Watching the growth point grow

Insa Röpke

CRC 673 "Alignment in Communication", B1-Project, Bielefeld University, Bielefeld, Germany

iroepke@uni-bielefeld.de

Abstract

For almost twenty years, McNeill's theory of generating co-occurring gesture and speech, *viz.*, the growth point (GP) theory, has been discussed. In doing so, one aspect seems to have received little consideration: his use of frame semantics in developing growth point theory ([4]: 254-256). This is the idea that the linguistic category of the GP could be linked to a frame. Additionally, grammatical patterns are attached to the frame which, in turn, lead to an utterance plus, optionally, a gesture. Since this idea is worthwhile to think about, I will explicate it.

Index Terms: growth point theory, frame semantics, speech and gesture production

1. Motivation

McNeill introduces the term "growth point" (GP) in order to explain the underlying cognitive process of producing co-occurring gesture and speech. This concept of growth points and speech-gesture production to be described later on led to several debates in the last two decades. Some researchers use this theory to explain the "between-person coordination of speech and gesture" ([3]: 349) or apply it for "[i]mplementing a non-modular theory of language production in an embodied conversational agent" ([8]: 425/426). Others, such as de Ruiter, have had doubts about McNeill's concept because it "does not give any account of how (in terms of processing) growth points develop into overt gestures and speech" ([2]: 306).

My interest follows a track similar to the concern expressed in the last quote: It seems to be *prima facie* not evident how exactly a GP is supposed to be able to "unpack" an utterance and a gesture. However, as I came across the passage in which McNeill links the growth point theory to frame semantics, my idea of the so-called "unpacking" became more feasible. Hence, my aim for this paper is to explicate the role of frame semantics within the development of an utterance from the GP, i.e., within the "unpacking" of a growth point.

For this purpose, I will first explain which kind of gestures are considered when we talk about generating gestures (section two) and what growth points are (section three). Since the context of an utterance plays an important role for its production, I will briefly say something about McNeill's theory of narrative, since his data are mostly narrative, and about "communicative dynamism" (section four). Finally, I will try to explicate how a GP can lead to an utterance and I will illustrate this by showing how someone could empirically infer the GP (section five).

2. The kind of gestures which were discussed

With the term "gestures" McNeill refers to gestures at the top of Kendon's continuum (see figure 1), i.e., gesticulation. Among other things, the continuum is set up by how much the properties of different gesture types resemble linguistic properties ([4]: 37). The bottom of it is made up of sign languages that have properties which come very close to verbal language, i.e., that they have standards of form, a syntax and so forth.

In contrast, gestures at top are "idiosyncratic" spontaneous movements which co-occur with speech ([4]: 37). The term "idiosyncratic" does not mean that gestures from different speakers share no characteristics or that they are unique ([7]: 157). However, it suggests that gestures have no standards of form and are rather created spontaneously and individually ([4]: 41). The idiosyncracy also includes that the form of the gestures is mostly driven by meaning and not by convention ([7]: 143). Moreover, gesticulation only occurs in the presence of speech, whereas emblems, such as the OK sign, are also used without speaking ([4]: 37).

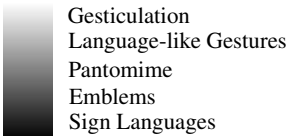

Figure 1: *Kendon's continuum (redrawn after [4]: 37)*

3. The concept of a growth point

McNeill assumes that an utterance and its structure do not emerge at once but develop in a certain order ([4]: 219). He is also impressed with the fact that iconic gestures show a high degree of similarity even across different languages ([4]: 221/222): When describing the same event, speakers used a similar gesture in the same temporal relation with a linguistic segment of an equivalent type, although they used different languages with substantially different grammars. This interesting common feature of speech-gesture synchrony indicates, from his point of view, that these speech-accompanying gestures appear at a level where utterance formation has a cross-linguistically common starting point which has something to do with thought, memory and imagery. Moreover, the speech-gesture synchrony of the stroke, i.e., the meaningful part of the gesture, and the co-occurring linguistic segment(s) seem to include semantic, pragmatic and phonological/kinesic synchrony ([4]: 26-29). Beyond that, different experiments give the impression that this synchrony is very persistent, even if, for instance, speech timing is manipulated ([7]: 145). All of this suggests that speech and gesture are closely linked to each other.

