, nonredundant combinations with the speech segments with which they synchronize, despite the fact that they are idiosyncratic and ephemeral. The implications of gestures for understanding of the nature of language itself are profound. The lack of linguistic properties in gesture combined with the phenomenon of speech-gesture synchrony opens a window into the mind that is otherwise curtained. In this paper, my focus will be on a theoretical analysis of the speech-gesture nexus, what it reveals of the human mind and its workings, how it enriches our view of language, and what form cultural differences take within a universal psychological system of speech-gesture linkage.

The Growth Point

The growth point (GP) is a unit of verbally engaged thinking based on speech-gesture correlations. It is a unit in which both imagery (from gesture) and language content (present in the form of linguistic categories) are combined. GPs are inferred from the totality of communicative events with a special focus on speech-gesture synchrony and co-expressivity. Following Vygotsky (1987), a GP is assumed to be a minimal psychological unit; that is, the smallest unit that retains the essential properties of a whole, in our case the whole of an image and a linguistically-codified meaning category, such as we see in the speech-gesture window. We use the gesture’s semantic content and its synchrony (that is, the synchrony of the gesture stroke phase\(^2\)) with speech to infer the GP.

A GP is called a growth point because it is meant to be the initial form of thinking out of which speech-gesture organization emerges. It is also called a GP since it is a theoretical unit with the potential for dynamic change during speech production. A final reason for calling it a GP is that it addresses the concept that there is a definite starting point for an idea unit. An idea unit is not always present; it comes into being at a certain moment. It has roots in the previous context, but the idea unit does not itself exist before this moment\(^3\).

The growth point is thus a point. A GP combines imagery and linguistic categorial content into a single unit. It is inferred empirically from on speech-gesture synchrony and co-expressivity and is validated by examining the context in which it occurs. The context must have been one in which the proposed GP could take form. A context that is inappropriate falsifies the hypothesized GP (McNeill & Duncan, 2000). The conditions that context must meet in order for a GP to take form are explained below. This paper will present an extended case study of a GP and analyze the relationship of this GP to its context. The case study is in no way special or atypical and we can draw general conclusions from it concerning the interrelations of language, imagery, and context.

Growth Point Differentiation and Psychological Predicates

To understand the relationship of context to a GP, it is first necessary to understand how the GP is a product of differentiation. Growth points

---

2) Gestures are temporally organized movements in which typically the limbs pass through up to 5 phases: preparation, prestroke hold, stroke, poststroke hold, and retraction. The stroke is the only obligatory phase. It carries the meaning of the gesture and has the quality of "effort" — a kinetic emphasis of the idea unit.

3) Dennett & Kinsbourne (1992) have argued against such a concept on the ground that it confounds the experiencing of a new conscious element with the temporal onset of the element itself. The temporal anatomy of a gesture however appears to answer part of their argument, namely, the onset of gesture preparation provides a definite locus for the beginning of a new thought element not confounded with the conscious experience of that element itself.
emerge as the newsworthy elements in the immediate context of speaking. They are, to adopt another concept from Vygotsky (1987), psychological predicates (not always grammatical predicates).

The concept of a psychological predicate illuminates the theoretical link between the GP and the context of speaking. Defining a psychological predicate (and hence a GP) requires reference to the context; this is because the psychological predicate and its context are mutually defining. The GP:

1. marks a significant departure in the immediate context; and
2. implies this context as a background.

We have in this relationship the seeds for a model of realtime utterance generation and coherent text formation.

Regarding the GP as a psychological predicate suggests a mechanism of GP formation in which differentiation of a focus from a background plays an essential part. Such differentiation is validated by a very close temporal connection of gesture strokes with the peak of acoustic output in speech. Nobe has documented this connection instrumentally: “The robust synchrony between gesture strokes and the peaks of acoustic aspects suggests that the information the gesture stroke carries has an intrinsic relationship with the accompanying speech information prominently pronounced with these peaks. The manifestation of the salient information seems to be realized through the synchronization of these two modalities” (Nobe, 1996, p. 35).

Regarding the GP as a psychological predicate also clarifies the sense in which we use the term ‘context’. This term has a host of meanings (cf. Duranti & Goodwin, 1992), but for our purposes ‘context’ is the background from which a psychological predicate is differentiated. This background indexes and is constrained by external conditions, both social and material, but an essential fact is that the background is also under the control of the speaker; it is a mental construction, part of the speaker’s effort to construct a meaning. The speaker shapes the background in a certain way, in order to make possible the intended significant contrast within it. Background and contrast are both necessary and are constructed together.

I will use the terms field of oppositions and significant (or newsworthy) contrast to refer to this constructed background and the differentiation of the psychological predicate. All of this is meant to be a dynamic, continuously updated process in which new fields of oppositions are formed and new GPs or psychological predicates are differentiated in ongoing cycles of thinking for speaking.