With regard to these assumptions, McNeill introduces the concept of a "growth point". According to him, a GP is the starting point or the primitive/earliest stage of an utterance, but usually not the first uttered segment ([4]: 219/220). Its particu-

lar property is that it is a unit which contains both imagery and linguistic categorial content which is owed to the tight linkage of gesture and speech. This hypothesized unit is the smallest unit that has properties of both kinds of information and cannot be further decomposed ([7]: 144). Hence, the GP is also called the speaker's irreducibly composite minimal idea unit ([4]: 219/220). The term "idea unit" further indicates that the GP is a unit of thought ([5]: 106), since, as we will see, its content is the novel thought in the current context ([4]: 220).

As mentioned above, the linguistic side of the GP is not necessarily the full surface utterance. It also need not include the phonological signifier but is a semantically interpreted, coded segment based on the categories of the speaker's language, hence, it is socially-constituted ([4]: 221). Beyond that, it is characterized as "analytic" and "segmented". It is "segmented" because in language the meaning of the whole depends on the meaning of its parts ([4]: 38) and "analytic" since distinct meanings are linked to distinct words ([4]: 19). The imagistic side is a holistic, idiosyncratic image which is schematic and thus, related to the speaker's idiosyncratic meaning system ([4]: 220/221 & 246). It is also characterized as "global" and "synthetic" since in case of gestures (and hence in case of the GP-image) the parts of gestures are determined by the whole gesture (global) and different parts are synthesized into a single gesture (synthetic) ([4]: 41). Moreover, the imagistic representation in the GP interacts with the linguistic one ([4]: 218) and thereby they influence each other.

Additionally, the GP is considered to be a "psychological predicate" and does not (always) coincide with the grammatical predicate. The term "psychological predicate" is adopted from the Soviet psychologist Vygotsky and it rather labels the newsworthy, contrasting element ([7]: 145) or, in other words, the content of the new departure of thought in the current context ([4]: 220). Besides, calling the GP a psychological predicate also highlights the relevance of context, since, trivially, without context there is no newsworthy element. The context or contextual background is, from McNeill's point of view, a speaker's mental construction and hence under his/her control. It is continuously shaped by the speaker to highlight the contrast between old elements and a possible new one ([7]: 145/146).

Above and beyond that, due to its property of uniting opposed representations (imagistic & global-synthetic *versus* linguistic & segmented-analytic), a GP is unstable ([4]: 218-220). It is unstable since the representations seem to contradict each other. This instability is crucial because it enables the GP to grow, i.e., to initiate a process that generates a surface linguistic constituent and, optionally, a gesture ([4]: 236). However, it is important to stress that not every GP automatically leads to a surface sentence structure, some utterance formations have underlying competing growth points, only one of which can win the competition ([4]: 230 & 234). This generation-process will be described in more detail in section five. For now, the concept of a growth point is summarized in figure 2.

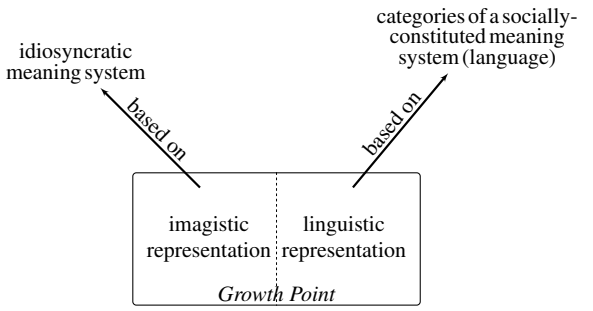

Figure 2: *Growth Point* $=_{Def.}$ *minimal idea unit, psychological predicate*

4. Gesture (types) and discourse

If the GP is the underlying starting point of utterance formation, the formation must have something to do with a differentiation from a background ([7]: 145), since the GP is the newsworthy element. Due to the fact that McNeill's data are almost exclusively narrative, the discourse context from which the examples are taken is a narrative one. Hence, it is important to give a short account of McNeill's idea of the theory of narrative:

As we saw in section two, McNeill deals with gesticulation, but even if such gestures are idiosyncratic, they can be categorized into different gesture types, such as iconic gestures. McNeill claims that these types reflect the discourse functions of the sentences they accompany: Gestures reveal something about the position the narrator takes, the level of narration and the "communicative dynamism" (CD) ([4]: 183/184). The CD is the "extent to which the message at a given point is 'pushing the communication forward'" ([4]: 207).

A narration has not only one but three levels. McNeill distinguishes between a narrative, a metanarrative and a paranarrative level ([4]: 185/186): Narrating at the *narrative level* means to refer to the (fictional) events of the world of the story. These events appear to the listener as presented in their actual order (story line). At the *metanarrative level* one does not refer to the fictional events but to the structure of the story. These narrations are not committed to the actual order of events but are rather comments on the story as a whole. The *paranarrative level*, at last, consists of narrations of the narrator's own experience of reading the book, watching the cartoon or film or telling the story. An example of the latter would be the utterance "Last night I saw a really funny movie", whereas the sentence "In the first scene, Sylvester chases Tweety" belongs to the metanarrative level.

McNeill states that not all gesture-types appear at all levels, but that some types are typical for certain narrative levels. This narratological structure can be seen in figure 3.

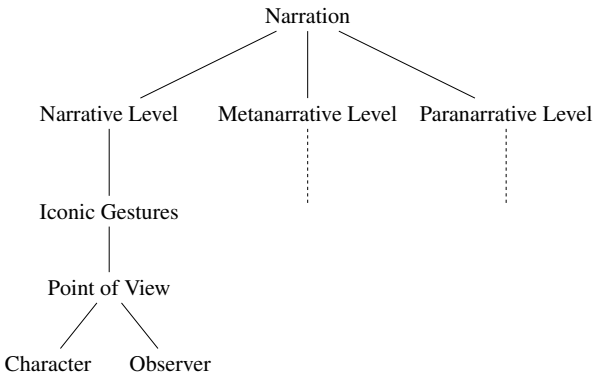

Figure 3: *Part of McNeill's narratological structure of gesture (based on [4]: 189)*

At the narrative level events are retold and this kind of "iconic" relationship is, according to McNeill, mostly accom-

panied by iconic gestures ([4]: 190-192). These gestures can further be differentiated into gestures which take a character's point of view (C-VPT) and ones which take an observer's point of view (O-VPT) and thereby reveal something about the narrator's position. In a gesture with a C-VPT the narrator's hand presents a part of the character's hand etc. and in a O-VPT gesture the hand shows the character as a whole. McNeill suggest that C-VPT gestures are somehow related to the salient events, while O-VPT appear when the event is more peripheral. The O-VPT gestures can also be differentiated with regard to perspective, however, to go into this in detail would take us too far afield. Likewise, I will not deal with the gesture types at the meta- and paranarrative level, since the example which I will use to explain the speech-gesture generation belongs to the narrative level.

The communicative dynamism can be seen on every narrative level and the peak of CD seems to occur when an element is focused upon and other elements fade into the background ([4]: 207). The CD can be measured, for instance, by looking at a scale of linguistic elements which can be presented in a continuum (see figure 4), organized from most continous/predictable (top) to least continous/predictable ones (bottom). The least predictable the element the higher the CD:

Figure 4: *Continuum of linguistic forms used for designating (based on [4]: 210/211)*

At the peak of CD gestures are likely to occur, whereas at low CD they do not occur, since the gestures accompany "rhematic" rather than "thematic" references ([4]: 208/209): The "theme" is, roughly speaking, the information which is known and the "rheme" is the new information. Since the GP is characterized as the newsworthy element, the GP seems to contain rhematic rather than thematic elements. McNeill concludes from this that a gesture indicates the highest CD in its respective utterance. It is important to note that when analyzing gestures, McNeill concentrates on the stroke, i.e., the meaningful part of the gesture. Hence, it is the stroke that indicates the peak of CD.

To summarize, McNeill is convinced that discourse functions, such as the CD and the level of narration, are an integral part of the utterance formation and that analyzing gestures is a good way to infer such discourse relationships ([4]: 184). If a gesture is used, its type (iconic, metaphoric etc.) reveals something about the level of narration and the stroke of the gesture suggests that the CD is high at that moment.