Case Study of a Growth Point

The narrative text in (1) through (9) below will be used to illustrate a number of features important for explaining the GP. The narrative is by an adult native English speaker recounting from memory the story of an animated color cartoon that she had just watched (a.c. 1950 Warner Brothers Tweety and Sylvester saga). The speaker and her listener were told that the listener would have to retell the story to a third person. There was no mention of gestures. The scene she is describing is outlined in the following:

- Sylvester crawls in drainpipe at bottom
- Tweety Bird dumps bowling ball into top of drainpipe
- Scene shifts to side of building, central portion of drainpipe
- Bulge moves down drainpipe
- Bulge explodes
- Sylvester falls into street with bowling-ball shaped bottom
- Sylvester rolls down street on bottom, feet rotating above ground
- Sylvester rolls down different part of street, looks upset
- Scene of empty street

4) Transcription by Elena Levy; see appendix of McNeill (1992) for the full transcript.
• Scene shifts to end of street; bowling alley
• Sylvester rolls into entrance of bowling alley
• Entrance of bowling alley without Sylvester; sounds of bowling pins crashing

Boldface in the gesture transcription marks the ‘stroke’ — the phase of gesture movement with semantic content (# shows a breath pause). Underlining marks holds (the hands held in place in midair). In Line 2 there is both a ‘pre-stroke’ hold on “drops” and a ‘post-stroke’ hold at the end of “down”, as the word is completed (Kita, 1990). Details of the method and transcription technique are provided in McNeill (1992).

Battle Plan for Vivian
5).

(1) he tries going [up the inside of the drain pipe and]

1hand: RH rises up 2X

(2) Tweety Bird runs and gets a bowling ba[ll and drops it down the drain|pipe #
Symmetrical: 2 similar hands move down

(3) [and / as he’s coming up]

Asymmetrical: 2 different hands, LH holds, RH up

(4) [and the bowling ball’s coming d[]

Asymmetrical: 2 different hands, RH holds, LH down

(5) [own he ss swallows it]

Asymmetrical: 2 different hands, RH up, LH down

(6) [ # and he comes out the bottom of the drain]

1hand: LH comes down

(7) [n pipe and he’s got this big bowling ball inside h|im

Symmetrical: 2 similar hands move down

(8) [and he rolls on down into a bowling all]

Symmetrical: 2 similar hands move forward 2X

(9) [ey and then you hear a sstri]ke #

Symmetrical: 2 similar hands move apart

To focus on one item for analysis, consider the utterance and gesture in (2) — the words “it down” plus a downward thrusting gesture. My purpose will be to show how this utterance-gesture combination can be explained utilizing the GP model. I have chosen this example in part because the linguistic elements “it” and “down” do not form a grammatical entity. While “it” is the direct object of “drops” and “down” is a satellite of this verb (cf. Talmy, 1985), “it” and “down” together do not comprise a grammatical unit. Nonetheless, in this example they are the inferred growth point. The challenge is to explain this GP and show why it and no other took form in the context of speaking.

First, to explain the GP itself. The gesture in (2) was made with two symmetrical hands — the palms loosely cupped and facing downward as if placed on top of a large spherical object, and the hands moved down during the linguistic segments “it do(n)”. The inferred GP is this image of downward movement plus the linguistic content of the “it” (i.e., the bowling ball) and the PATH particle “down”. The GP is both image and linguistic categorial content: an image, as it were, with a foot inside the door of language. Such imagery is important, since it grounds the linguistic categories in a specific visuo-spatial context. It may also provide the GP with the property of ‘chunking’, a hallmark of expert performance, (cf. Chase & Ericsson, 1981), whereby a chunk of linguistic output is organized around the presentation of an image. The downward content of the gesture is a specific case of “down”, the linguistic category — a specific visualization of it — in which imagery is the context of the category and possibly the unit of performance. The linguistic categorization is also crucial, since it brings the image into the system of categories of the language, which is both a system of classification and a way of patterning action. The speech and its synchronized gesture are the key to this theoretical unit.

5) Transcription is by many hands — Susan Duncan is chiefly responsible for this version. The transcription has been simplified in certain respects for ease of reading.
Table 1  Catchment Structure in Vivian’s Battle Plan

<table>
<thead>
<tr>
<th>Ln</th>
<th>Catchment</th>
<th>Utterance</th>
<th>Gesture Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>C1</td>
<td>he tries going <strong>[up the inside of the drainpipe and]</strong></td>
<td>1-hand (right)</td>
</tr>
<tr>
<td>2</td>
<td>C2</td>
<td>Tweety Bird runs and gets a bowling ball and drops <strong>it down</strong> the drainpipe</td>
<td>2-similar hands</td>
</tr>
<tr>
<td>3</td>
<td>C3</td>
<td><strong>[and as he’s coming up]</strong></td>
<td>2-different hands</td>
</tr>
<tr>
<td>4</td>
<td>C3</td>
<td><strong>[and the bowling ball’s coming down]</strong></td>
<td>2-different hands</td>
</tr>
<tr>
<td>5</td>
<td>C3</td>
<td>own he <strong>swallows it</strong></td>
<td>2-different hands</td>
</tr>
<tr>
<td>6</td>
<td>C1, C3</td>
<td><strong>[and he comes out the bottom of the drain]</strong></td>
<td>1-hand (left)</td>
</tr>
<tr>
<td>7</td>
<td>C2</td>
<td>[npipe and he’s got this big bowling ball inside hjim]</td>
<td>2-similar hands</td>
</tr>
<tr>
<td>8</td>
<td>C2</td>
<td><strong>[and he rolls on down into a bowling alley]</strong></td>
<td>2-similar hands</td>
</tr>
<tr>
<td>9</td>
<td>C2</td>
<td>[ey and then you hear a strike]</td>
<td>2-similar hands</td>
</tr>
</tbody>
</table>