5. From growth point to gesture and utterance and *vice versa*

McNeill has the general idea that the utterance structure is built up around the GP und is unpacked into a hierarchical, linear-segmented linguistic structure ([4]: 222). In this section, we will see how this utterance formation may work. Moreover, I will illustrate this theory by using an example to show how to get from an utterance to the underlying GP.

5.1. Gesture and utterance formation

With regard to the previous sections, we are able to note four aspects that must be taken into account when explaining the speech-gesture production, i.e., the unpacking of a GP: (i) *context-dependence*: The unpacking must be related to the discourse context. (ii) *language-dependence*: The unpacking of an utterance depends on the respective language of the speaker, since each language has different grammatical patterns for certain concepts ([4]: 222). Thus, there is no necessary association between certain kinds of GP and types of grammatical structure ([4]: 230): In case of a similar GP the resulting utterances from two speakers do not have to be similar. (iii) *interaction*: The unpacking must explain how the image representations and the linguistic ones interact while generating the surface utterance and gesture. (iv) *preservation of core statement*: While unpacking, the main significance of the idea unit, i.e., the GP, must be preserved.

First of all, it is important to state that, according to McNeill, the order of utterance formation is not necessarily the word order of the surface utterance ([4]: 219). Instead, the formation starts with the GP. Its linguistic side need not be the first word uttered, but can be nearly every linguistic element, from verb particles to proper names and noun phrases to adverbials and so forth ([4]: 227). Thus, as we can see, McNeill's model differs fundamentally from approaches which include a left-to-right generating structure and from those which assume that the core of a sentence is its verbal phrase ([4]: 232 & 247).

The part of McNeill's approach I found difficult to understand was how a grammatical structure can evolve from something like a growth point if the verb can be excluded from the GP, thus implying that the verb is not always the element from which things are unpacked. If the linguistic element of the GP was always the verb, an unpacking could be imagined, since a verb, such as "climb up", could, at least, unpack an agent and a location as its arguments. The position of the elements in the surface utterance can be indicated by indices; the elements can be filled by contextual information or the imagistic part of the GP:

$$climb - up_{\mathrm{V}_2}(Agens_{\mathrm{NP}_1}, Location_{\mathrm{NP}_3})$$

In the end, the verb "climb up" plus an image of, for instance, the cartoon character Sylvester climbing a mountain could unpack the following utterance:

$$Sylvester_{\mathrm{NP}_1}\ climbs\ up_{\mathrm{V}_2}\ a\ mountain_{\mathrm{NP}_3}$$

Since utterance formation does not work in that way, we should have a closer look at McNeill's conception of the unpacking of a GP: According to McNeill, the beginning of an utterance formation could be located when the speaker's focus shifts, i.e., when some aspects are more interesting for him than others. The source of such shifts are desires, needs, interests and emotions which are called "disruptive force" and activate linguistic segments as relevant, interesting or appropriate ([4]: 238/239). At this point either a single GP or a set of competing growth points emerges. The "fittest" GP survives, where "fitness" means that the selected GP incorporates new information at a greater rate or makes better use of imagery and language fragments *et cetera* ([4]: 234). Then, the instability of the (selected) GP leads to a conflict between the imagistic and the linguistic side which is resolved at the level of surface utterance ([4]: 247). A conflict arises because the representations seem to contradict each other, as we saw at the end of section three.

The linguistic side of the GP is essential for solving the conflict: It can be understood as linked to a *frame* ([4]: 253).

In cognitive science frames are conceptual units of knowledge which consist of open slots, default values and fillers ([9]: 441). For instance, the linguistic segment "buy" is related to a "commercial event frame" which provides open slots for a subject (the buyer), a direct object (the goods) and constructions like "from x" (the seller) and "for x" (medium of exchange) ([4]: 255). There are theoretically many open slots in the frame "buy", for example, the location at which the purchase takes place, the reason for the purchase, and so forth. However, the default values, i.e., the values that are typically attached to the frame, are "the buyer", "the goods", "the seller" and "medium of exchange". Since, dependent on context, there may be different combinations of relevant open slots, it is more plausible to think of a set of (commercial event) frames rather than of one (commercial event) frame, even if McNeill does not talk about sets of frames. For instance, someone could only think about "the buyer" and "the goods" and could leave other possible open slots aside. It also may be that frames have mandatory and optional open slots. The slot "the buyer" might be mandatory rather than "medium of exchange". The buy-frame which has open slots for the buyer, the goods and the seller may be as follows:

$$buy\ (the\ buyer, the\ good, the\ seller)$$

The relevant open slots can have different concrete fillers: The open slot "the buyer" can be filled by a proper noun, such as "Simon", and "the goods" may be filled by the noun "apples".