Incorporating Context — Catchments

A GP is a psycholinguistic unit based on contrast, and this concept brings in context as a fundamental component. As mentioned above, I use the terms field of oppositions and significant contrast to analyze the role of context in thinking while speaking. A significant contrast and the field of oppositions within which it is drawn are linked meaning structures, under the creative control of the speaker at the moment of speaking. Control by the individual ensures that GPs establish meanings true to the speaker’s intentions and memory. The formation of a GP is the highlighting or differentiation of what is novel in such a field. The field defines the significance of the contrast, it establishes what is meaningful about it; the contrast itself is the source of the GP.

To explain the utterance in (2) and why it has the growth point we infer, we must consider the complete context above and study how the different parts of it came together in the utterance. A useful approach to this analysis is by means of catchments — a phenomenon first noted by Kendon in 1972 (although he did not use the term ‘catchment’). See Table 1 for the catchments apparent in Vivian’s Battle Plan. A catchment is recognized from a recurrence of gesture features over a stretch of discourse. It is a kind of thread of consistent visuospatial imagery running through a discourse segment that provides a gesture-based window into discourse cohesion. The logic of a catchment is that discourse themes produce gestures with recurring features; these recurrences give rise to the catchment. Thus, working backwards, the catchment offers clues to the cohesive linkages in the text with which it co-occurs. Catchments are recognized from 2 or more gestures (not necessarily consecutive) with partially or fully recurring features of shape, movement, space, orientation, dynamics, etc.

The example illustrates 3 catchments. The catchments in this case are recognized from handedness and shape, and each of the features is based on the thematic content that motivates it: C1 is a about a single moving entity and its recurring gesture feature is a single moving hand; C2 is about a bowling ball and what it does, and its recurring feature is a rounded shape (in gesture transcription terms, ‘2 similar hands’ with shape details added); C3 is about the relative positions of two entities in a drainpipe and its recurring feature involves two hands in the appropriate spatial configuration (‘2 different hands’). C1. The first is the catchment of one-
handed gestures in items (1) and (6). These gestures accompany descriptions of Sylvester's motion, first up the pipe then out of it with the bowling ball inside him. Thus C1 ties together references to Sylvester as a solo force. This one-handed catchment differs from the two-handed gestures, which in turn divide into two other catchments:

**C2.** Two-handed symmetrical gestures in (2), (7), (8) and (9). These gestures group descriptions where the bowling ball is the antagonist, the dominant force. Sylvester becomes what he eats, a kind living bowling ball, and the symmetric gestures accompany the descriptions where the bowling ball asserts this power. In (2) the bowling ball is beginning its career as antagonist. The rest of the catchment is where it has achieved its result. A two-handed symmetric gesture form highlights the shape of the bowling ball or its motion, an iconicity appropriate for its antagonist role.

**C3.** Two-handed asymmetrical gestures in items (3), (4) and (5). This catchment groups items in which the bowling ball and Sylvester mutually approach each other in the pipe. Here, in contrast to the symmetric set, Sylvester and the bowling ball are equals differing only in their direction of motion.

With these catchments, we can analyze the real-time origins of the utterance and gesture in (2) in a way that incorporates context as a fundamental component.

The occurrence of (2) in the symmetrical catchment shows one of the factors that comprised its field of oppositions at this point — the various guises in which the bowling ball appeared in the role of an antagonist. This catchment set the bowling ball apart from its role in C3 where the bowling ball was on a par with Sylvester. The significant contrast in C2 was the downward motion of the bowling ball toward Sylvester. Because of the field of oppositions at this point, this downward motion had significance as an antagonistic force. We can write this meaning as Antag-

onistic Force: Downward toward Sylvester; this was the context and contrast. Thus, “it down”, unlikely though it may seem as a unit from a grammatical point of view, was the cognitive core of the utterance in (2) — the “it” indexing the bowling ball and the “down” indexing the significant contrast itself in the field of oppositions.

The verb “drops”, therefore, was excluded from this GP. Exclusion is evidenced in the fact that the stroke did not synchronize with the verb; in fact, it was withheld from the verb by the pre-stroke hold. We can explain this as follows. The verb describes what Tweety did, not what the bowling ball did (it went down), and thus was not a significant contrast in the field of oppositions involving the bowling ball. The core idea at (2) was the bowling ball and its action, not Tweety and his.