A main advantage of frame semantics is that not only verb phrases but nearly every linguistic constituent can evoke frames ([9]: 294). This is very important, since the starting point of an utterance, i.e., the GP, can consist of very different kinds of linguistic constituents. Moreover, frames are compositional. This means that an open slot can be filled with an element that is also linked to a frame. For instance, the noun "apples" is attached to "apple-frames".

The frame concept introduced above is linked to *linguistic information* in various ways. However, McNeill needs to extend the frame concept in order to use imagistic information as well. Additionally, he adds to the concept that frames are attached to certain lexical and syntactical patterns ([4]: 254). In this way, he tries to provide a kind of interface of semantics and syntax. With reference to our last example, the attached patterns can be coded by indices: The slots "the buyer", "the goods", "the seller" and "medium of exchange" are complemented by grammatical categories, such as "nominal phrase", plus a number which indicates the order of the elements. The resulting frame may be the following:

$$buy_{\mathrm{V}_2}(the\ buyer_{\mathrm{NP}_1}, the\ goods_{\mathrm{NP}_3}, the\ seller_{\text{from-PP}_4})$$

Filling the open slots, a resulting sentence may be as follows:

$$Simon\ [the\ buyer_{\mathrm{NP}_1}]\ buys_{\mathrm{V}_2}\ apples\ [the\ goods_{\mathrm{NP}_3}]$$
$$from\ Michael\ [the\ seller_{\text{from-PP}_4}]$$

Since this way of using frames differs from the concept mentioned above, we must keep in mind that from now on the notion "frame" is used to designate a data structure which can contain open slots, i.e., an interface for the interaction of different types of (underspecified) information, e.g., syntax, semantics, pragmatics and gesture information.

If the linguistic meaning category of a GP is linked to a set of frames we can think of the unpacking of a GP as follows: With regard to the imagistic side of the GP and the contextual background the relevant open slots and thus the "fittest" frame of this set might be selected. This means that if, for instance, a frame requires an inanimate subject but the imagistic side provides only an animate one this frame would be excluded. The open slots of the selected frame pattern can then be filled by activating further information from the contextual background and from the imagistic side of the GP.[1] If not all needed information can be extracted from the GP and the background additional conceptualizations are required. The aspects of the image that don't fit into the frame remain for the gesture and thereby complement the utterance or provide fragments for further utterance formations ([4]: 255/256), for instance, in form of background information. McNeill concludes from this that "[t]hinking[2] thus is driven, in part, by the requirements of the frame and its syntactic pattern" ([4]: 256).

In addition, the utterance formation is not unrestricted but, for instance, the rhythmical patterns of the preceding utterances influence the unpacking-process ([4]: 235). If the previous utterance had a strong-weak alternating stress pattern, it is more likely than not that the utterance "under construction" has the same stress pattern. Among these restrictions seems to be a rhythmical pulse as well, which controls the periodicity of the process in intervals between 1 and 2 seconds. McNeill characterizes this pulse as "the motor" for the utterance formation which helps to integrate the stroke with the co-expressing linguistic segment and could thus explain the speech-gesture synchrony ([4]: 241-244).

At the final stage of utterance formation there is a synthesis of gesture and speech in the stroke (if there is a gesture at all). At this point speech, which is a linear-segmented presentation, and the gesture, which is a global one, are combined into one unified presentation of meaning ([4]: 246). If an unpacking of a GP has only a linguistic constituent and no gesture as a result, this indicates that the degree of CD is very low, as we saw in section four, or that there was not enough time to combine gesture and utterance ([4]: 236).

To summarize we should have a look if and how the four requirements which were mentioned at the beginning of this section are fulfilled: (i) *context-dependence*: The unpacking is indeed related to the context, since the information needed for the frames are extracted from it. (ii) *language-dependence*: Due to the fact that each language has different grammatical patterns, the unpacking is actually dependent on the language. (iii) *interaction*: The interaction of the two sides of the GP is given by the requirements of the frame patterns which get information to fill their slots from the imagistic side. (iv) *preservation of core statement*: The main significance of the GP remains because the slots of the frame are automatically semantically coherent with it. A summary is also given in figure 5.