**One Utterance, Several Contexts**

That “drops” was excluded from the GP yet was included in the utterance points to a second context at play in the origins of (2). The utterance, though a single grammatical construction, grew out of two distinct contexts and gained oppositional meaning from each.

The first context we have already analyzed; it was the C2 theme in which the bowling ball was an antagonistic force. The second context can be seen in that the gesture at (2) also contrasted with C1 — the preceding one-handed gesture depicting Sylvester as a solo force. This significant contrast led to the other parts of the utterance in (2) via a partial repetition of the utterance structure of (1). Contrasting verbal elements appeared in close to equivalent slots (the match is as close as possible given that the verb in (2) is transitive while that in (1) is intransitive):

| (1') | (Sylvester) | up | in “he tries going up the inside of the drainpipe” |
| (2') | (Tweety) | down | in “and (Tweety) drops it down the drainpipe” |
The thematic opposition in this paradigm is counter forces — Tweety-down vs. Sylvester-up. Our feeling that the paradigm is slightly ajar is due to the shift from spontaneous to caused motion with “drops”. This verb does not alter the counter forces paradigm but transfers the counter force from Tweety to the bowling ball, as appropriate for the gesture with its downward bowling ball imagery.

The parallel antagonistic forces in (1’) and (2’) made Tweety the subject of (2’), matching Sylvester as subject of (1’). The contrast of (2’) with (1’) thus had two effects on our target utterance. It was the source of the verb, “drops,” and was also why the subject was “Tweety,” rather than “bowling ball.” The verb expressed Tweety’s role in the contrast and shifted the downward force theme to the field of oppositions about the bowling ball. The subject identity expressed the antagonistic forces paradigm. The prestroke hold over “drops” is thus also explained: the verb, deriving from an antagonistic forces context, was propaedeutic to the GP, and the stroke was withheld until the way had been prepared for it (the poststroke hold on the second half of “down” derives differently, as Kita (1990) observed: the mechanical downward stroke was over before the word and was held until the co-expressive content could be discharged).

Let’s summarize how (2) came into being:

1. The field of oppositions in which the significance of the downward motion of the bowling ball was that of an antagonistic force — the contrast of (2) with (3), (4), (5): this gave the growth point core meaning centered on “it down”. It’s noteworthy that the preparation for the gesture in (2) began in preceding clause, concurrent with mentioning the bowling ball for the first time (“Tweety Bird runs and gets a bowling ball[ and drops it down the dra[pipe”]). That is, the new growth point embodying the idea of the bowling ball in its role as the antagonist to Sylvester began to take form as soon as the bowling ball itself entered into the discourse.

2. The field of oppositions in which the significance was the counter forces of Sylvester-up vs. Tweety-down. This gave a sentence schema that included the words “drops”, “down”, “drain-pipe”, and the repetition of the sentence structure with Tweety in the subject slot.

The choice of verb in (2) was “drops,” rather than “throws”, “thrusts” or some other caused-motion option for a downward trajectory from a manual launch, possibly because it, among these options, corresponds most closely to the force-dynamics of how Tweety made use of gravity to launch the bowling ball\(^6\). Thus, a further aspect of the context of (2) is this force-dynamics. If this is the case, we have a further argument for the analysis above in which “drops” and “it down” are contrasts in different contexts. This comes from the hand shape of the gesture in (2). The speaker made the gesture with her hands facing down, in a thrusting position. They were not positioned for a simulation of Tweety’s hands when he exploited gravity to launch the bowling ball. Tweety in the cartoon stimulus held the bowling ball from the bottom and chucked it into the pipe, allowing gravity to do the rest. The GP image altered this force-dynamics by making the launch into a thrust. The gesture, that is, reflected the speaker’s reconceptualizing of the cartoon. The new force-dynamics is not appropriate to Tweety, but it does fit the field of oppositions that concentrated on the force-dynamics of the bowling ball in its persona as antagonist\(^7\).

The question of how a GP leads to a grammatically allowable surface form can be answered in part with filters or templates. The idea unit in a GP, being a point of differentiation from

---

6) Pointed out by Karl-Erik McCullough.
7) Sotaro Kita has pointed out that the first-person viewpoint inherent in the gesture at (2) in which the speaker’s hands depict Tweety’s hands is deducible from the agentivity tier (in Jackendoff’s 1990 terminology) in the utterance. While the Sylvester-Tweety opposition is not a new contrast at this point and so cannot be a psychological predicate, the agentivity tier does carry new information in that Tweety now has ‘done something’. And this fact is embodied in a first-person character viewpoint gesture.
a background, must be realized in a surface position consistent with its communicative dynamism, but this fact does not, by itself, specify a position in a sentence framework. The concepts of ‘construction grammar’ (Goldberg, 1995) and of ‘emergent grammar’ (Hopper, 1987) may apply here. The contextual weight generated for the initially non-grammatical pair, “it down” plus gesture, could be completed by accessing a caused-motion construction,

<table>
<thead>
<tr>
<th>Subj</th>
<th>V</th>
<th>Obj</th>
<th>Obl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tweety drops it (b-ball) down</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This construction type, learned from childhood and functioning as a unit, could ease the GP, despite its double contextual base, into a single grammatical format and guide the speaker to a well-formed output. This is ‘emergent’ in the sense that the grammatical form was not part of the input (this was the GP, the non-grammatical “it down” plus gesture), yet the output was this grammatical pattern. Thus, constructions, retrievable as wholes, can provide part of the template that the speaker needs.