5.2. Applying the unpacking to an example while empirically inferring the growth point

Since the growth point is a theoretical concept, we must look at co-occurring gesture and speech to validate it. According to McNeill, the GP hypothesis is falsifiable, this means that if,

[1] Someone could reply that it is not plausible to extract something from a *holistic* image since it is not further decomposable. However, McNeill seems to use the term "holistic" rather as "global and synthetic" ([4]: 412). The term "global" means that the parts of the gesture (and hence of the image) are determined by the whole gesture and "synthetic" means that the different parts are synthesized into a single gesture (image). With respect to this interpretation, we could think of parts of the image from which we can extract information, even if the meaning of them depends on the whole image.

[2] He talks about "thinking", since he is convinced that thinking is closely connected to the utterance formation ([4]: 247).

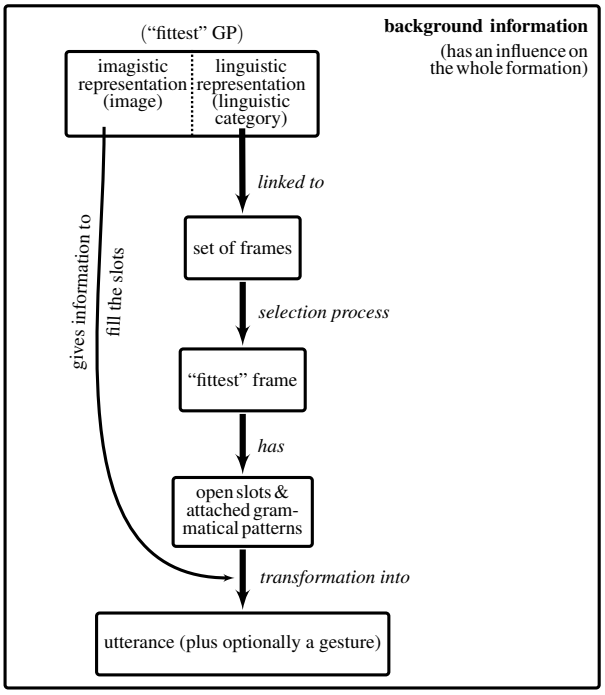

Figure 5: *The Formation of utterances and gestures*

for instance, synchronized speech and gesture cannot be viewed as co-expressing the same idea unit, i.e., when the gesture's meaning clearly differs from the uttered one, the GP hypothesis would be falsified ([6]: slide 39). An example of it would be when the gesture shows a downward-movement while something, such as "up", is uttered and the speaker repairs neither the gesture nor the utterance. Likewise, to make the theory feasible we must also consider real speech-accompanying-gestures.

In order to infer the GP empirically, we must look at the gesture stroke, i.e., the carrier of the meaning of a gesture, and the linguistic segment(s) with which it co-occurs and optionally the following segments if they preserve semantic und pragmatic synchrony ([4]: 220/221): The GP is seen in the image and the co-occurring linguistic segment(s) with which the gesture stroke coincides. Words that precede the stroke would not belong to the empirically inferred GP but they could be unpacked from it. Certainly, it is not always possible to identify the parts of the GP precisely but an approximation of the GP could help as well to conclude something about the underlying utterance formation. Since the description of utterance formation in the last subsection may have been difficult to follow, I will apply the unpacking to an example or, in other words, try to empirically infer the GP of an utterance. The example which I will use for illustration will be the utterance:

"he tries going [up the inside] of the drainpipe" ([4]: 106)

In this example the stroke phase which consists of an iconic gesture showing a blob hand rising (vertically) up with an extended forefinger ([4]: 110) is enclosed in square brackets and the pronoun "he" refers to Silvester. The context of this utterance is the retelling of a Sylvester-Tweety-cartoon ([4]: 191) and, more precisely, the utterance is embedded in a retelling of the scenes in which Sylvester tries twice to catch Tweety while climbing a drainpipe. At first, Sylvester climbs the outside of the pipe, but he is not successful in catching Tweety. Afterwards, he goes up the inside of it ([4]: 367).