Unpacking

‘Unpacking’ is the process of articulating the implications of the core idea and using these implications as a guide toward a well-formed surface structure. A key idea is that meaning continues to evolve during this process. The GP is not a mere input to a sentence formulator that leads to a sentence output (cf. Levelt, 1989); unpacking rather is regarded an ongoing process of growth. The GP is the starting point: meaning continues to emerge and as it does so channels the structure into surface forms, and this is a temporally extended process.

In the case study, the emergence of “drops” is an illustration of the continued transformation of meaning during the genesis of the utterance. The accessing of the caused-motion construction is another. In both cases, new meanings — “drops”, the idea of caused-motion — emerged as part of the process of unpacking the core meaning of the bowling ball as an antagonistic force. The verb of the sentence and a construction such as caused-motion semantic are typically regarded as foundation stones of the sentence structure. However, taking into account imagery the and catchment content of the utterance, we can see that they were, in this case, derived, secondary features, generated through the unpacking of the core meaning that was neither “drops” nor caused motion but an antagonist force. The unpacking sequence would have begun with $C2$ and then turned to $C1$. The word order (“drops it down”) obviously does not correspond to this sequence. The speaker was thinking in terms of “it down” plus the downward thrusting image as the realization of antagonistic force, and then came up with “drops” and caused motion as a means of presenting the core meaning.

Comparison to Information Packaging

Kita (2000) has proposed a new hypothesis similar in some respects to the GP, that he term the Information Packaging Hypothesis (IPH). I will suggest below that the GP and IPH hypotheses can be combined, but I will start by describing how they differ. An important difference is the concept of a unit. Despite the Information Packaging Hypothesis name, it does not have units as they exist in the GP. This is because speech and gesture are held to be independent systems in the IPH. The concept resembles that in Paivio, 1971. Imagery and language intertwine as continuous threads and are available as alternate modes of representation, so that blockage

---

8) An impetus for unpacking is that the GP is inherently unstable and incomplete, and is driven toward closure or repose. A grammatical construction is almost by definition a state of repose. Thus a grammatically complete surface structure is the target and stop-order for the unpacking process. The process continues through interruptions and re-starts, if need be, until a well-formed surface structure or locally acceptable approximation to one is reached; then unpacking.
of one mode can often be compensated for by the other mode. The speaker uses imagery to seek information packages that are suitable for linguistic encoding. Thus if encoding encounters an obstacle, imagery can guide the process to a point where linguistic encoding can recommence. In the GP, as I have explained, analytic and analogic meaning systems are simultaneous and packaged into bundles with onsets and offsets (the GP unit, while bundled, is far from isolated from context, as explained above).

The unbounded IPH offers advantages in accounting for certain phenomena, especially mismatches of spoken and gestured depictions in ongoing speech performance and in children's conservation and arithmetic performances, as described by Goldin-Meadow and colleagues (Goldin-Meadow et al., 1993). It is easy to see with the IPH how one thread is able to take over and represent meanings not accessible to the other thread. The concept of a GP on the other hand is more appropriate for expressing communicative dynamism and differentiation in context since it provides something that can be differentiated as a unit from the background. A possible relationship between the IPH and GP is to see them as points along a continuum of communicative and operational effectiveness. It could be that GPs break down temporarily during mismatch phases of development, whereupon the situation described by the IPH takes over. While the GP is modeled on the mental growth of children it may, paradoxically, not apply during children’s transition points in mental growth. The conceptual instability that characterizes these phases could be a force breaking the GP apart. After negotiating the mismatch (e.g., attaining a new level of cognitive organization), GP units reappear at the new level, that is, the new mental growth dimensions become available on the micro-level. This relates the IPH and GP as successive stages. An empirical prediction from this form of relationship is that measures of GP binding power — synchrony during DAF (McNeill, 1992), synchrony during stuttering (Mayberry & Jaques, 2000), and the interchangeability in memory of speech- and gesture-based information (Cassell et al., 1999) — should be weakened in areas of the mismatch. Another implication is that during mismatches communicative dynamism should be conveyed by gesture or speech but not by both (with gesture the more likely vehicle, according to Goldin-Meadow et al., 1993). A third implication is that cognitive instability (e.g., from fatigue or distraction) could erode the GP and bring the IPH to the fore. Finally, in children, GPs should be structured differently before and after mismatches.