The utterance seems to belong to the narrative level of the narratological structure (fig. 2), since an event of the world of the story is described. Appropriately enough, the observed gesture is an iconic one. Moreover, taking the narrative context into account, the interiority seems to be the newsworthy element and thus is highlighted by the stroke.

An application of McNeill's concept may work as follows: The beginning of this utterance formation starts with the speaker's shift of focus from the previous sentence topic (which could have been Sylvester's failure to catch Tweety Bird) to Sylvester's new attempt to climb up the pipe on the *inside*. This newsworthy element of interiority together with the image of upward-climbing constituted the growth point. It consists of a noun phrase, such as "the inside", and an adverb, such as "up". The imagistic side could contain the image of Sylvester climbing up the pipe vertically. Since we have two linguistic categories, two sets of frames are attached which interact with each other: The set of "up"-frames interact with the set of "inside"-frames. An "up"-frame can have open slots for a verb of motion, an inanimate (e.g., a lift) or animate agent, a location, speed, and a direction (since upwardness need not include straight "verticalness") *et cetera*. With regard to the image of Sylvester climbing the pipe vertically a frame should be selected which has, at least, open slots for a verb of motion, an agent, a location and a direction. The attached syntactical pattern of this frame which includes the linguistic category and an order can furthermore be coded by indices:

$$up_{\mathrm{P}_3}(Agens_{\mathrm{NP}_1}, Motion_{\mathrm{V}_2}, Location_{\mathrm{NP}_4}, Direction)^3$$

The selected frame for "inside" has, at the least, an open slot for a location which specifies which inside is described ("of x"):

$$the\ inside_{\mathrm{NP}_1}(Location_{\text{of-PP}_2})$$

The interaction of these two frames is an unification in which the argument "location" of the "up"-frame is filled with the "inside"-frame. Then there remain the following required elements: an agens, a verb of motion, a location and a direction. These pieces of information can be extracted from the imagistic side. The image of Sylvester provides an agens and related to the narrative context it gives us an unstressed pronoun rather than a proper noun. Since Sylvester's name was introduced before, it is not a newsworty information and is hence more likely to provide a low CD (see figure 4) which explains the use of an unstressed pronoun. The verb of motion can also be taken from the image: With reference to the climbing-motion the speaker can use verbs like "climb", "go(ing)", "come", "crawl", or "barreling" (cf. McNeill's examples [4]: 106-108). The decision which verb is selected presumably depends on the preceding utterance and the speaker's vocabulary. One reason why the speaker selected "going up" may be that he could have used the verb "climbing" while describing Sylvester's first attempt and sees no need to choose this word again. Likewise, the location can be extracted from the image and the designation of the drainpipe also depends on similar things.

The function of the gesture is to highlight both that Sylvester crawls *up* and that he does this on the *inside* of the pipe. Moreover, it should show the direction of the upward-movement which should be vertical but this is not expressed in

[3]The slot "direction" has no index since it is not filled by a surface construction but by the speech-accompanying gesture, as we will see later on.

the utterance, since an upward-climbing need not to be straight vertical, but can be "sloped". To explain the gesture formation, we should extend the frame idea to gestures as well. However, it would be speculative to develop a frame *ad hoc*, but we are able to state that some properties of the arising gesture are not accidental but due to the growth point. In doing so, I will concentrate on the properties of a one-handed gesture, but, certainly, properties for a two-handed gesture could be considered as well.

First of all, we can state that the gesture must include a movement of the wrist with the direction "up" and this movement should be in straight line which can be described by the value "line". In combination this is a vertical movement which fills the last required element of the frame. To highlight that Sylvester climbs up the *inside* of the drainpipe, the gesture needs to fulfill further properties: According to McNeill's narratological structure (the accuracy of which will not be discussed in this paper), the emerging gesture should be an iconic one (as it was the case in our example), since the speaker wants to say something on the *narrative* level. Generally, an iconic gesture phrase can have different practices (for the following descriptions cf. [1]: 7-9). One practice is, for instance, "drawing" which is the use of a finger in order to draw the contours of an object. However, the only practice which can be applied to our "upward-inside"-gesture is "modeling & shaping". This category contains dynamic gestures which both model an object and shape something in the gestural space. Thus, gestures of that category are able both to capture the upward-movement (since they are dynamic) and to show the interiority by modeling an object in a certain way. A general example of a "modeling & shaping"-gesture would be the following: The gesticulating person clenches his/her fist in order to model a car and uses this fist as well to illustrate the curvy path which the car must drive along while moving the fist along an arc. "Modeling & shaping" are further gestures from the O-VPT, since the hand is not used to pantomime the action of someone. Beyond fixing the practice category, the arising iconic-gesture is unspecified with regard to highlighting the interiority, since different realizations are possible. For instance, the person in our example used his extended index finger while moving his wrist and may thereby show the interiority ([4]: 108). But certainly, this is not the only way to realize it.

The required properties of the gesture are presented by an attribute value matrice (AVM) in figure 6, which, in principle, can be translated into a frame.

"upward-inside"-gesture	
WRIST MOVEMENT	*up*
PATH OF WRIST	*line*
PHRASE	*iconic*
PRACTICE	[O-VPT *modelling & shaping*]

Figure 6: *AVM for the "upward-inside"-gesture (unspecified with regard to the "inside"-character)*

To sum up, we have tried to empirically infer the GP of the utterance "he tries going [up the inside] of the drainpipe" while assuming that the stroke and the co-occurring linguistic elements "up" and "the inside" together constituted the GP. The unpacking of the GP can be described as considering their linguistic element as linked to a set of frames. The open slots of the selected frame require elements which can be taken from the imagistic side of the GP and the contextual background. The syntax of the utterance can also be gathered from the frame, since it can include indices which fix the order of the elements and their linguistic categories. Due to the requirements of the frame, the gesture is not unrestricted, but must fulfill certain properties, such as to include the upward-movement of the wrist. Moreover, while showing the vertical direction, the gesture complements the utterance.

6. Conclusion

As we saw in the last summary, frame semantics together with attached grammatical patterns provide a possibility to get from growth points to utterances and gestures. Certainly, frame semantics may not be the only way to fill the gaps between growth points and the surface utterance and the co-occurring gesture. Similar approaches which deal with syntax and semantic interfaces may fit as well.

Anyhow, future work is needed to elaborate either a more detailed account of the role of frame semantics or to try to apply another syntax-semantic-interface. Of course, these approaches would have to develop a more accurate characterization of the selection processes underlying the "fittest GP" and the "fittest frame" than I did in this paper. However, I've tried to show that using such interfaces might be an interesting concept for speech-gesture production, especially with regard to the growth point theory.

7. Acknowledgements

First and foremost, I'd like to thank Hannes Rieser for his support and important impulses. I'd also like to thank Florian Hahn, Shane Sale and two anonymous reviewers for their helpful comments and suggestions. Last but not least, the work on this paper has been supported by the CRC 673 "Alignment in Communication".

8. References

[1] Bergmann, K., Fröhlich, C., Hahn, F., Kopp, S., Lücking, A., and Rieser, H., "Grobannotationsschema", Ms., Bielefeld University, 2007.

[2] de Ruiter, J., "The production of gesture and speech", in McNeill, D. [Ed], "Language and gesture", 284-311, Cambridge University Press, 2000.

[3] Furuyama, N., "Prolegomena of a theory of between-person coordination of speech and gesture", International Journal of Human-Computer Studies 57, 347–374, 2002.

[4] McNeill, D., "Hand and Mind", The University of Chicago Press, 1992.

[5] McNeill, D., "Gesture and Thought", University of Chicago Press, 2005.

[6] McNeill, D., "Growth Points and Modeling", McNeill Lab, University of Chicago. Online: mcneill-lab.uchicago.edu/pdfs/modeling_or_not.pdf, accessed on 22 May 2011.

[7] McNeill, D. and Duncan, S., "Growth points in thinking-for-speaking", in McNeill, D. [Ed], "Language and gesture", 141-161, Cambridge University Press, 2000.

[8] Sowa, T., Kopp, S., Duncan, S., McNeill, D. and Wachsmuth, I., "Implementing a non-modular theory of language production in an embodied conversational agent", in Wachsmuth, I., Lenzen, M. and Knoblich, G. [Eds], Embodied Communication in Humans and Machines, 425-450, Oxford University Press, 2008.

[9] Ziem, A., "Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz", Walter de Gruyter, 2008.