Kita also applies the IPH to adult mismatches. A striking phenomenon is a gesture that is repeated several times by a speaker with modifications in a process that leads eventually to a match with speech. The examples show transitory states of disorganization where imagery and categorial content have trouble combining. The ultimate convergence can be analyzed from a GP point of view in terms similar to that offered for child mismatches but with a time scale of disruption and resolution that is over in a few seconds. Consider one of Kita’s examples: an arcing hand showing the path of a catapulted weight which first appeared during a speech hesitation pause, then was repeated in modified form while speech articulated the situation, “um and he gets clobbered by the weight.” In GP terms, the speaker is attempting to converge on a GP which she does manage to do in the final utterance — the image of the flying weight, until then only imagery, is brought into the system of language (categorized) as “gets clobbered.” The IPH applies during the repetition phase, the GP to the convergence. The modification of the gesture in the final GP could show that convergence involved an image-categorial content dialectic (see Dray & McNeill, 1990 for other cases of gesture modification during dialectics).
Other Means of Incorporating Context — Foreshadowing

The IPH takes the form of intertwining threads because this allows gesture to be a path-finder in conceptualization for speaking. Gesture can explore alternative organization of information, which analytic thinking may not be capturing at the moment. This exploratory mode manifests itself as loose semantic coordination between speech and gesture, namely in a so-called mismatch. Another seeming mismatch is the phenomenon of *gestural foreshadowing*, observed in Turkish cartoon story narrations by Aslı Özyürek (personal communication). Foreshadowing might be prominent in languages with so-called ‘free’ or pragmatic word order. In foreshadowing, a gesture is maximally co-expressive with the *following* linguistic segment, not with the segment with which it is synchronous. The phenomenon seems made to order for Kita’s IPH. There seemingly is an initial disconnection of imagery and linguistic content followed by a convergence on a final speech-gesture combination, guided there by gesture. On the other hand, for the GP, foreshadowing might seem problematic. For, there is a systematic separation of imagery and the language categorial meaning that forms the GP. However, I would like to see if there is something structured that points to GPs; that is, if the speaker is thinking in terms of coherent idea units at each moment, and what these units might be and how they take form.

If GPs take part in gestural foreshadowing, we see a new mechanism whereby word order in so-called free word order languages can be controlled. Words in Turkish speech are not uttered at random or one at a time but in a definite controlled sequence with an overarching prosodic contour that suggests an action plan covering the whole sequence of words. The sequence varies with pragmatic value and is not specified by grammatical rules, but it is neither random nor programmed a word at a time. The proposed control process is that the speaker’s *gestures* may act as bridges between the successive components of the utterance. This makes the resolution of a word order a process of thinking while speaking. With gestural foreshadowing half of the GP is in one place and the other half is in a second, later place. The GPs are divided and the gesture bridges the parts. This trick may be especially prominent in ‘free’ word order regimes.

An example. The following is an example of gestural foreshadowing as transcribed by Özyürek (arrows, which have been added, link gestures to their maximally co-expressive linguistic segments):

1. [top bi sekil-de] \(\rightarrow\) Hands hop in place, palms down = *Manner only*
   *ball in one way*

2. [zipla-ya zipla-ya] \(\rightarrow\) Hands hop and move right = *Path + Manner*
   *while hopping*

3. [yuvar-lan-a yuvar-lan-a] \(\rightarrow\) Hand moves right = *Path only*
   *while rolling itself*

4. [sokak-tan] \(\rightarrow\) Hand moves right = *Path only*
   *on the street*

5. [gid-iy0] \(\rightarrow\) Hand moves right = *Path only*
   *goes*

   “ball somehow, hopping, rolling, goes on the street”.
Table 2  GPs in the Turkish example

<table>
<thead>
<tr>
<th>Text Line</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gst. Content</td>
<td>Cause of action</td>
<td>Cause of action</td>
<td>Effect of action</td>
<td>Effect of action</td>
<td>Effect of action</td>
</tr>
<tr>
<td>Lg. Category</td>
<td>Total episode</td>
<td>Cause of action</td>
<td>Cause of action</td>
<td>Effect of action</td>
<td>Ground</td>
</tr>
<tr>
<td>Speech Items</td>
<td>hop in place</td>
<td>hop &amp; move</td>
<td>move to right</td>
<td>move to right</td>
<td>move to right</td>
</tr>
<tr>
<td></td>
<td>ball somehow</td>
<td>hopping</td>
<td>rolling</td>
<td>on street</td>
<td>go</td>
</tr>
</tbody>
</table>

Lines 1 through 5 comprise a single sentence, “top bi sekil-de, zipla-ya zipla-ya, yuvar-lan-a yuvar-lan-a, sokak-tan gid-ijo” ‘ball somehow, while hopping, while rolling, goes on the street’. The terms ‘manner’ and ‘path’ denote motion event components. According to Talmy’s (1985) semantic analysis, information about motion events (such as rolling down the street) can be better analyzed into the fact of motion itself; the directed path along which the motion is taking place; the backdrop or ground with respect to which the path is situated; and the manner in which the motion is carried out (there are other components not figuring in this discussion). Languages differ in what components are fused with others in the lexical patterns of the language. Gestures also can be analyzed as carrying motion event information. As the speaker of our example began Line 1 her hands bounced up and down but stayed in place; they conveyed the manner of motion but not the ball’s path or shape. Then in Line 2 she said, “while hopping”, a linguistic segment co-expressive with the gesture of Line 1. At the same time, she performed a new gesture forming a path-manner combination. This gesture was co-expressive, in turn, not with its own line, but with Line 3, “while rolling itself.” Line 3, in its turn, was accompanied by a pure path gesture that was co-expressive (probably with “on the street” in Line 4 (the Line 4 gesture was also co-expressive with the Line 4 linguistic segment, and this brought gestural foreshadowing to an end). Thus, up to Line 4, gesture imagery was maximally co-expressive with the following linguistic segment at each step. The final 2 lines were co-expressive with their own gestures, as might be expected as the utterance came to an end.

**The example is about cause and effect.** Table 2 lays out the example in a way that reveals more internal structure than might be at first apparent. And it reveals another means by which GPs incorporate information from context. The table is set up to show the thematic elements of the target utterance. To state this theme in summary form, the sentence is an attempt by the speaker to analyze the events in the drainpipe and beyond causally. The several GPs present steps in her analysis of a causal structure. The theme of causes and effects is the source of GPs and ties the sentence together.

The sentence starts at Line 1 with a statement of cause and effect but the cause is on one level and the effect is on another, and the two do not form a semantic unity. From this initial disparity the speaker lays out the causal relations of the motion event. At Line 1, however, the ball’s manner of motion (hopping) is combined with “ball somehow (in one way)”. Hopping does not necessarily cause moving (i.e., translational)

† The quaint terminology, ‘Cause of action’ and ‘Effect of action’, is used to bring out the causal analysis in the speaker’s sentence. In more usual terminology, these are the motion event components ‘manner’ and ‘fact of motion’, respectively.
motion. The "somehow" or "in one way" appears to index the episode of rolling down the street as a whole. The single motion, rolling, is only one component of this total episode. Hence there was a disparity of levels at Line 1 — the gesture was at the motion level while the linguistic content was at the total episode level.

The next step, where the gesture combines both path and manner, brings out another element of cause and effect. Moving in a rolling manner can be analyzed into sub-predicates corresponding to different semantic components (Talmy, 1975; Goldberg, 1997). One component is moving in a certain way or "rolling". A semantically separate component is the fact of motion itself. A further important concept is that one of these motion event components is the cause of the other (Croft, 1999) — the rolling causes the moving. The speaker of the Turkish example seems to have been carrying out an impromptu causal analysis of this kind. Her gesture in Line 1 presented the cause component (manner) and she completed the paradigm in Line 2 by adding the effect component to make a path+manner gesture — cause with effect. In terms of the speaker’s causation project, the gesture in Line 2 was the logical next step.

The gesture in Line 1 was categorized by the linguistic segment "hopping" in Line 2 (the first arrow). The gesture at Line 2 was in turn categorized in Line 3 with "while rolling." The crucial point is that there was the same cause-effect sequence in both imagery and linguistic categorial content. Just as the gesture sequence went from manner (=cause of movement) to manner + path (=cause + effect), speech also went through the same sequence, one step behind. This double pattern suggests a linkage of imagery and categorial content and thus thinking in terms of GPs.

**Inferring the GPs of the Turkish example.** What could the GPs have been and how did one lead to the next? The speaker was thinking of the total motion event from the start, and this theme continued throughout the sentence. The *succession* of GPs was crucial. This succession is the speaker’s causal analysis of the motion event as a whole. Each GP was a departure from its immediate context. Line 1 embodied the disparity of a cause on one level with an effect on a different level. The disparity itself was categorized with the "in some way" hedge, thus suggesting that a disparity of cause and effect was part of the GP plan. The Line 2 verb ("hopping") and its linked path + manner gesture provided the effect for the cause presented in the Line 1 gesture. The Line 2 gesture was in turn categorized in Line 3 to complete the image-category combination. From this point on, only the fact of motion and the path semantic components — that is, only the effect — appeared in gesture form, suggesting that achieving closure on the motion event level cause-effect link had been satisfied.

The remaining lines of the example filled out the causal paradigm for the full scene, with additional linguistic content for the ground and for the fact of motion. Thus starting from the semantic disparity of cause on the motion level and effect on the whole episode level, the sentence ended with a no disparity of level and a motion event representation with complete semantic dressing.

The gesture-speech pairings therefore were not random. The speaker was not loosely aligning linguistic categories and gestural imagery but was forming definite idea units based on a causal analysis of the motion event. The shape of the GPs both on the linguistic categorial side and the imagistic side are predictable from the unfolding context (the causation context), similar to the case in Viv.’s example (2). The speaker, at each moment, was working out a causal logic. The Gesture Content and Language Category rows of Table 2 track this logic, and this logic would have been the basis of the speaker’s thinking in coherent GPs.

**Foreshadowing and word order.** How did the speaker control word order? The ut-
terance was an integrated output, not a sputter of single-words by single-words. Though called ‘free’, the word order is programmed, as evidenced by a single, smooth output. A clue to what might have been going on is provided by the adverbials “while” in Lines 2 and 3. They linked the verbs “hopping” (Line 2) and “rolling” (Line 3) to the preceding gesture — that is, performed a kind of anaphoric function in which the line of text was linked to the imagery that came before. Thus the gestures in one line could be tied to verbs on the next line, verbs that, in fact, were needed to complete the GPs of which the gestures were the imagery side. The linguistic categorization part of each GP was always second. Thus we have the following idea, that in this sentence a GP initiated in Line 1 with the image of something hopping in place was completed in Line 2 with the categorization “hopping” and the GP initiated with the gesture of Line 2 (hopping and moving laterally) was categorized in Line 3 (“rolling”). In other words, GPs were temporally spread out over different speech elements. This explains why gestural foreshadowing might be especially prominent in ‘free’ word order regimes like Turkish. It is a mechanism for controlling word order sequences without grammatical rule systems. Gesture itself can be a source of word order, as it was in this example (Japanese may show similar uses of gesture).

The place of context. As an illustration of how GPs incorporate context, this Turkish example shows that a succession of GPs can diagram larger discourse themes — here, causation. The additional discovery is foreshadowing itself, the use of gestures to guide the succession of GPs.

By delaying the categorization of imagery until the next step and with that step adding more imagery, the sequence of idea units was systematically controlled to move from initial disparity to final causal resolution.

The Moral

All of this implies that every utterance, even though seemingly self-contained, contains content from outside of its own structure. This was seen in both the English and Turkish analyses. This other content ties the utterance to the context at the level of thinking. The GP predicts the context, rather than adds a context as a parameter (the treatment of context in Levelt, 1989, for example). A successful prediction, in turn, validates the GP analysis of the utterance and its origins by achieving the desired quality of falsification: a GP analysis is rejected if the required context cannot be shown. That two contexts could collaborate in English to form one grammatical structure also implies that a sense of grammatical form enters into utterances in piecemeal and oblique ways that do not necessarily follow the rule-governed patterns of the utterance’s formal linguistic description. In Turkish, gestures may have guided thinking to the next GP, also a way of building up a sense of grammatical form piecemeal but with a systematic sequencing in time rather than, as in English, a concentration of contextual oppositions into single moments of speaking, like our case study example (2).

Conclusion — Synthesis

From the Turkish example we see that co-expressive imagery and categorial content are not always simultaneous. What does this imply? When does the GP come into play? An essential question to ask is: Is imagery a thought process in and of itself (as in the IPH) or is it incomplete until categorized linguistically (as in the GP)? In the GP model, imagery and linguistic categorial content must be bound into a unit. Co-occurrence automatically produces the

9) Pointed out by Nobuhiro Furuyama.
10) There is also reverse foreshadowing — linguistic segments foreshadowing gestures. Such cases have also been observed (Özyurek personal communication). It is not clear what this kind of foreshadowing accomplishes for speech — it cannot control a word order envelope — but this is not to deny that it might have other functions as yet unknown. Reverse foreshadowing would not appear to fit the IPH scenario of mental spaces explored with visuospatial thinking before language.
necessary simultaneity for binding categorial and imagistic thinking, but the regularity of temporal foreshadowing in the Turkish example suggests that binding can take place in other ways as well. In the Turkish example, ‘simultaneity’ occurred when memory combined imagery and current linguistic content. The continuity of memory and thought could produce the conditions needed for the binding of the GPs. Languages may differ in what mechanisms they use to produce simultaneity (co-occurrence is the only means in English as far as is known) and thus in the criteria for inferring growth points, at least in part. Perhaps the most fundamental conclusion is that language is inseparable from imagery. This is as true of Turkish as English, and as true of the IPH as of the GP. Contrary to a text-based tradition that has dominated linguistics and psycholinguistics from their historical beginnings, language contains analogical elements that are as defining of human language capacity as the time-honored analytic elements of syntagmatically related words, phrases, clauses, sentences, and texts. The postulate that language and imagery are inseparable explains the power of gestures to open a window onto language and the mind and to reveal the cross-linguistic differences that have been described in this paper. But it is an insight that cannot be deduced from the text-based tradition and requires theoretical concepts of language in new domains.

References


bridge University Press.

(Received 15 Dec. 1999)
(Accepted 15 Jan. 2000)

David McNeill
David McNeill was educated at the University of California, Berkeley (A.B. 1953, Ph.D. 1962). He has taught at Harvard University, the University of Michigan, and the University of Chicago, where he has been Professor of Psychology and Linguistics since 1969. His research on the speech-gesture nexus started in 1980 and has focused on spontaneous gestures in narrative discourse of children, adults, patients with neurological injuries (aphasics, right-hemisphere injured, split-brains, and a totally deafferented speaker who continues to gesture), and monolingual and bilingual speakers of several languages — English, Spanish, Chinese, Japanese, Korean being the major samples. Current work emphasizes ‘growth points’ in speech and gesture production, cross-linguistic differences in thinking for speaking, the gestures of world leaders, and a system of motion detection and measurement to recover three-dimensional gesture, speech, and gaze cues for discourse segmentation — the latter project in collaboration with computer engineers.